

AMLaP 2018

Berlin, Germany

Architectures and Mechanisms for Language Processing 2018

Proceedings

Editor: Pia Knoeferle
Humboldt-Universität zu Berlin
Germany

<https://amor.cms.hu-berlin.de/knoeferp/AMLaP2018/Home.html>

Contents

1 Preface	3
2 Organization	4
3 Reviewer committee	4
4 Important Information	5
5 Sponsors	6
6 Social program	7
7 Oral Presentations at a Glance	8
8 Poster Presentations at a Glance - Thursday Sept 06	12
9 Poster Presentations at a Glance - Friday Sept 07	18
10 Poster Presentations at a Glance - Saturday Sept 08	24
11 Oral Presentations - Abstracts	30
12 Poster Presentations - Abstracts Thursday	68
Poster Presentations - Abstracts Friday	158
Poster Presentations - Abstracts Saturday	245
Index	334

1 Preface

I am delighted to welcome you to Berlin for the 2018 conference on Architectures and Mechanisms for Language Processing. This year counts the 24th edition of AMLaP.

Proceedings The proceedings contain the abstracts of the keynotes, and of the accepted oral and poster presentations. The abstracts were included in the format submitted by the authors.

AMLaP 2018 unites experimental, computational, and theoretical perspectives on the mechanisms implicated in human language processing. Contributions which explicitly relate experimental findings to computational mechanisms were especially encouraged - as were contributions that bring together different methodological approaches (e.g., response times, eye tracking, EEG, corpora, fMRI and any combination of these and related approaches).

Selection criteria AMLaP 2018 had a total of 415 submissions. Of these, 81 had to be rejected. To qualify for an oral presentation, submission had to score > 5 averaged across the reviewing categories (on a scale = 1-7, with 1 being very low and 7 very high in assessment). I in addition considered reviewer recommendations for talks, and the topic fit.

Keynotes I am delighted to welcome our four invited speakers, Sonja Kotz, Asifa Majid, Matthew W. Crocker, and Ken McRae (Figure 1).

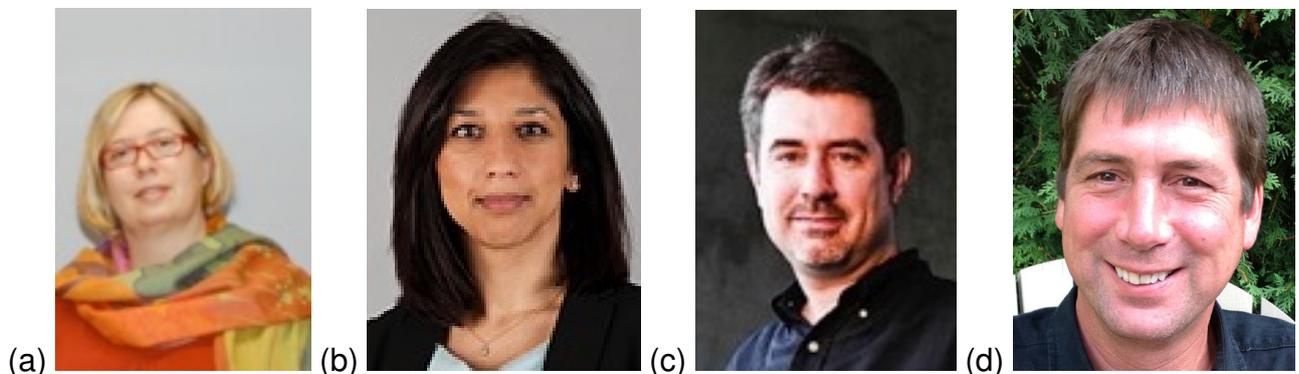


Figure 1: (a) Sonja Kotz, (b) Asifa Majid, (c) Matthew W. Crocker, and (d) Ken McRae

Enjoy the conference!

Pia Knoeferle (Conference Chair)

2 Organization

- Conference Chair: Pia Knoeferle
- Technical support: Dipl.-Ing. Carsten Schlieve and in-house support from the Titanic hotel
- Organization and registration: supported by the Humboldt-Innovation GmbH (<https://www.humboldt-innovation.de/de/aboutus.html>, Ms Stefanie Karbe)
- Further Acknowledgements: I would like to thank the onsite organization team for their assistance with printing the program overview, setting up poster stands, helping out at the registration desk and in the auditorium (Jakob Wunsch, Daniela Palleschi, Camilo Rodriguez Ronderos, Maria Nella Carminati, and Luise Henneberg).

3 Reviewer committee

I am very grateful to the members of the AMLaP 2018 reviewing committee, whose dedicated and timely reviews were vital in ensuring a high-quality program.

Adrian Staub, Adriana Hanulikova, Aine Ito, Alan Garnham, Alba Rodriguez, Amit Almor, Andrea E. Martin, Andrew Kehler, Annie Tremblay, Anouschka Foltz, Asifa Majid, Barbara Hemforth, Carson Schütze, Charles Clifton, Cheryl Frenck-Mestre, Chris Cummins, Christina Kim, Christoph Scheepers, Claudia Felser, Daphna Heller, Diogo Almeida, Don Mitchell, Douglas Roland, Duane Watson, E. Matthew Husband, Edward Gibson, Ellen Gurman Bard, Emmanuel Chemla, Ernesto Guerra, Eva Belke, Eva Wittenberg, F.-Xavier Alario, Florian Schwarz, Franklin Chang, Francois Rigalleau, Hannah De Mulder, Hannah Rohde, Hans Rutger Bosker, Harm Brouwer, Heather Ferguson, Helene Kreysa, Hiroko Yamashita, Hugh Rabagliati, Ian Cunnings, Jean-Pierre Koenig, Jelena Mirkovic, Jennifer Culbertson, Jonathan Brennan, Juhani Jrvikivi, Katherine White, Katja Mnster, Katy Carlson, Ken McRae, Kiel Christianson, Lars Konieczny, Laura Speed, Laurence White, Leigh Fernandez, Les Sikos, Manon Jones, Mante S. Nieuwland, Marc Brysbaert, Marc Swerts, Margaret Grant, Maria Pinango, Maria Staudte, Markus Bader, Martin Corley, Masako Hirotani, Matt Goldrick, Matthew Crocker, Michael Wagner, Michele Burigo, Mikel Santestaban, Ming Xiang, Nayoung Kwon, Neal Pearlmutter, Nicola Molinaro, Nicole Gotzner, Nina Kazanina, Pdraic Monaghan, Pdraig O'Seaghda, Paul Engelhardt, Petra B. Schumacher, Pia Knoeferle, Robert Hartsuiker, Ruth Filik, Sandra Pappert, Sarah Brown-Schmidt, Saveria Colonna, Shane Lindsay, Shanley Allen, Shari Speer, Shelia Kennison, Shravan Vasishth, Sonja Kotz, Stephani Foraker, Tessa Warren, Thomas Pechmann, Titus von der Malsburg, Vicky Lai, Victor Ferreira, Whitney Tabor, William Horton, Wing-Yee Chow, Youngon Choi, Yuki Kamide, Zhenguang Cai

4 Important Information

- Venue: AMLaP 2018 takes place at the Titanic Hotel Chaussee Berlin (Chausseestrasse 30, 10115 Berlin, Germany).
- Registration: The registration desk is open from 8am-4pm on Thursday and Friday and from 9am-10am on Saturday. You can pick up your badge and the short program there. Please let the registration desk know in case you need confirmation of attendance when you pick up your badge.
- Poster presentations take place in the auditorium. Poster stands each carry a number that you can also find in the program. Please put up your poster at the poster stand (check your number in the program) and present them over lunch. Don't forget to remove your poster at the end of the conference day.
- Coffee breaks and lunches will take place in front of the auditorium. They are included in the registration.
- Free Internet is available at the venue.

Travel

- From Tegel to conference hotel:
 - Public transport: The best connection into the city center on public transport is the Bus TXL. It runs every 5-6 minutes and it takes approximately 30-40 minutes to reach the main train station ("Hauptbahnhof"). From the Hauptbahnhof take the M8 or M10 (direction "Warschauerstr" or "Ahrensfelde", exit at "Naturkundemuseum").
 - Taxi: typically around 20-25 Euros, ca. 20-30 min. ride
- From Schönefeld to conference hotel: You can take the S9 (direction "Spandau Bhf") to "S+U Friedrichstr. Bhf" and then change to the U6 (direction "Alt-Tegel", 1 intermediate stop); exit at "Naturkundemuseum". Alternatively, take the S45 (direction "Südkreuz Bhf") to "S+U Tempelhof". At Tempelhof change to the subway (Ubahn) and take the U6 to "Alt-Tegel"; exit at "Naturkundemuseum".
- Public transport stations close to the conference hotel:
 - Ubahn: The subway station ("Ubahn") closest to the hotel is "Naturkundemuseum";
 - SBahn: the closest Sbahn stop is "Nordbahnhof".
 - On tram: The hotel is located a short ride on the M8 or M10 from the main train station (from ?Hauptbahnhof?, direction "Warschauerstr" or "Ahrensfelde", 2 stops, exit at "Naturkundemuseum").
 - On foot: The hotel is located a 10 minute walk from the main train station ("Hauptbahnhof").
 - From conference hotel to Friedrichstr.; Friedrichstrasse Bahnhof is 2 subway stops away (U6, direction "Alt Mariendorf").

5 Sponsors

I thank the *German Research Council (DFG)*, the *Berlin School of Mind and Brain*, the *Sprach- und Literaturwissenschaftliche Fakultät at the Humboldt-Universität zu Berlin*, and last but not least the *Humboldt-Universität zu Berlin* (Office of the Vice President for Finance, Human Resources and Operations) for supporting the conference.



6 Social program

Conference Dinner The conference dinner takes place on Friday, September 07, 2018 at the restaurant AIGNER (Gendarmenmarkt) in Berlin Mitte. You can find directions in Figure 2.

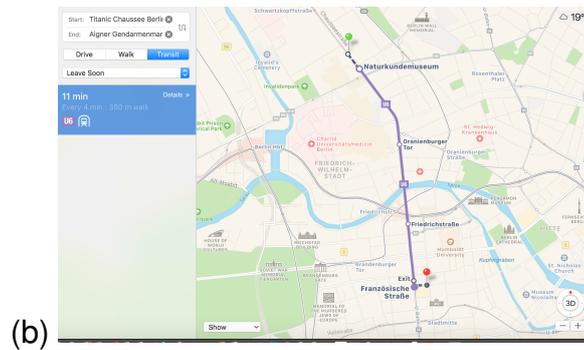
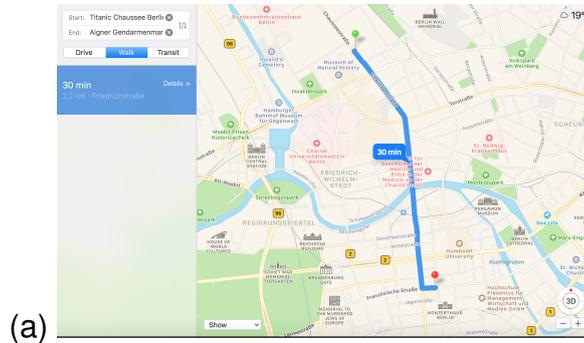


Figure 2: Conference venue to the conference dinner (a) pedestrian and (b) public transport

7 Oral Presentations at a Glance

Thursday, September 06, 2018

from 8:00-16pm

Collecting name tags

08:50am–09:00am		Welcome	
09:00am–10:00am	Keynote	Probabilistic architectures and neurocomputational mechanisms for language processing	Matthew W. Crocker
10:00am-10:30am		Coffee	
10:30am-10:50am		RESISTANCE TO ENVIRONMENTAL VARIABILITY IN LANGUAGE LEARNING: CROSS-SITUATIONAL WORD LEARNING FROM MULTIPLE CUES	Padraic Monaghan, James Brand and Rebecca Frost
10:50am-11:10am	Oral presentations	IMPLICIT CONCURRENT LEARNING OF ADJACENT AND NONADJACENT DEPENDENCIES IN CHILDREN	Lai-Sang Lao, Jens Roeser, Lucy Justice and Gary Jones
11:10am-11:30am		INVESTIGATING IMPLICIT LEARNING AND SURPRISAL EFFECTS IN A STRUCTURALLY BIASED LANGUAGE OVER DEVELOPMENT	Alina Kholodova, Michelle Peter, Caroline Rowland and Shanley Allen
11:30am-11:50am		PREDICTABLE WORDS LEAVE PRODUCTION-LIKE TRACES IN MEMORY	Joost Rommers, Gary S. Dell and Aaron S. Benjamin
		<i>PERCEPTUAL PRIMING AND SYNTACTIC CHOICE IN RUSSIAN LANGUAGE: MULTIMODAL STUDY.</i>	Mikhail Pokhoday, Yury Shtyrov, Christoph Scheepers and Andriy Myachykov
		<i>ABOUT THE IMPORTANCE OF IMPLICIT CAUSALITY TO UNDERSTAND RELATIVE CLAUSE PROCESSING</i>	Céline Pozniak and Barbara Hemforth
		<i>WORD ORDER BEATS MORPHOSYNTAX: INCREMENTAL THEMATIC ROLE ASSIGNMENT IN TAGALOG L1 ACQUISITION</i>	Rowena Garcia, Jens Roeser and Barbara Höhle
11:50-12:00	1st SLAM: 1 minute / slide per poster	<i>AUDITORY STATISTICAL LEARNING RAPIDLY SUPPORTS THE PROCESSING OF LARGER LINGUISTIC CHUNKS ACROSS EARLY CHILDHOOD.</i>	Evan Kidd, Joanne Arciuli, Michael Smithson, Erin Isbilen and Morten Christiansen
		<i>PRIOR LEARNING OF ACOUSTIC CUES BLOCKS LEARNING OF NEW CUES IN NON-NATIVE SPEECH ACQUISITION (DIS-)CONFIRMATION OF LINGUISTIC PREDICTION BY NON-LINGUISTIC CUES</i>	Jessie S. Nixon
		<i>SYNTACTIC AND SEMANTIC PROCESSING IN POOR COMPREHENDERS: EVIDENCE FROM EYE-TRACKING AND COMPUTATIONAL MODELING</i>	Torsten Jachmann, Heiner Drenhaus, Maria Staudte and Matthew Crocker
		<i>LINGUISTIC PREDICTION AND VISUAL ATTENTION DO NOT INTERACT IN READING</i>	Luca Campanelli, Nicole Landi and Julie Van Dyke
			Suzanne Jongman, Marta De Pedis and Ashley Lewis
12:00pm-14:00:pm		Lunch and posters 1	
14:00pm-14:20pm		DIFFICULTIES TRACKING ROLE-REFERENT SWITCHES CAN HELP TO EXPLAIN THE SUBJECT/OBJECT RELATIVE CLAUSE ASYMMETRY	Andrew Jessop and Franklin Chang
14:20pm-14:40pm	Oral presentations	LONG-TERM EFFECTS OF STRUCTURAL REPETITION ON SENTENCE PLANNING?	Zsofia Stefan and Agnieszka Konopka
14:40pm-15:00pm		LINEAR VS. STRUCTURAL INCREMENTALITY IN THE FACE OF SENTENCE PRODUCTION IN CONTEXT	Xiaogang Wu and Johannes Gerwien

15:00pm-15:20pm		EXPECTATION MANAGEMENT IN ONLINE DIALOGUE COMPREHENSION: AN ERP INVESTIGATION OF DUTCH INDERDAAD 'INDEED' AND EIGENLIJK 'ACTUALLY'	Marlou Rasenberg, Joost Rommers and Geertje van Bergen
15:20pm-15:50pm		Coffee	
15:50pm-16:10pm		PARTICIPANT ASSIGNMENT TO THEMATIC ROLES IN TZELTAL: EYE-TRACKING EVIDENCE FROM SENTENCE COMPREHENSION IN A VERB-INITIAL LANGUAGE	Gabriela Garrido Rodriguez, Elisabeth Norcliffe, Falk Huettig, Penelope Brown and Stephen C. Levinson
16:10pm-16:30pm	Oral presentations	ASSESSING THE EFFECT OF ADDRESSEE'S LINGUISTIC COMMUNITY ON SPEAKERS' LEXICAL CHOICES	Anita Tobar, Hugh Rabagliati and Holly Branigan
16:30pm-16:50pm		HIERARCHICAL STRUCTURE PRIMING FROM MATHEMATICAL EQUATIONS TO TWO- AND THREE-SITE RELATIVE CLAUSE ATTACHMENTS IN RUSSIAN	Andriy Myachykov, Christoph Scheepers, Anastasia Galkina and Yury Shtyrov
16:50pm-17:50pm		Beyond language for visual objects	Asifa Majid

Friday, September 07, 2018

from 8:00am-16pm

Collecting name tags

09:00am-10:00am		Event knowledge and semantic processing	Ken McRae
10:00am-10:30am		Coffee	
10:30am-10:50am		CONTRASTING FACILITATION PROFILES FOR AGREEMENT AND REFLEXIVES REVISITED: A LARGE-SCALE EMPIRICAL EVALUATION OF THE CUE-BASED RETRIEVAL MODEL	Lena Jäger, Daniela Mertzén, Julie Van Dyke and Shravan Vasishth
10:50am-11:10am	Oral presentations	HOW THE INPUT SHAPES THE ACQUISITION OF INFLECTIONAL MORPHOLOGY: COMPUTATIONAL MODELLING ACROSS THREE HIGHLY INFLECTED LANGUAGES	Felix Engelmann, Joanna Kolak, Sonia Granlund, Ben Ambridge, Julian Pine, Anna Theakston and Elena Lieven
11:10am-11:30am		N400 AMPLITUDES REFLECT CHANGE IN A PROBABILISTIC REPRESENTATION OF MEANING, EVEN IF INDUCED BY DIFFERENCES IN WORD FORMS: A NEURAL NETWORK MODEL	Milena Rabovsky
11:30am-11:50am		THE MIRRORING EFFECTS OF INFORMATION THEORY BASED AND DISCRIMINATION LEARNING BASED QUANTIFICATIONS OF INFLECTED ADJECTIVES	Dusica Filipovic Durdevic and Petar Milin
11:50-12:00		FALSE POSITIVES IN GROWTH CURVE ANALYSIS OF VISUAL WORLD PARADIGM DATA	Yujing Huang and Jesse Snedeker
		REFERENCE TO QUANTIFIED EXPRESSIONS IN SWEDISH: AN ERP STUDY	Fredrik Heinat and Eva Klingvall
		VARIATION IN FRENCH PARTIAL INTERROGATIVES: SOCIAL MEANING AS A KEY FACTOR TO UNDERSTAND SOCIOLINGUISTIC NORM VIOLATIONS	Gabriel Thiberge and Barbara Hemforth
	2nd SLAM: 1 minute / slide per	LEARNERS GENERATE UNENCOUNTERED NOVEL STRUCTURES DRAWING ON KNOWLEDGE OF A UNIVERSAL	Adam Morgan and Victor Ferreira

	poster	<i>STAIRS4WORDS: A NEW ADAPTIVE TEST FOR ASSESSING RECEPTIVE VOCABULARY SIZE IN ENGLISH, DUTCH AND GERMAN</i>	Florian Hintz, Suzanne Jongman, Marjolijn Dijkhuis, Vera van t' Hoff, Markus Damian, Sascha Schroeder, Marc Brysbaert, James McQueen and Antje Meyer
		<i>GENDER-BIASES IN LANGUAGE PROCESSING: EXPLICIT BELIEFS ABOUT EVENT OUTCOMES VS. IMPLICIT LINGUISTIC EXPECTATIONS</i>	Titus von der Malsburg, Veronica Boyce, Till Poppels and Roger Levy

12:00pm-14:00pm

Lunch and posters 2

14:00pm-14:20pm		CATCHING YOUR EYE: LOW-LEVEL PERCEPTUAL CUES INFLUENCE PRESCHOOLERS' SENTENCE FORMULATION	Laura Lindsay, Holly Branigan and Hugh Rabagliati
14:20pm-14:40pm	Oral presentations	ERPS DO NOT SHOW THAT LEXICAL ACCESS DURING WORD PRODUCTION BEGINS WITHIN 200 MS	Daniel Kleinman and Kara Federmeier
14:40pm-15:00pm		GESTURE INCONGRUITY EFFECTS ON SPEECH PRESERVED WITH VERBAL BUT NOT VISUOSPATIAL WM LOAD: AN ERP STUDY	Seana Coulson and Jacob Mommsen
15:00pm-15:20pm		EFFECT OF DISCOURSE AND ACTION ON VISUAL ATTENTION DURING LANGUAGE PROCESSING	Ruth Maddeaux, Margaret Grant and Daphna Heller

15:20pm-15:40pm

Coffee

15:40pm-16:00pm		VISUAL CUES AND THE GRADED REDUCTION OF REFERENTIAL UNCERTAINTY	Mirjana Sekicki and Maria Staudte
16:00pm-16:20pm	Oral presentations	PROCESSING OF AD HOC METONYMY: EVIDENCE FROM CO-REGISTRATION OF EYE MOVEMENTS AND ERPS	Petra B. Schumacher, Jenny Knowles, Andrea Krott and Steven Frisson
16:20pm-16:40pm		L2 LEARNERS PREDICT AT THE LEVEL OF THE DISCOURSE: EVIDENCE FROM ERP	José Alemán Bañón, Elena Fano and Clara Martin
16:40pm-17:00pm		FORMING UNGRAMMATICAL STRUCTURES SHORT-SIGHTEDLY: LOCAL COHERENCE EFFECTS IN THE VISUAL WORLD PARADIGM AND READING	Yuki Kamide and Anuenue Kukona

19:00pm-22:00pm

Conference Dinner: Aigner am Gendarmenarkt, <http://aigner-gendarmenmarkt.de>

Saturday, September 08, 2018

from 9-10am

Collecting name tags

09:00am-10:00am	Why time and rhythm matter in speech/language comprehension	Sonja Kotz
-----------------	--------------------------------------------------------------------	-------------------

10:00am-10:30am

Coffee

10:30am-10:50am		THE LATERALIZATION OF EXPECTATIONS: EVIDENCE FROM A DIVIDED VISUAL FIELD ERP STUDY	Yoana Vergilova, Les Sikos and Matthew Crocker
10:50am-11:10am	Oral presentations	CONTEXTUAL SPEECH RATE INFLUENCES MORPHOSYNTACTIC PREDICTION AND INTEGRATION	Greta Kaufeld, Wibke Naumann, Anna Ravensschlag, Andrea E. Martin and Hans Rutger Bosker
11:10am-11:30am		ATTENDING FAST AND SLOW 'COCKTAIL PARTIES': UNATTENDED SPEECH RATES INFLUENCE PERCEPTION OF AN ATTENDED TALKER	Hans Rutger Bosker, Eva Reinisch and Matthias Sjerps

11:30am-11:50am		DOES STRESS CLOSE THE LANGUAGE GATE? INHIBITION OF UNCONSCIOUS L1 ACTIVATION	Jennifer Lewendon, Anouschka Foltz and Guillaume Thierry
11:50-12:00	3rd SLAM: 1 minute / slide per poster	<i>A CRACK IN THE CRYSTAL BALL: EVIDENCE AGAINST PRE-ACTIVATION OF GENDER FEATURES IN SENTENCE COMPREHENSION</i>	Ernesto Guerra, Bruno Nicenboim and Andrea Helo
		<i>EFFECTS OF EMOTIONAL SPEAKER FACIAL EXPRESSIONS ON SENTENCE PROCESSING: AN ERP STUDY</i>	Katja Münster, Johanna Kißler and Pia Knoeferle
		<i>LEARNING CONSISTENT GENDER ERRORS IN NON-NATIVE SPEECH</i>	Thomas St. Pierre and Jean-Pierre Koenig
		<i>PARALLEL, CASCADED, INTERACTIVE PROCESSING OF WORDS DURING SENTENCE READING: THE SENTENCE-SUPERIORITY N400 EFFECT</i>	Yun Wen, Joshua Snell and Jonathan Grainger
		<i>SEMANTIC INTERFERENCE AND MORPHOLOGICAL FACILITATION IN NOUN-NOUN COMPOUND PRODUCTION: EVIDENCE FROM EVENT-RELATED BRAIN POTENTIALS</i>	Antje Lorenz, Stefanie Regel, Pienie Zwitserlood and Rasha Abdel Rahman
		<i>NEURAL MECHANISM FOR PRONOUN RESOLUTION IN CHINESE DURING NATURALISTIC LISTENING</i>	Jixing Li, Murielle Fabre, Wen-Ming Luh and John Hale
12:00pm-14:00pm		Lunch and posters 3	
14:00pm-14:20pm		IT'S ALIGNMENT ALL THE WAY DOWN - BUT NOT ALL THE WAY UP	Rachel Ostrand
14:20pm-14:40pm	Oral presentations	WHEN ATTENTION DISTRACTION HELPS RULE INDUCTION: AN ENTROPY MODEL	Silvia Radulescu, Mridhula Murali, Sergey Avrutin and Frank Wijnen
14:40pm-15:00pm		MODELLING OF MISMATCH NEGATIVITY RESPONSE AND NON-NATIVE STATISTICAL LEARNING	Jacolien van Rij, Jessie S. Nixon and Tomas O. Lentz
15:00pm-15:20pm		YOUR EARS OR YOUR BRAIN? NOISE STRUCTURE CAN HIDE GRAMMATICAL PREFERENCES	Suhas Arehalli and Eva Wittenberg
15:20pm-15:50pm		Coffee	
15:50pm-16:10pm		HOW SPEECH RATE NORMALIZATION AFFECTS LEXICAL ACCESS	Merel Maslowski, Antje S. Meyer and Hans Rutger Bosker
16:10pm-16:30pm	Oral presentations	ON PREDICTION OF PHONOLOGICAL AND GRAMMATICAL GENDER INFORMATION	Aine Ito, Chiara Gambi, Martin Pickering, Kim Fuellenbach and E. Matthew Husband Brett Myers and Duane Watson
16:30pm-16:50pm		METER AND PHONOLOGICAL ENCODING DURING SPEECH PRODUCTION	

8 Poster Presentations at a Glance - Thursday Sept 06

Thursday, 6.9.2018

- 1 Perceptual Adaptation to Segmental Sounds in Non-native Speakers
Hiroki Fujita, Ruri Ueda and Ken-ichi Hashimoto (h.fujita@pgr.reading.ac.uk)
Status: Accept-Poster
- 2 Mapping Speech Segmentation to Linguistic Properties – What Counts?
Anna Mauranen and Alena Konina (alena.konina@gmail.com)
Status: Accept-Poster
- 3 Two Mechanisms of Scalar Implicature in Comparatively Modified Numerals
Christoph Hesse and Anton Benz (christoph.hesse@ed-alumni.net)
Status: Accept-Poster
- 4 Connectives as Processing Cues in Tracking Information Source: Evidence from Visual World Paradigm
Yipu Wei, Pim Mak, Jacqueline Evers-Vermeul and Ted Sanders (yipu.wei@outlook.com)
Status: Accept-Poster
- 5 Expressing Subjectivity in Discourse Relations: Evidence from Collocational Analyses
Yipu Wei, Jacqueline Evers-Vermeul and Ted Sanders (yipu.wei@outlook.com)
Status: Accept-Poster
- 6 Word Order Beats Morphosyntax: Incremental Thematic Role Assignment in Tagalog L1 Acquisition
Rowena Garcia, Jens Roeser and Barbara Höhle (rgarcia@uni-potsdam.de)
Status: Accept-Poster
- 7 The Ambiguity Disadvantage Effect in Word Comprehension: Rt and Eeg Evidence for Semantic Competition
Greg Maciejewski and Ekaterini Klepousiotou (psgm@leeds.ac.uk)
Status: Accept-Poster
- 8 Language Planning in Writing Resembles Planning in Speech
Jens Roeser, Mark Torrance and Thom Baguley (jens.roes@gmail.com)
Status: Accept-Poster
- 9 Expected Utility of Later Context Mediates Maintenance of Subcategorical Information
Wednesday Bushong and T. Florian Jaeger (wbushong@ur.rochester.edu)
Status: Accept-Poster
- 10 What Is “Good-Enough” about the Processing of Garden-Path Sentences in Czech
Jan Chromý (jan.chromy@ff.cuni.cz)
Status: Accept-Poster
- 11 Lexical Access on Behalf of Task Partner: Electrophysiological Insights from Joint Picture Naming
Anna Kuhlen and Rasha Abdel Rahman (anna.kuhlen@hu-berlin.de)
Status: Accept-Poster
- 12 Prior Learning of Acoustic Cues Blocks Learning of New Cues in Non-native Speech Acquisition
Jessie S. Nixon (jess.s.nixon@gmail.com)
Status: Accept-Poster
- 13 Novel Labels Increase Category Coherence, but Only in Coordinative Contexts
Ellise Suffill, Holly Branigan and Martin Pickering (esuffill@ed.ac.uk)
Status: Accept-Poster
- 14 The Reality of Hierarchical Morphological Structure in Multimorphemic Words
Yoosang Song, Youngah Do, Jongbong Lee, Arthur Thompson and Eileen Waegemaekers (song@cuhk.edu.hk)
Status: Accept-Poster
- 15 People with Smaller Social Networks Are Better at Talker Identification
Shiri Lev-Ari (shirilevari@gmail.com)
Status: Accept-Poster
- 16 Semantic Bias and Topicality in Pronoun Resolution
Yvonne Portele and Markus Bader (bader@em.uni-frankfurt.de)
Status: Accept-Poster
- 17 Features Matter in Computational Models of Word Reading
Stephan Tulkens, Dominiek Sandra and Walter Daelemans (stephan.tulkens@uantwerpen.be)

- Status: Accept-Poster
- 18 Focus Particle Position, and Accents, Affect Attachment
David Potter and Katy Carlson (k.carlson@moreheadstate.edu)
Status: Accept-Poster
- 19 Retention of Surface Information During L1 and L2 Reading: an Eye-tracking Study
Denisa Bordag, Andreas Opitz, Max Polter and Michael Meng (denisav@uni-leipzig.de)
Status: Accept-Poster
- 20 Being Nice Is Hard: Underspecification & Coercion in Copula Sentences
Anna Prysłowska (anna.pryslowska@uni-tuebingen.de)
Status: Accept-Poster
- 21 A Verbal Illusion without the Verb: Derailed Compositional Interpretation in Sentence Completions
Dario Paape and Shravan Vasishth (paape@uni-potsdam.de)
Status: Accept-Poster
- 22 Is Memory of Corrected Information Suppressed in Text Processing?
Scott McKenzie and Roger van Gompel (scott.mckenzie1@hotmail.com)
Status: Accept-Poster
- 23 Access of Demonstratives in Discourse: Evidence for Direct Access
Klaus von Heusinger and Andreas Brocher (abrocher@uni-koeln.de)
Status: Accept-Poster
- 24 A Neural Network Model of Adaptation in Reading
Marten van Schijndel and Tal Linzen (marten.vanschijndel@gmail.com)
Status: Accept-Poster
- 25 Accessibility and Affectedness of Bare Direct Objects
Elyesa Seidel (elyesa.seidel@uni-koeln.de)
Status: Accept-Poster
- 26 Is Quotation Articulated? Quotational Constructions and Their Acoustic Properties
Marcel Schlechtweg and Holden Härtl (marcelschlechtweg@gmail.com)
Status: Accept-Poster
- 27 Orofacial Expressions and Acoustic Cues in Whispered and Normal Speech
Marzena Zygis, Susanne Fuchs and Katarzyna Stoltmann (zygis@zas.gwz-berlin.de)
Status: Accept-Poster
- 28 Representation of Number in Agreement Comprehension
Yingzhao Zhou, Rhea T. Eskew, Jr. and Neal Pearlmutter (zhou.yingz@husky.neu.edu)
Status: Accept-Poster
- 29 Contribution of Grammatical Functions and Semantic Roles to Discourse Prominence in Turkish
Gökben Konuk and Klaus von Heusinger (goekben.konuk@googlemail.com)
Status: Accept-Poster
- 30 Linguistic Prediction and Visual Attention Do Not Interact in Reading
Suzanne Jongman, Marta De Pedis and Ashley Lewis (suzanne.jongman@mpi.nl)
Status: Accept-Poster
- 31 Resting State EEG Power Predicts Cognitive and Language Skills
Suzanne Jongman (suzanne.jongman@mpi.nl)
Status: Accept-Poster
- 32 Cross-linguistic Investigation of the Representations Driving the Choice between Pronouns and Nouns
Kumiko Fukumura, Coralie Herve, Sandra Villata, Francesca Foppolo and F-Xavier Alario
(kumiko.fukumura@stir.ac.uk)
Status: Accept-Poster
- 33 Normalizing Vowels at a Cocktail Party
Hans Rutger Bosker, Eva Reinisch and Matthias Sjerps (hansrutger.bosker@mpi.nl)
Status: Accept-Poster
- 34 The Effects of Context and Literality on L1 and L2 Idiom Processing: Evidence from Self-paced Reading
Sara D. Beck and Andrea Weber (sara.beck@uni-tuebingen.de)

- Status: Accept-Poster
- 35 Run-speaking? Simulations of Rate Control in Speech Production
Joe Rodd, Hans Rutger Bosker, Mirjam Ernestus, Antje Meyer and Louis ten Bosch (joe.rodd@mpi.nl)
Status: Accept-Poster
- 36 Are All Invalid Parafoveal Previews Created Equal?
Leigh Fernandez, Christoph Scheepers and Shanley Allen (leigh.fernandez@sowi.uni-kl.de)
Status: Accept-Poster
- 37 Morphological and Stereotypical Gender in Processing Agreement
Anastasiia Generalova and Natalia Slioussar (slioussar@gmail.com)
Status: Accept-Poster
- 38 Incremental Generation Drives "Efficient" Language Production
Spencer Caplan (spcaplan@sas.upenn.edu)
Status: Accept-Poster
- 39 Chinese Wh-in-situ and Islands: a Formal Judgement Study
Jiayi Lu, Cynthia K. Thompson and Masaya Yoshida (jiayilu2019@u.northwestern.edu)
Status: Accept-Poster
- 40 Auditory Statistical Learning Rapidly Supports the Processing of Larger Linguistic Chunks across Early Childhood.
Evan Kidd, Joanne Arciuli, Michael Smithson, Erin Isbilen and Morten Christiansen (evan.kidd@mpi.nl)
Status: Accept-Poster
- 41 Aspects of Semantics and Their Effects on Aphasic Picture Naming
Leonie Lampe, Solène Hameau, Nora Fieder and Lyndsey Nickels (leonie.lampe@students.mq.edu.au)
Status: Accept-Poster
- 42 Deliberative Process in Sharing Information with Different Audiences: Eye-tracking Correlates
Beatriz Martin Luengo, Yury Shtyrov and Andriy Myachykov (beatriz.martin.luengo@gmail.com)
Status: Accept-Poster
- 43 (Specifically) Language-impaired Processing of Relative Clauses in German
Flavia Adani, Yair Haendler, Romy Lassotta, Anne Adelt, Nicole Stadie and Frank Burchert (adani@uni-
Status: Accept-Poster
- 44 What Use for Wrong Guesses? Disconfirmed Predictions Boost Novel Word Learning
Chiara Gambi, Martin Pickering and Hugh Rabagliati (GambiC@cardiff.ac.uk)
Status: Accept-Poster
- 45 Prediction Errors Due to Tense Biases Do Not Affect Structural Priming
Roger van Gompel (r.p.g.vangompel@dundee.ac.uk)
Status: Accept-Poster
- 46 Does Noun Capitalization Help? an Eye-tracking Study on German
Margreet Vogelzang, Esther Ruigendijk, Tobias Mundhenk and Nanna Fuhrhop (margreet.vogelzang@uni-
oldenburg.de)
Status: Accept-Poster
- 47 Spoken Idioms Are Decomposed: Evidence from Erp and Eye Tracking
Ruth Kessler and Claudia K. Friedrich (ruth.kessler@uni-tuebingen.de)
Status: Accept-Poster
- 48 Eye-movement Control in the Visual World Paradigm
Anna Laurinavichyute and Anastasiya Lopukhina (annlaurin@gmail.com)
Status: Accept-Poster
- 49 Lexical Boost from the Matrix Verb
Laura Wakeford, Leila Kantola and Roger van Gompel (ljwakeford@dundee.ac.uk)
Status: Accept-Poster
- 50 Against Reactivation of Syntactic Traces in Filler-gap Dependencies in the Visual World Paradigm
Anna Laurinavichyute, Olga Dragoy, Mariya Khudyakova and Irina Sekerina (annlaurin@gmail.com)
Status: Accept-Poster
- 51 High Verb Frequency as an Accessibility Parameter Promoting Early Verb Placement in Main Clauses in Three
Semi-free Word Order Languages

- Gerard Kempen and Karin Harbusch (gerard.kempen@mpi.nl)
Status: Accept-Poster
- 52 About the Importance of Implicit Causality to Understand Relative Clause Processing
Céline Pozniak and Barbara Hemforth (celine.pozniak@gmail.com)
Status: Accept-Poster
- 53 Executive Function Adaptation and Syntactic Priming
Edith Kaan, Emma Leone and Yucheng Liu (kaan@ufl.edu)
Status: Accept-Poster
- 54 Active Antecedent Search in Cataphora Processing: Insights from Neural Oscillations
Ashley Lewis, Dave Kush, Andrew Jahn, Luca Campanelli, Clinton L. Johns and Julie Van Dyke (stashly@gmail.com)
Status: Accept-Poster
- 55 The Statistical Significance Filter Leads to Overoptimistic Expectations of Replicability
Shravan Vasishth, Daniela Merten, Lena Jäger and Andrew Gelman (vasishth.shravan@gmail.com)
Status: Accept-Poster
- 56 Integration and Anticipation Processes of the Speaker and Meaning in Adults with and without Autism Spectrum Disorder: Evidence from Eye-tracking and Erps
Mahsa Barzy, Jo Black, David Williams and Heather Ferguson (mm951@kent.ac.uk)
Status: Accept-Poster
- 57 Comprehending Information Structure by Korean English L2ers: an ERP Study
Wonil Chung and Myung Kwan Park (parkmk@dgu.edu)
Status: Accept-Poster
- 58 The Processing of Focus Alternatives: Evidence from Neuroimaging
Katharina Spalek and Yulia Oganian (katharina.spalek@hu-berlin.de)
Status: Accept-Poster
- 59 L2 Lexical Engagement of Grammatical Functions in Recently Learned Words: Insights from an Eye-tracking
Veronica Garcia-Castro and Leah Roberts (vgc505@york.ac.uk)
Status: Accept-Poster
- 60 The Role of Expectations in the Production Bias towards Short Dependencies
Idoia Ros, Adam Zawiszewski and Itziar Laka (idoia.ros@gmail.com)
Status: Accept-Poster
- 61 Effects of Cathodal Tdcs on Online Acquisition of Novel Word Forms
Daria Gnedych, Nadezhda Mkrtychian, Diana Kumakaeva, Evgenii Blagoveschenskii, Svetlana Kostromina and Yury Shtyrov (daria-gn@yandex.ru)
Status: Accept-Poster
- 62 Not All Islands Are Created Equal: Speeded Judgments in Spanish
Claudia Pañeda, Sol Lago and Claudia Felser (claudiapaneda@gmail.com)
Status: Accept-Poster
- 63 Effects of Syntactic and Semantic Structure on Production Planning
Monica Do and Elsi Kaiser (monicado@usc.edu)
Status: Accept-Poster
- 64 Parallelism Effects in Pronoun Dependency Resolution
Kathleen Hall and Masaya Yoshida (kathleenhall2018@u.northwestern.edu)
Status: Accept-Poster
- 65 Evaluating Prediction-by-production: Production-like Access to Orthographic and Phonological Forms of Predictable Words
Aine Ito (aineito@gmail.com)
Status: Accept-Poster
- 66 Cue Reliability and Adaptive Re-Weighting in Spoken Word Recognition
Wednesday Bushong and T. Florian Jaeger (wbushong@ur.rochester.edu)
Status: Accept-Poster
- 67 Korean Negative Polarity Items Meet an ERP Study
Myung Kwan Park, Wonil Chung and Sanghoun Song (parkmk@dgu.edu)

- Status: Accept-Poster
- 68 (Dis-)Confirmation of Linguistic Prediction by Non-linguistic Cues
Torsten Jachmann, Heiner Drenhaus, Maria Staudte and Matthew Crocker (jachmann@coli.uni-saarland.de)
Status: Accept-Poster
- 69 Are 'Great' and 'Not Great' the Same? Erp Evidence on the Processing of Irony and Sentence Negation
Stefanie Regel and Thomas C. Gunter (regel@cbs.mpg.de)
Status: Accept-Poster
- 70 Syntactic and Semantic Processing in Poor Comprehenders: Evidence from Eye-tracking and Computational Modeling
Luca Campanelli, Nicole Landi and Julie Van Dyke (campanelli.l@gmail.com)
Status: Accept-Poster
- 71 On Duration and Complexity: the Horse Raced Faster When Embedded
Nino Grillo, Miriam Aguilar, Leah Roberts, Andrea Santi and Giuseppina Turco (nino.grillo@york.ac.uk)
Status: Accept-Poster
- 72 Event-based Eye Blink Rate as an Index of Working Memory Gating and Updating: Predictive Pre-updating and Individual Differences in Working Memory Capacity
Tal Ness and Aya Meltzer-Asscher (talness@mail.tau.ac.il)
Status: Accept-Poster
- 73 Early Sensitivity to Number Agreement: What Pupillometry Reveals about L1 Acquisition of German
Assunta Süß, Tom Fritzsche, Petra Hendriks and Barbara Höhle (assuess@uni-potsdam.de)
Status: Accept-Poster
- 74 What Makes a House a Home? Mechanisms of Lexical Alignment in Preschoolers' Referential Communication
Laura Lindsay, Zoe Hopkins and Holly Branigan (s1036164@sms.ed.ac.uk)
Status: Accept-Poster
- 75 Online Processing of Case in Auditory and Written Korean Sentences as Revealed by Eye Movements in Native Speakers and L2 Learners
Cheryl Frenck-Mestre, Seung Kyung Kim, Hyeree Choo, Alain Ghio and Sungryong Koh (cheryl.frenck-mestre@univ-amu.fr)
Status: Accept-Poster
- 76 Adaptation to Variable Use of Expressions of Uncertainty
Sebastian Schuster and Judith Degen (sebschu@stanford.edu)
Status: Accept-Poster
- 77 Eye Movement Data and the Causes of Relative Clause Difficulties
Douglas Roland, Yuki Hirose and Gail Mauner (doug.roland@gmail.com)
Status: Accept-Poster
- 78 Approaching Scalar Diversity through (RSA with) Lexical Uncertainty
Chao Sun and Richard Breheny (uczlsun@ucl.ac.uk)
Status: Accept-Poster
- 79 Erps Differentiate between Type of Linguistic Information During Working Memory Maintenance of Sentences
Matteo Mascelloni, Roberto Zamparelli, Francesco Vespignani, Thomas Gruber and Jutta Mueller (jutta.mueller@uos.de)
Status: Accept-Poster
- 80 Verbal and Non-verbal Predictors of Word Comprehension and Word Production
Florian Hintz, Suzanne Jongman, James McQueen and Antje Meyer (florian.hintz@mpi.nl)
Status: Accept-Poster
- 81 The Effects of Visual Information in Bilingual Language Processing. Eye-tracking Study
Dato Abashidze, Pavel Trofimovich, Kim McDonough and Matthew Martin (dato.abashidze@concordia.ca)
Status: Accept-Poster
- 82 Taking It a Level Higher: the LEIA Model of Complex Word Recognition
João Veríssimo (joao.verissimo@uni-potsdam.de)
Status: Accept-Poster
- 83 Investigating Factors That Influence the Interpretation of Ambiguous Phrases as Literal or Sarcastic

- Ruth Filik, Christina Ralph-Nearman and Rachel Giora (ruth.filik@nottingham.ac.uk)
Status: Accept-Poster
- 84 Bilingual'S Referential Choice in Cognitively Demanding Situations
Carla Contemori and Iva Ivanova (ccontemori@utep.edu)
Status: Accept-Poster
- 85 On the Relationship between Lexical Processing Speed and Vocabulary in Toddlers
Seamus Donnelly and Evan Kidd (seamus.donnelly@anu.edu.au)
Status: Accept-Poster
- 86 The "Production P2" Effect Primarily Reflects Training in Picture Naming
Agata Wolna, Jakub Szewczyk, Patrycja Kałamała, Jonas Walther and Zofia Wodniecka
(jakub.szewczyk@gmail.com)
Status: Accept-Poster
- 87 Perceptual Priming and Syntactic Choice in English Language: Multimodal Study.
Mikhail Pokhoday, Yury Shtyrov, Christoph Scheepers and Andriy Myachykov (mikhail.pokhoday@gmail.com)
Status: Accept-Poster
- 88 Head Directionality Interacts with Dependency Length
Himanshu Yadav, Ashwini Vaidya and Samar Husain (yadavhimanshu059@gmail.com)
Status: Accept-Poster
- 89 The discourse behavior of weak definites in German
Frederike Weeber & Klaus von Heusinger (University of Cologne) fweeber1@uni-koeln.de
Status: Accept-Poster

9 Poster Presentations at a Glance - Friday Sept 07

Friday, 7.9.2018

- 1 (Event) Semantic Cues for Pronoun Realization
Semra Kizilkaya (semra.kizilkaya@uni-koeln.de)
Status: Accept-Poster
- 2 Wernicke Area Stimulation Differentially Affects Acquisition of Novel Concrete and Abstract Semantics
Diana Kurmakaeva, Nadezhda Mkrtchian, Daria Gnedykh, Evgenii Blagoveschenskii, Svetlana Kostromina and Yury Shtyrov (diana.s-pb@mail.ru)
Status: Accept-Poster
- 3 Age Differences in the Use of Syntactic and Semantic Associations During Sentence Processing
Caroline Beese, Markus Werkle-Bergner, Ulman Lindenberger, Angela D. Friederici and Lars Meyer (beese@cbs.mpg.de)
Status: Accept-Poster
- 4 Association of Speech Perception and Production in 2-Month-olds: Relating Even-related Brain Potential and Vocal Reactivity Measures
Gesä Schaadt, Angela D. Friederici, Hellmuth Obrig and Claudia Männel (schaadt@cbs.mpg.de)
Status: Accept-Poster
- 5 ERP Indices of Encoding Effects in Wh-Dependency Processing
Sergio López-Sancio and Ellen Lau (sergio.lopez-sancio@ehu.eus)
Status: Accept-Poster
- 6 Individual (Non-)Variability of Prosodic Cue Production in Coordinate Structures
Clara Huttenlauch, Carola de Beer, Sandra Hanne and Isabell Wartenburger (huttenlauch@uni-potsdam.de)
Status: Accept-Poster
- 7 German Demonstratives Are Sensitive to Perspective-taking
Stefan Hinterwimmer and Umesh Patil (stefan.hinterwimmer@uni-koeln.de)
Status: Accept-Poster
- 8 Can German Demonstrative Pronouns Really Be Bound?
Stefan Hinterwimmer and Umesh Patil (stefan.hinterwimmer@uni-koeln.de)
Status: Accept-Poster
- 9 Reference to Quantified Expressions in Swedish: an ERP Study
Fredrik Heinat and Eva Klingvall (evaklingvall@englund.lu.se)
Status: Accept-Poster
- 10 Antecedent Retrieval for Referential and 'Donkey' Pronouns
Dave Kush and Ragnhild Eik (dave.kush@ntnu.no)
Status: Accept-Poster
- 11 Children with SLI Can Use Number Agreement in Object-initial Sentences to Overcome Their Difficulties with Case Marking
Maja Stegenwallner-Schütz and Flavia Adani (stegenwa@uni-potsdam.de)
Status: Accept-Poster
- 12 Referential Overspecification: from Egocentricity to Rationality
Elli Tourtouri, Les Sikos and Matthew Crocker (elli@coli.uni-saarland.de)
Status: Accept-Poster
- 13 Language-induced Effects on Event Memory
Yaqi Wang, Gareth Gaskell and Silvia Gennari (yw1308@york.ac.uk)
Status: Accept-Poster
- 14 Semantic Attraction in Sentence Processing
Anna Laurinavichyute and Titus von der Malsburg (annlaurin@gmail.com)
Status: Accept-Poster
- 15 Knowledge of Academic Words Predicted by a Combination of Tests, Despite Cognate Inflation
Agnieszka Otwinowska-Kasztelanica and Breno Silva (brenotesol@gmail.com)
Status: Accept-Poster
- 16 Electrophysiological Correlates of Implicit and Explicit Acquisition of Novel Words

- Olga Scherbakova, Alexander Kirsanov, Elizaveta Nikiforova, Margarita Filippova, Evgeny Blagoveshchensky and Yury Shtyrov (o.scherbakova@gmail.com)
Status: Accept-Poster
- 17 Implicit Causality Biases and Thematic Roles in ASL Verbs
Anne Therese Frederiksen and Rachel I. Mayberry (atfreder@ucsd.edu)
Status: Accept-Poster
- 18 Animacy-driven Expectations in Norwegian Relative Clause Processing
Dave Kush and Ragnhild Eik (dave.kush@ntnu.no)
Status: Accept-Poster
- 19 Frequency Effects of Multiword Sequences in Spoken Mandarin Chinese
Ching Chu Sun, Peter Hendrix and Harald Baayen (peter.hendrix@gmail.com)
Status: Accept-Poster
- 20 Preservation of Phonological Contrast Does Not Block but Does Attenuate Phonetic Imitation
Václav Jonáš Podlipský, Šárka Šimáčková and Filip Smolík (vaclav.j.podlipsky@upol.cz)
Status: Accept-Poster
- 21 Coherence and Finiteness Effects in Extraction from Adjunct Islands in English
Christiane Müller, Damon Tutunjian and Anna-Lena Wiklund (christiane.muller@nordlund.lu.se)
Status: Accept-Poster
- 22 'Gedownloadet' or 'Downgeloadet'? Participle Formation for Multimorphemic English Loan Verbs in German
Britta Schulte and Ulrike Freywald (bschulte@uni-potsdam.de)
Status: Accept-Poster
- 23 Variation in French Partial Interrogatives: Social Meaning as a Key Factor to Understand Sociolinguistic Norm Violations
Gabriel Thiberge and Barbara Hemforth (gthg@tuta.io)
Status: Accept-Poster
- 24 Construal in Language: a Visual World Approach
Srdan Medimorec, Petar Milin and Dagmar Divjak (s.medimorec@sheffield.ac.uk)
Status: Accept-Poster
- 25 Learners Generate Unencountered Novel Structures Drawing on Knowledge of a Universal
Adam Morgan and Victor Ferreira (adam.milton.morgan@gmail.com)
Status: Accept-Poster
- 26 Cascading Activation: Evidence from Determiner Competition
Anja Riemenschneider, Tine Mooshammer and Katharina Spalek (anja.riemenschneider@iqb.hu-berlin.de)
Status: Accept-Poster
- 27 What Defines Grammatical Gender of Russian Expressive Nouns?
Natalia Chuprasova, Varvara Magomedova and Natalia Slioussar (slioussar@gmail.com)
Status: Accept-Poster
- 28 A Cross-linguistic Investigation of Response Time Distributions in Lexical Decision
Peter Hendrix (peter.hendrix@gmail.com)
Status: Accept-Poster
- 29 The Impact of Information Structure on Language Change: an Experimental Study
Shira Tal, Kenny Smith, Jennifer Culbertson, Eitan Grossman and Inbal Arnon (shira.tal1@mail.huji.ac.il)
Status: Accept-Poster
- 30 A Word or Two about Nonwords
Peter Hendrix (peter.hendrix@gmail.com)
Status: Accept-Poster
- 31 Learning from Speech Sounds' Probability Distributions Is Constrained by Prior Language Experience
Kateřina Chládková and Šárka Šimáčková (k.chladkova@uva.nl)
Status: Accept-Poster
- 32 Explicit Encoding and Fast Mapping of Novel Concrete Words: Behavioural Evidence of Equal Efficiency
Aleksandr Kirsanov, Olga Scherbakova, Evgeny Blagoveshchensky, Margarita Filippova, Nikiforova Elizaveta and Yury Shtyrov (isbraitwister@gmail.com)

- Status: Accept-Poster
- 33 L2 Speakers Are Not More Rational than L1 Speakers When It Comes to Loss Aversion
Zoe Schlueter, Chris Cummins and Antonella Sorace (zoe.schlueter@ed.ac.uk)
Status: Accept-Poster
- 34 Agreeing to Disagree: Agreement Attraction Effects in Resumptive Pronouns
Mor Ovadia, Maayan Keshev and Aya Meltzer-Asscher (maayankeshev@mail.tau.ac.il)
Status: Accept-Poster
- 35 Subject-verb Agreement Affects the Processing of a Subsequent Reflexive Pronoun
Maayan Keshev and Aya Meltzer-Asscher (maayankeshev@mail.tau.ac.il)
Status: Accept-Poster
- 36 Awareness of Linguistic Competence Influences Structural Priming
Christina Kim and Gloria Chamorro (c.s.kim@kent.ac.uk)
Status: Accept-Poster
- 37 Can L2 Speakers Acquire New Morphological Distinctions? Evidence from Temporal Morphological
Production in Mandarin Speakers of English
Qingyuan Gardner, Holly Branigan and Vicky Chondrogianni (s1360390@sms.ed.ac.uk)
Status: Accept-Poster
- 38 Sticky Labels: Formulation vs Reconceptualisation Effort in the Use of Referential Precedents
Lucia Castillo, Kenny Smith and Holly Branigan (s1354576@sms.ed.ac.uk)
Status: Accept-Poster
- 39 Watch Your Words: the Effects of Visual Manipulations in Event Scenes on Language Production
Yulia Esaulova, Sarah Dolscheid and Martina Penke (yulia.esaulova@uni-koeln.de)
Status: Accept-Poster
- 40 Stairs4Words: a New Adaptive Test for Assessing Receptive Vocabulary Size in English, Dutch and German
Florian Hintz, Suzanne Jongman, Marjolijn Dijkhuis, Vera van t' Hoff, Markus Damian, Sascha Schroeder,
Marc Brysbaert, James McQueen and Antje Meyer (florian.hintz@mpi.nl)
Status: Accept-Poster
- 41 Addressee Identity and Grammatical Processing: the Case of Basque Allocutive Agreement
Simona Mancini, Max Wolpert, Dana Scarinci and Sendy Caffarra (s.mancini@bcbl.eu)
Status: Accept-Poster
- 42 Individual Parsing Strategies in Complex Verb-final Structures. Evidence from Memory Interference
Katja Suckow and Jana Häussler (katja.suckow@gmail.com)
Status: Accept-Poster
- 43 The Validity of the Lexical Decision Task in Beginning Readers
Pauline Schroeter and Magdalena Bartelt (pauline.schroeter@iqb.hu-berlin.de)
Status: Accept-Poster
- 44 Co-speech Gestures and Co-present Objects: Attentional Allocation in Referential Communication
Raheleh Saryazdi and Craig Chambers (raheleh.saryazdi@mail.utoronto.ca)
Status: Accept-Poster
- 45 Factors Influencing Semantic Competition During Real-time Language Processing
Raheleh Saryazdi and Craig Chambers (raheleh.saryazdi@mail.utoronto.ca)
Status: Accept-Poster
- 46 German Children'S Processing of Non-canonical Word Order Sentences: the Role of Temporal Ambiguity
Chiara Boila, Tom Fritzsche and Barbara Höhle (chiara.boila@uni-potsdam.de)
Status: Accept-Poster
- 47 Lost in a Story, Detached from the Words: Absorbed Readers Are Less Sensitive to Word Characteristics
During Narrative Reading
Roel Willems, Lynn Eekhof, Moniek Kuijpers, Xin Gao, Emiel van den Hoven, Myrthe Faber and Marloes Mak
(roel.willems@donders.ru.nl)
Status: Accept-Poster
- 48 Erp Correlates of Semantic and Syntactic Processing in Pre-verbal Cochlear Implant Users.

- Luca Artesini, Francesco Pavani, Debora Musola, Giuseppe Nicolò Frau and Francesco Vespignani
(luca.artesini@unitn.it)
Status: Accept-Poster
- 49 The Interaction between Memory Retrieval and Expectations During Sentence Processing
Luca Campanelli, Julie Van Dyke and Klara Marton (campanelli.l@gmail.com)
Status: Accept-Poster
- 50 Did John Call Her Mother? the Role of Native-language Processing Strategies in Second-language Production Errors
Wing-Yee Chow, Esther Jesús Ortiz and Shasha Jin (wingyee.chow@ucl.ac.uk)
Status: Accept-Poster
- 51 Implicit Causality Affects the Choice of Anaphoric Form
Oliver Bott, Torgim Solstad and Anna Prysłowska (oliver.bott@uni-tuebingen.de)
Status: Accept-Poster
- 52 Effects of Referential Gaze in Spoken Language Comprehension: Human Speaker Gaze vs. Virtual Agent Listener Gaze
Eva Maria Nunnemann, Kirsten Bergmann, Helene Kreysa and Pia Knoeferle (enunnemann@techfak.uni-bielefeld.de)
Status: Accept-Poster
- 53 The Relationship between Reading Skill and Linguistic Prediction: an Investigation of Dyslexia
Norhafizah Mohd Taha, Michelle Yuen and Paul Engelhardt (p.engelhardt@uea.ac.uk)
Status: Accept-Poster
- 54 An Ear (and Eye) for Language: Predictors of Incidental and Explicit Foreign Language Vocabulary Learning
Marie-Josée Bisson, Anuenu Kukona and Angelos Lengeris (marie-josee.bisson@dmu.ac.uk)
Status: Accept-Poster
- 55 Why Two Is Not Always Better than One – an ERP Study on Minimality-based and Prosodic Predictions in German Discourse Processing
Petra Augurzky and Nadja Schaffler (petra.augurzky@uni-tuebingen.de)
Status: Accept-Poster
- 56 Rates of Scalar Inferences beyond ‘Some – a Corpus Study
Chao Sun, Ye Tian and Richard Breheny (uczlsun@ucl.ac.uk)
Status: Accept-Poster
- 57 Aspectual Make-up Reduces Relative Clause Avoidance
Miriam Aguilar and Nino Grillo (maguilar@fcs.unl.pt)
Status: Accept-Poster
- 58 Bilingual Priming of Pragmatic Enrichment
Lewis Bott and Susanne Muenz (bottla@cardiff.ac.uk)
Status: Accept-Poster
- 59 Planning until the End of the Sentence?
Agnieszka Konopka and Nele Ots (agnieszka.e.konopka@gmail.com)
Status: Accept-Poster
- 60 The Role of Alternative Constructions for Quantifier Scope Ambiguities: a Comparative Study
Barbara Hemforth and Lars Konieczny (barbara.hemforth@linguist.univ-paris-diderot.fr)
Status: Accept-Poster
- 61 Trajectories in Bilingual Production of Grammatical Gender Agreement: Language Experience and Cross-language Influence
Hamutal Kreiner, Eva Smolka and Tamar Degani (hamutalk@ruppin.ac.il)
Status: Accept-Poster
- 62 Statistical Learning in Infants, and Its Relationship with Language Development: a Study of Nonadjacent Dependency Learning.
Rebecca Frost, Caroline Rowland, Samantha Durrant, Michelle Peter, Amy Bidgood and Padraic Monaghan (rebecca.frost@mpi.nl)
Status: Accept-Poster

- 63 Complex Mandarin Motion Event Descriptions Are Not Serializations of Verbs: Evidence against the Equipollently-framed View
Ziwei Li, Qili Wang and Johannes Gerwien (gerwien@idf.uni-heidelberg.de)
Status: Accept-Poster
- 64 Processing Singular They with Generic and Specific Antecedents
Lauren Ackerman (lauren.ackerman@ncl.ac.uk)
Status: Accept-Poster
- 65 Morphological Generalization beyond Surface Similarity: Argument Structure and Inflectional Classes in Hebrew
Yael Farhy (farhy@uni-potsdam.de)
Status: Accept-Poster
- 66 Topicality Is Not a Prerequisite of Topic Drop: Evidence from a Rating Study on German
Lisa Schäfer, Robin Lemke and Ingo Reich (lisa.schaefer@uni-saarland.de)
Status: Accept-Poster
- 67 Discourse Accessibility of Semantically Integrated Referents
Eva Wittenberg (ewittenberg@ucsd.edu)
Status: Accept-Poster
- 68 Interference in the Processing of Grammatical Sentences: the Case of Multiple Negations
Iria de-Dios-Flores (iriadediosflores@gmail.com)
Status: Accept-Poster
- 69 Can Entropy Explain Successor Surprisal Effects in Reading?
Marten van Schijndel and Tal Linzen (marten.vanschijndel@gmail.com)
Status: Accept-Poster
- 70 Putting Prefixes in Front: Morphological Priming in L1 and L2 German
Laura Anna Ciaccio (ciaccio@uni-potsdam.de)
Status: Accept-Poster
- 71 Gender-biases in Language Processing: Explicit Beliefs about Event Outcomes vs. Implicit Linguistic Expectations
Titus von der Malsburg, Veronica Boyce, Till Poppels and Roger Levy (malsburg@posteo.de)
Status: Accept-Poster
- 72 Child-like Adults: Testing Distributivity Using a Dual Task
Anna de Koster, Jennifer Spenader and Petra Hendriks (a.m.b.de.koster@rug.nl)
Status: Accept-Poster
- 73 Agreement Attraction in German Sov Structures: an Erp Study
Robin Schäfer, Sol Lago and Titus von der Malsburg (rschaefer@uni-potsdam.de)
Status: Accept-Poster
- 74 Can Gapping Be Embedded? a Crosslinguistic Perspective
Gabriela Bilbie, Pegah Faghiri, Israel de la Fuente and Anne Abeillé (gabriela.bilbie@gmail.com)
Status: Accept-Poster
- 75 Evidence from Eye-tracking Shows Qualitatively Similar Processing of Novel Items by L1 and L2 Speakers
Anna Tsiola and Kiel Christianson (tsiola2@illinois.edu)
Status: Accept-Poster
- 76 The Effect of Event Depictions on Second Language Learning
Huong Thi Thu Nguyen, Katja Münster, Carsten Schlieve and Pia Knoeferle (nguyetxh@hu-berlin.de)
Status: Accept-Poster
- 77 The Role of Expectations in Referential Ambiguity Processing: Evidence from Russian
Veronika Prokopenya and Ekaterina Saenko (veronika.info@gmail.com)
Status: Accept-Poster
- 78 Structural Priming of the German Passive in Language Production
Yvonne Portele (y.portele@gmail.com)
Status: Accept-Poster
- 79 Beta-band Activity to the Rescue of Non-native Processing

- Laurent Dekydtspotter, Kate Miller, Mike Iverson, Yanyu Xiong, Kyle Swanson and Charlene Gilbert
(Idekydts@indiana.edu)
Status: Accept-Poster
- 80 False Positives in Growth Curve Analysis of Visual World Paradigm Data
Yujing Huang and Jesse Snedeker (yujinghuang@fas.harvard.edu)
Status: Accept-Poster
- 81 Investigating the Influence of Similarity Structure of Language Networks on Visual Word Recognition:
Insights from Megastudies
Cynthia Siew (cynsiewsq@gmail.com)
Status: Accept-Poster
- 82 Talker-specific Adaptation of Inferences Based on Scalar Adjectives
Cameron Morgan, Bethany Gardner, Rebecca Lawrence and Chigusa Kurumada
(ckuruma2@ur.rochester.edu)
Status: Accept-Poster
- 83 (Not) Forgetting Verbs in Hindi Doubly Center-embedded Structures
Samar Husain and Sakshi Bhatia (samar@hss.iitd.ac.in)
Status: Accept-Poster
- 84 Working Memory Constraints Override Prediction in Processing Hindi Center Embedded Constructions
Apurva Apurva and Samar Husain (apurva@hss.iitd.ac.in)
Status: Accept-Poster
- 85 Orthography in Second Language Word Learning and Pronunciation: Friend or Foe?
Pauline Welby, Audrey Bürki and Elsa Spinelli (buerki@uni-potsdam.de)
Status: Accept-Poster
- 86 Processing Inferred Result States in Discourse
Sarah Hye-yeon Lee and Elsi Kaiser (sarahl@usc.edu)
Status: Accept-Poster
- 87 The Role of Prediction Error in Linguistic Generalization and Item-based Learning
Masa Vujovic, Michael Ramscar and Elizabeth Wonnacott (masa.vujovic.15@ucl.ac.uk)
Status: Accept-Poster

10 Poster Presentations at a Glance - Saturday Sept 08

Saturday, 8.9.2018

- 1 The Role of Vowel Duration for Perceived Vowel Quality of Czech Vowels: Data from Native and Non-native Listeners
Nikola Paillereau and Radek Skarnitzl (nikola.paillereau@mac.com)
Status: Accept-Poster
- 2 Testing Implicit Learning with Case Marking Variation
Heeju Hwang, Jeong-Ah Shin, YooLae Kim and Bailiang Li (heejuhwang@gmail.com)
Status: Accept-Poster
- 3 Item-bound vs category-based generalization: an entropy model
Radulescu, Efi Giannopoulou, Sergey Avrutin and Frank Wijnen (S.Radulescu@uu.nl)
Status: Accept-Poster
- 4 Attention toward Shape and Color Is Affected by Linguistic Structures
Maximilian Paulus, Veronica Mazza, Giulia Calignano and Francesco Vespignani (francesco.vespignani@unitn.it)
Status: Accept-Poster
- 5 Independent versus Shared Syllable-representations in Late Spanish-german Bilinguals
Kiara Abad, Annett B. Jorschick and Joana Cholin (joana.cholin@uni-bielefeld.de)
Status: Accept-Poster
- 6 Calques from English Are Processed like Well-formed Collocations by Native Speakers of Polish: Evidence from N400
Marta Marecka, Agnieszka Otwinowska-Kasztelanic, Joanna Durlik, Jakub Szewczyk, Marcin Opacki and Zofia Wodniecka (marta.marecka@uj.edu.pl)
Status: Accept-Poster
- 7 An Ear for Language: Basic Auditory Skills Are Linked to More Efficient Novel Word Learning
Marta Marecka, Tim Fosker, Jakub Szewczyk, Patrycja Kałamała and Zofia Wodniecka (marta.marecka@uj.edu.pl)
Status: Accept-Poster
- 8 Parallel, Cascaded, Interactive Processing of Words During Sentence Reading: the Sentence-superiority N400 Effect
Yun Wen, Joshua Snell and Jonathan Grainger (yun.wen@univ-amu.fr)
Status: Accept-Poster
- 9 Case and (Mis)interpretation in Number Attraction: Evidence from Eastern Armenian
Serine Avetisyan, Sol Lago and Shravan Vasishth (serine.avetisyan@gmail.com)
Status: Accept-Poster
- 10 Understanding Changes in Garden-Paths as Expectation Adaptation
Wednesday Bushong, Zachary Burchill and T. Florian Jaeger (wbushong@ur.rochester.edu)
Status: Accept-Poster
- 11 Code-switching Patterns En Un Modelo Computacional: Simulating Code-switching in a Bilingual Sentence-production Model.
Chara Tsoukala, Stefan L. Frank, Mirjam Broersma and Antal van den Bosch (c.tsoukala@let.ru.nl)
Status: Accept-Poster
- 12 Is Pupillometry Sensitive to Island Violation Strength?
Christen Madsen II, Ian Phillips, Gita Martohardjono and Richard Schwartz (iphillips@gc.cuny.edu)
Status: Accept-Poster
- 13 This Is Not Only about Decomposition: L2 Learners Process Inflected Words Differently from Native Speakers
Kira Gor, Anna Chrabaszcz and Svetlana Cook (kiragor@umd.edu)
Status: Accept-Poster
- 14 Sensitivity to Language Statistics in 1St and 2Nd Language Reading
Stefan L. Frank and Robin Thompson (s.frank@let.ru.nl)
Status: Accept-Poster

- 15 A Crack in the Crystal Ball: Evidence against Pre-activation of Gender Features in Sentence Comprehension
Ernesto Guerra, Bruno Nicenboim and Andrea Helo (ernesto.guerra@ciae.uchile.cl)
Status: Accept-Poster
- 16 Semantic Interference and Morphological Facilitation in Noun-noun Compound Production: Evidence from Event-related Brain Potentials
Antje Lorenz, Stefanie Regel, Pienie Zwitserlood and Rasha Abdel Rahman (antje.lorenz@hu-berlin.de)
Status: Accept-Poster
- 17 Low-level Vocal Cues Affect the Acquisition of Hierarchical Structure
Antony Scott Trotter, Padraic Monaghan and Rebecca Frost (t.trotter@lancaster.ac.uk)
Status: Accept-Poster
- 18 Auditory-perceptual Gestalts Support the Processing of Phrase Structure in Comprehension
Antony Scott Trotter, Padraic Monaghan and Rebecca Frost (t.trotter@lancaster.ac.uk)
Status: Accept-Poster
- 19 Online Response to Perspective-taking in Narratives
Sara Meuser, Umesh Patil and Stefan Hinterwimmer (smeuser@uni-koeln.de)
Status: Accept-Poster
- 20 Evidence for Syntactic Transfer from Language to Music
Mythili Menon and Drew Colcher (mythilim@usc.edu)
Status: Accept-Poster
- 21 Placing Pronouns Rhythm Affects Word Order Preferences in German
Isabelle Franz, Markus Bader and Gerrit Kentner (isabelle.franz@ae.mpg.de)
Status: Accept-Poster
- 22 Processing in Parallel: Single-Trial EEG at the Phonology-Morphology Interface
Laurel Lawyer (l.lawyer@essex.ac.uk)
Status: Accept-Poster
- 23 Incremental Learning in Word Production: Tracing the Fate of Non-Selected Alternative Picture Names
Jörg D. Jescheniak, Franziska Kurtz, Herbert Schriefers and Andreas Mädebach (jdg@uni-leipzig.de)
Status: Accept-Poster
- 24 Incremental Strategies in Children'S Language Production
Jessica Brough, Holly Branigan, Chiara Gambi and Hugh Rabagliati (j.brough@ed.ac.uk)
Status: Accept-Poster
- 25 Lexical Prediction Does Not Interact with Morphophonological Prediction During Early Stages of Sentence Processing.
Mikel Santesteban, Paolo Lorusso, Anna Hatzidaki, Adam Zawiszewski and Itziar Laka (mikel.santesteban@ehu.eus)
Status: Accept-Poster
- 26 Prediction Overrides Syntactic Priming: Evidence from Hindi
Samar Husain and Himanshu Yadav (yadavhimanshu059@gmail.com)
Status: Accept-Poster
- 27 How Good Is Prediction in Head-final Languages?
Apurva Apurva and Samar Husain (apurva@hss.iitd.ac.in)
Status: Accept-Poster
- 28 Traces of Traces
Massimo Burattin, Francesca Foppolo and Carlo Cecchetto (francesca.foppolo@unimib.it)
Status: Accept-Poster
- 29 Is French Masculine Gender Overrated ? a Closer Look at Closest Conjunct Agreement
Aixiu An, Simon Duverger and Anne Abeille (aixiu0806@gmail.com)
Status: Accept-Poster
- 30 Only and Clefts: the Incremental Processing of Presupposition

- Jérémy Zehr, Francesca Foppolo, Daniele Scanzi and Florian Schwarz (francesca.foppolo@unimib.it)
Status: Accept-Poster
- 31 Do We Project Sluicing (and by Extension Ellipsis) Wherever Possible?
Emilia Molimpakis, Masaya Yoshida and Andrea Santi (emilia.molimpakis@ucl.ac.uk)
Status: Accept-Poster
- 32 The Impact of Stereotypes and Noun Endings on Processing Gender in English: Comparing Native and Non-Native Performance
Julia Müller, Lars Konieczny and Verena Haser (julia.mueller@pluto.uni-freiburg.de)
Status: Accept-Poster
- 33 Effects of Emotional Speaker Facial Expressions on Sentence Processing: an ERP Study
Katja Münster, Johanna Kißler and Pia Knoeferle (katja.muenster@hu-berlin.de)
Status: Accept-Poster
- 34 When Switching Language Is Cost-free
Michela Mosca, Chaya Manawamma and Kees de Bot (mosca@uni-potsdam.de)
Status: Accept-Poster
- 35 Relationship between the Processing of Semantically Anomalous Interpretation and Suppression Mechanism
Manabu Arai (manabu-arai@seijo.ac.jp)
Status: Accept-Poster
- 36 On-line Sensitivity to Tense and Tense/aspect Mismatches L1 and L2 English
Leah Roberts, Norbert Vanek and Josje Verhagen (leah.roberts@york.ac.uk)
Status: Accept-Poster
- 37 Distinguishing Competition from Task Demands: an ERP Study of Name Agreement in Timed Picture Naming
Evangelia Balatsou, Guillaume Thierry and Gary Oppenheim (e.balatsou@bangor.ac.uk)
Status: Accept-Poster
- 38 Is the Tendency to Lexically Entrain Stable across Time and Interlocutors?
Anita Tobar, Hugh Rabagliati and Holly Branigan (anita.tobar@ed.ac.uk)
Status: Accept-Poster
- 39 Avoiding Gaps in Romance
Francesca Foppolo, Ingrid Konrad, Massimo Burattin, Adrian Staub, Carlo Cecchetto and Caterina Donati (francesca.foppolo@unimib.it)
Status: Accept-Poster
- 40 Retrieval Errors as Common Source of Misinterpretations and Repetition Errors
Markus Bader and Michael Meng (bader@em.uni-frankfurt.de)
Status: Accept-Poster
- 41 Processing Correlates of Action Verb Specificity
Margit Scheibel (scheibel@phil-fak.uni-duesseldorf.de)
Status: Accept-Poster
- 42 Do Cross-linguistic Patterns of Morpheme Order Reflect a Cognitive Bias?
Carmen Saldana and Jennifer Culbertson (c.c.saldama@sms.ed.ac.uk)
Status: Accept-Poster
- 43 Bilingual – and Monolingual? – Language Control
Iva Ivanova (iva.m.ivanova@gmail.com)
Status: Accept-Poster
- 44 The Influence of the Animacy of Direct Objects in Brazilian Portuguese
João Vieira, Brenda Arruda de Souza and Elisângela Teixeira (elisteixeira@letras.ufc.br)
Status: Accept-Poster
- 45 Right-lateralization of Verbal Collocations
Shohini Bhattasali, Murielle Fabre and John Hale (sb2295@cornell.edu)
Status: Accept-Poster

- 46 It's Time to Prime Time! Structural Priming Shows Interrelation between Viewpoint Aspect and Event Structure
Monique Flecken and Johannes Gerwien (monique.flecken@mpi.nl)
Status: Accept-Poster
- 47 Anti-locality Effect without Verb-final Dependencies
Mingya Liu and Ming Xiang (lmingya@uos.de)
Status: Accept-Poster
- 48 Acquisition of Resultative Event Representations in Dutch: Does Describing Events Aid Memory of Event Culmination?
Ciara Hobbelink, Miguel Santín and Angeliek van Hout (miguel.santin@rug.nl)
Status: Accept-Poster
- 49 Discourse Inert? Implicit Objects Can Be as Accessible and Persistent as Overt Ones
Ana Besserman and Elsi Kaiser (pianibes@usc.edu)
Status: Accept-Poster
- 50 Processing Quantity Implicatures under QUDs
Eszter Ronai and Ming Xiang (ronai@uchicago.edu)
Status: Accept-Poster
- 51 Cross-varietal Lexical Alignment
Marie-Anne Morand and Constanze Vorweg (constanze.vorweg@csls.unibe.ch)
Status: Accept-Poster
- 52 First Language Processing of Compounds in Late Bilinguals
Serkan Uygun and Ayşe Gürel (serkanuygun2014@gmail.com)
Status: Accept-Poster
- 53 How Much Does Verb Semantics Determine Verb Syntax?
Mariela Jennings, Martha Palmer and Joshua Hartshorne (marielajennings@gmail.com)
Status: Accept-Poster
- 54 Pushkin: an Open-source Engine for Social Science at Scale
Mariela Jennings and Joshua Hartshorne (marielajennings@gmail.com)
Status: Accept-Poster
- 55 The Construction That the Reader Never Learns: ORCs and Adaptation
Caroline Andrews, Brian Dillon and Adrian Staub (ceandrews@linguist.umass.edu)
Status: Accept-Poster
- 56 Perceptual Uncertainty Effects and Referential Contrast
Adriana Baltaretu and Craig Chambers (craig.chambers@utoronto.ca)
Status: Accept-Poster
- 57 Learning Novel Morphosyntactic Features During Visual Action-events: Eye-tracking
Yang Gao, Dato Abashidze, Pavel Trofimovich and Kim McDonough (dato.abashidze@concordia.ca)
Status: Accept-Poster
- 58 The Role of Variability in Linguistic Generalization: Evidence from a Computerized Language Training Game with 7-Year-olds
Elizabeth Wonnacott, Masa Vujovic and Chantal Miller (masa.vujovic.15@ucl.ac.uk)
Status: Accept-Poster
- 59 Morphological Regularity and Processing Difficulty in an fMRI Study on Russian
Natalia Slioussar, Maxim Kireev, Alexander Korotkov and Svyatoslav Medvedev (slioussar@gmail.com)
Status: Accept-Poster
- 60 Phonemic Prediction at the Root for Arabic Words with Prefixes and Infixes
Samantha Wray and Alec Marantz (samantha.wray@nyu.edu)
Status: Accept-Poster
- 61 Gender Attraction in Modern Greek
Anastasia Paspali (paspali.anastasia@gmail.com)
Status: Accept-Poster

- 62 The Development of Idiom Knowledge across the Life Span
Simone Sprenger and Jacolien van Rij (s.a.sprenger@rug.nl)
Status: Accept-Poster
- 63 Mental State Verbs, Endorsement Readings, and Theory of Mind
Natalia Talmina and Kyle Rawlins (natalia.talmina@gmail.com)
Status: Accept-Poster
- 64 Dissecting Structural Priming: Differential Priming of Structural Features in Translation and Repeating
Robert M Maier (robert.maier@phil.uni-augsburg.de)
Status: Accept-Poster
- 65 Delayed Online Attachment for Parenthetical Relative Clauses
Marju Kaps, Alexandra Lawn and Jesse Harris (alawn@ucla.edu)
Status: Accept-Poster
- 66 Age of Acquisition Ratings Validated by Actual Vocabulary Scores
Shalom Zuckerman and Manuela Pinto (s.zuckerman@uu.nl)
Status: Accept-Poster
- 67 Wh-questions Are Understood before Polars
Sara Moradlou, Xiaobei Zheng, Ye Tian and Jonathan Ginzburg (sara.moradlou@gmail.com)
Status: Accept-Poster
- 68 Uniform Information Density Constrains Omissions in Fragments
Robin Lemke, Lisa Schäfer and Ingo Reich (robin.lemke@uni-saarland.de)
Status: Accept-Poster
- 69 Voice Mismatches in Vp Ellipsis Are Licensed by Syntactic Cues
Robin Lemke, Lisa Schäfer and Ingo Reich (robin.lemke@uni-saarland.de)
Status: Accept-Poster
- 70 Null Pronoun Is Always Better than Overt. Behavioral and Eye-tracking Evidence on Anaphora Resolution in Polish Language.
Agata Wolna, Joanna Durlik, Jakub Szewczyk, Michał Remiszewski and Zofia Wodniecka (agatawuwu@gmail.com)
Status: Accept-Poster
- 71 Learning Consistent Gender Errors in Non-native Speech
Thomas St. Pierre and Jean-Pierre Koenig (tastpier@buffalo.edu)
Status: Accept-Poster
- 72 Passive Sentence Difficulty: It'S Not about Argument Order but a State of Mind
Caterina Paolazzi, Claudia Cera, Nino Grillo, Artemis Alexiadou and Andrea Santi (a.santi@ucl.ac.uk)
Status: Accept-Poster
- 73 Neural Mechanism for Pronoun Resolution in Chinese During Naturalistic Listening
Jixing Li, Murielle Fabre, Wen-Ming Luh and John Hale (jl2939@cornell.edu)
Status: Accept-Poster
- 74 Underspecification in Relative Clause Attachment
Pavel Logacev and Noyan Dokudan (pavel.logacev@gmail.com)
Status: Accept-Poster
- 75 Development of a Language-independent System for Automatic Evaluation of L2 Oral Reproduction Tasks Using a Deep Learning Algorithm
Yutaka Yamauchi, Nobuaki Minematsu, Kayoko Ito, Megumi Nishikawa, Kay Husky and Aki Kunikoshi (yutaka@soka.ac.jp)
Status: Accept-Poster
- 76 Processing Dynamics of Subject-Verb Agreement: a Speed-Accuracy Tradeoff Analysis
Ryan King and McElree Brian (ryan.king@nyu.edu)
Status: Accept-Poster
- 77 Priming Ungrammatical Structures across Languages
Ian Phillips (iphillips@gc.cuny.edu)
Status: Accept-Poster

- 78 Wasted Power and False Interactions in Reading Time Analyses
Zachary Burchill, Wednesday Bushong and T. Florian Jaeger (zach.burchill@gmail.com)
Status: Accept-Poster
- 79 Disentangling Intra- and Inter-talker Variability in L2 Phonetic Production: L2 Speech, but Not Talkers, Is More Variable
Xin Xie, Ruolan Li and T. Florian Jaeger (xxie13@ur.rochester.edu)
Status: Accept-Poster
- 80 Modeling Prior Knowledge in the Perception of Native and Foreign-accented Speech
Xin Xie and T. Florian Jaeger (xxie13@ur.rochester.edu)
Status: Accept-Poster
- 81 Comparing Models of Unsupervised Adaptation During Speech Perception
Shaorong Yan and T. Florian Jaeger (syan13@ur.rochester.edu)
Status: Accept-Poster
- 82 Individual Differences in L2 Sentence Processing: Effects of Working Memory, Language Experience
Sven Hintzen, Elma Kerz, Daniel Wiechmann and Stella Neumann (sven.hintzen@rwth-aachen.de)
Status: Accept-Poster
- 83 English Resumptive Pronouns Do Not Help the Comprehender
Adam Morgan, Titus von der Malsburg, Victor Ferreira and Eva Wittenberg
(adam.milton.morgan@gmail.com)
Status: Accept-Poster
- 84 A Corpus of Native and Non-native Speech for Speech Production Research
Ruolan Li, Xin Xie and T. Florian Jaeger (xxie13@ur.rochester.edu)
Status: Accept-Poster
- 85 The handwriting of Chinese characters: a psycholinguistic data base
Ruiming Wang, Shuting Huang and Zhenguang Cai (zhenguangcai@gmail.com)
Status: Accept-Poster
- 86 Perceptual Priming and Syntactic Choice in English Language: Multimodal Study.
Mikhail Pokhoday, Yury Shtyrov, Christoph Scheepers and Andriy Myachykov
Status: Accept-Poster
- 87 Language-specific statistical computations in adults and infants
Luca Onnis, Erik Thiessen, Soo-Jong Hong and Kyung-Sook Lee (LucaO@ntu.edu.sg)
Status: Accept-Poster
- 88 ENHANCED IMPLICIT LEARNING IN BILINGUALS
Luca Onnis (LucaO@ntu.edu.sg)
Status: Accept-Poster

11 Oral Presentations - Abstracts

1

Keynote: *Probabilistic architectures and neurocomputational mechanisms for language processing*

Matthew W. Crocker

Department of Language Science & Technology
Saarland University, Germany.

The human language comprehension system maps the unfolding linguistic signal into rich meaning representations. Cognitive (computational) models of this system seek to account for how this is accomplished, in a manner that is both informed by - and also explains - online behavioural and neurophysiological measures that are sensitive to language. Salient indices include the relationship of a word's contextually-determined likelihood (or, *surprisal*) with that word's reading time (Hale, 2001; Levy, 2008), as well as the lexical, revision and integration processes indexed by event-related potentials such as the N400 and P600.

I will present a neurocomputational model based on the above assumptions and observations. I begin by motivating and introducing the *neural semantics* framework (Frank, Koppen, Noordman & Vonk, 2003; Venhuizen, Crocker & Brouwer, 2018), which not only supports the representation of complex logical forms, but also encodes their likelihood in the world, and supports probabilistic, knowledge-driven inference. Critically, these representations further support the computation of a 'meaning-centric' notion of surprisal, on a word-by-word basis. Using a simple recurrent network (Elman, 1990) trained to recover neural semantic representations for input utterances, I then demonstrate that the online surprisal of a word reflects both the *linguistic expectancy* of that word (as determined by training frequencies), and the *likelihood of the meaning* it induces (Venhuizen et al, 2018).

Next, I present our ongoing integration of neural semantic representations into the neurocomputational model of Brouwer, Crocker, Venhuizen & Hoeks (2017), which identifies a clear linking hypothesis to both the N400 (lexical retrieval) and P600 (semantic integration) ERP components. Key predictions of the model are that semantic integration difficulty should (a) result in increased surprisal, which should further (b) be manifest as an increased P600 amplitude, while (c) the N400 reflects contextually-driven retrieval processes. Findings from a recent ERP experiment that was conducted to test these predictions, against those of alternative accounts, are then presented (Delogu, Brouwer & Crocker, 2018). The findings confirm that the P600 component indexes general semantic integration and not, e.g., syntactic revision processes alone, while the N400 is predominantly driven by contextually-driven retrieval (i.e. priming), rather than integration processes. Taken together, the presented model and ERP findings offer a theoretically, computationally and empirically compelling account of both the processes involved in recovering utterance meaning, and their manifestation in behavioural (surprisal-driven reading times), and electrophysiological (N400 as retrieval, and P600 as integration/surprisal) measures.

Brouwer, H., Crocker, M., Venhuizen, N., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41, 1318-1352.
Delogu, F., Brouwer, H., & Crocker, M.W. (2018). The P600 - not the N400 - indexes semantic integration. *AMLaP Asia*.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211.
Frank, S. L., Koppen, M., Noordman, L. G., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, 27(6), 875-910.
Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings NAACL*, 1-8.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
Venhuizen, N., Crocker, M., & Brouwer, H. (2018): Expectation-based Comprehension: Modeling the Interaction of World Knowledge & Linguistic Experience, *Discourse Processes*.

¹Abstracts for the poster slams are listed in the poster section.

**RESISTANCE TO VARIABILITY FROM THE ENVIRONMENT IN LANGUAGE LEARNING:
CROSS-SITUATIONAL LEARNING OF WORDS FROM MULTIPLE CUES**

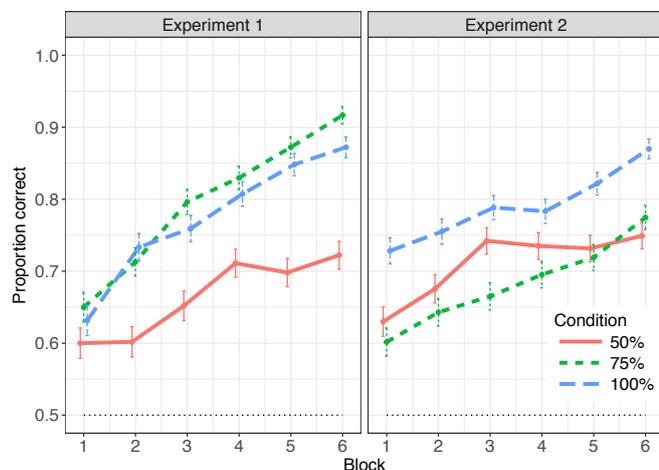
Padraic Monaghan (Lancaster University & MPI Psycholinguistics), James Brand (Lancaster University), & Rebecca Frost (MPI Psycholinguistics)
p.monaghan@lancaster.ac.uk

Studies of language acquisition have moved from considering the internal structure of language toward embedding language learning in rich, multimodal environments, involving consideration of both verbal and non-verbal communication, as well as interactions with objects and events that occur around the learner. This changed perspective extends our understanding of the sources of information potentially available to constrain acquisition, including distributions between words, prosody, and gesture accompanying speech. However, multiple environmental cues are also replete with noise and variability in their occurrence (Clerkin et al., 2017) which increases the computational cost of processing them, but also provides opportunities for learning. Dynamic systems theory predicts that noise in a computational system can facilitate learning, and the source of this advantageous noise can result from the environmental input to the system by increasing salience of individual cues, and reducing reliance on a single cue during learning (Monaghan, 2017). In two experiments, we tested the prediction that multiple cues are processed by the learner, and that variability in cues may even elicit an advantage for learning.

In Experiment 1, adults learned referents for 10 words on a cross-situational word learning task. In each trial, two spoken words and two objects were presented, and the task was to determine which object the speech referred to. Over multiple trials, one of the words and objects always co-occurred. No feedback was given. For the distributional cue, the referring word was preceded by a marker word. For the prosodic cue, the referring word was increased in amplitude. For the gestural cue, the target object was indicated by a finger point. Between subjects, we manipulated the variability of individual cues, whether they occurred in 100%, 75%, or 50% of trials. After training, participants' ability to identify the referent from the word with no cues present was tested. Results indicated that participants were sensitive to individual distributional, prosodic and gestural cues. Learning for 75% variability was quicker and 50% variability was slower than 100%, $p = .034$, $p < .001$, respectively (see Figure), indicating that variability in cue presence reduced reliance on particular cues and best supported learning.

In natural language learning, cues may vary in their presence (leading) or absence but they may also be (accidentally) misleading (e.g., pointing to an object whilst talking about another). In Experiment 2, we manipulated whether cues were present, absent, or misleading. The 100% condition was as in Experiment 1. In the 75% condition, in $\frac{3}{4}$ of trials the cue was present and leading, and for the remaining $\frac{1}{4}$ of trials the cue was misleading. In the 50% condition, in $\frac{1}{2}$ the trials, the cue was leading, in $\frac{1}{6}$ of trials, the cue was misleading, and in the remaining $\frac{1}{3}$ of trials the cue was absent. This time, misleading cues did not significantly affect pace of learning (100%:75%, $p = .448$, 100%:50%, $p = .079$). Overall accuracy was lower for 75% than 100%, $p = .009$, but not when variability was even higher (100%:50%, $p = .183$).

Variability in environmental cues can enhance word learning, and learning is robust to noise in the environment not only for the presence or absence of cues, but even when those cues are accidentally misleading to the learner.



References: Clerkin, E.M., Hart, E., Rehg, J.M., Yu, C., & Smith L.B. (2017). *Phil Trans Royal Soc London B*, 372. Monaghan, P. (2017). *Topics Cog Sci*, 9, 21-34.

IMPLICIT CONCURRENT LEARNING OF ADJACENT AND NONADJACENT DEPENDENCIES IN CHILDREN

Lai-Sang Iao, Jens Roeser, Lucy Justice & Gary Jones (Nottingham Trent University)
lai-sang.iao@ntu.ac.uk

Our environment is permeated with statistical regularities, occurring among adjacent elements (e.g., the syllable /pre/ is more likely to be followed by /ti/ than /on/) and nonadjacent elements (e.g., the morphosyntactic rule *is X-ing* where the intervened X is a verb). Importantly, both adjacent and nonadjacent dependencies occur simultaneously in language and other domains. Previous research has investigated how humans acquire these adjacent and nonadjacent dependencies (e.g., Newport & Aslin, 2004). Concurrent learning of both dependencies has recently been shown in adults in extended and multiple learning sessions using auditory stimuli (e.g., Creel, Newport, & Aslin, 2004). This study extends this line of research by testing the hypothesis that concurrent learning rapidly occurs in children without extensive exposure to visual stimuli.

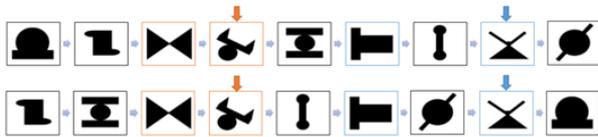


Fig. 1: Examples of adjacent (orange) and nonadjacent (blue) dependencies for training blocks 1–6.

Sixty-two children aged 10 to 11 years ($M = 131$ months, $SD = 3.3$) were tested first in a serial reaction time (SRT) task in which they were trained on materials comprising equally probable adjacent and nonadjacent dependencies by pressing keys that corresponded to each stimuli shown on screen as quickly and accurately as possible for 6–8 minutes. Fig. 1 illustrates two sequences of the SRT task: stimuli were shown one by one on screen in sequences over 6 blocks. Each sequence involved both adjacent and nonadjacent dependency. Examples of adjacent dependency are marked by orange boxes whereas examples of nonadjacent dependency are marked by blue boxes. These dependencies were violated in block 7 to test implicit learning. In a subsequent explicit judgement task, children were then required to discriminate between trained and untrained dependencies.

The results obtained from the SRT task (see Fig. 2) showed a speed-up from block 1 to 2 and subsequent blocks which indicates learning of key-image associations for adjacent/nonadjacent dependencies and other items that are not part of dependencies; slow-down from block 6 to 7 for adjacent and nonadjacent items but not the remaining items indicates that children learned and noticed the violation of learned dependencies (in block 7). Therefore, the reaction-time data show that children quickly developed sensitivity to both types of dependencies. This sensitivity was not found in the data from the explicit judgement task.

These findings suggest that implicit concurrent learning of both types of dependencies occurs rapidly in school-aged children. This shows that children rely on implicit mechanisms that allow them to simultaneously extract adjacent and nonadjacent regularities from their environment. However, children are not consciously aware of these regularities as suggested by the explicit judgement data.

While learning statistical structures from exposure may first be implicit, it might become explicit over time and development. These conclusions are consistent with Daltrozzo and Conway's (2014) statistical-sequential learning model and Cleeremans' (2006) model of unconscious cognition.

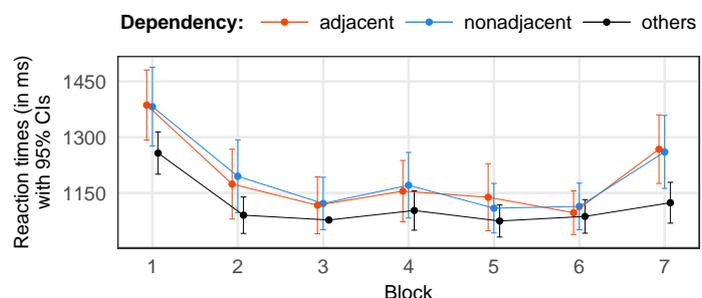


Fig. 2: Summary of SRT task data across blocks.

Investigating Implicit Learning and Surprisal Effects in a Structurally Biased Language over Development

Alina Kholodova ((University of Kaiserslautern), Michelle Peter (Univ. of Liverpool),
Caroline Rowland (MPI PsychoLinguistics) and Shanley Allen (University of
Kaiserslautern)

Syntactic priming is the tendency for an individual to unconsciously reproduce a previously experienced sentence structure, irrespective of lexical information (Bock 1986). This effect is often investigated using the Dative Alternation (DA), in which an individual can either use a prepositional object structure or a double object structure (*The girl gave the book to the boy(PO) / the boy the book(DO)*) to express the same message. Numerous priming studies on the DA in English and Dutch have provided valuable insights into the processing mechanisms of both children and adults, and have demonstrated the early emergence of abstract representations by the age of three years.

However, in English and Dutch the two dative structures are more or less balanced in use, and presumably also in strength of mental representations. Little research on the DA has been carried out in languages where one structural option is strongly preferred. Investigating priming effects in speakers of such structurally biased languages is important because they have considerably different mental strengths of the two syntactic representations. This is likely to have important implications for processing theories on priming.

In the present study, we investigate priming in German, a language in which the default structure is DO and where young children are hardly ever exposed to POs. We determined whether syntactic priming can enhance the dispreferred PO production in German-speaking children and adults. If so, this would support the *Implicit Learning account* (Bock & Griffin, 2000) according to which language input is processed via an *error-based learning* mechanism (Chang et al., 2000), and in which more surprising structures (i.e., less frequent structures like the German PO), are predicted to result in more priming (so-called *surprisal effects*; Jaeger & Snider, 2007).

To test this prediction, we adapted for German a video description priming task from Rowland et al.'s (2012) study with English-speaking children. We primed monolingual German-speaking children aged 3-4 years (N=42) and 5-6 years (N=31), as well as adults (N=37), with PO and DO structures (e.g., *Micky schickte den Fisch zu Minnie / Minnie den Fisch* 'Mickey sent the fish to Minnie / Minnie the fish'). The two structures also appeared either in the same verb (SV) or different verb (DV) condition in prime and target in order to test a secondary prediction that adults, but not children, would show enhanced priming effects under conditions of lexical overlap (i.e., lexical boost effect; Pickering & Branigan, 1998). Additionally, we incorporated a baseline condition containing intransitive (neutral) primes to assess baseline rates of PO/DO production without priming.

The results from a mixed design ANOVA with Post Hoc Tests revealed a significant 30% priming effect in 3- to 4-year-olds, a 13% priming effect in 5- to 6-year-olds, and an 18% priming effect in adults for PO production following a PO prime compared to a (baseline) intransitive prime. This suggests that a strong language-specific bias for the DO structure can be overridden by immediate exposure resulting in implicit learning for all groups. Moreover, the size of the priming effect in 3- to 4-year-old children (double that of the older children and adults) is predicted by the Implicit Learning account, and can be explained in terms of stronger surprisal effects due to weaker representations and little exposure to POs. Within the priming paradigm, we found a 13-16% increase in PO production after a PO prime compared to a DO prime in the DV condition, supporting abstract representations in all groups. However, when participants were primed in the SV condition, only adults showed a lexical boost effect (31%; $p < .001$). This finding is in line with Rowland et al. (2012) as well as Peter et al. (2015), but not with Morris and Scheepers (2015) and Branigan et al. (2016). According to our results, children unlike adults do not experience a facilitation in retrieving structural representations when lexical cues are involved. This is in line with the Implicit Learning Account's claim that the lexical boost stems from explicit memory retrieval, which is underdeveloped in children. In sum, our study shows the impact of structural infrequency on structural priming, providing more evidence for the existence of strong prime surprisal effects and error-based learning.

PREDICTABLE WORDS LEAVE PRODUCTION-LIKE TRACES IN MEMORY

Joost Rommers (Donders Institute, Nijmegen), Gary S. Dell (University of Illinois),
Aaron S. Benjamin (University of Illinois)
j.rommers@donders.ru.nl

Predictions about likely upcoming input may enable rapid language processing. However, the mechanisms by which predictions are generated remain unclear. One hypothesis is that comprehenders use their production system to covertly produce what they would say if they were the speaker (e.g., Chang, Dell, & Bock, 2006; Federmeier, 2007; Pickering & Garrod, 2007). The present study investigated whether predictable words leave production-like traces in memory. Two experiments capitalized on the *production effect*, which is the observation that words are remembered better by 10-25% when read aloud than when read silently (e.g., MacLeod et al., 2010). If reading predictable words involves covert production, the memory improvement from actually producing the words should be smaller for predictable words than for unpredictable words.

In Experiment 1, 70 participants read 96 sentences word-by-word in a 2x2 design (Predictability x Production). Half of the sentences were strongly constraining and ended in a predictable word (cloze probability 87%; "He swept the floor with a broom"), the other half were weakly constraining and ended in an unpredictable word (cloze 1%; "He noticed the man with the broom"). Participants read the final word aloud or silently, depending on the color of the words in the lead-in sentence (blue or red, counterbalanced between participants). Predictability shortened naming latencies by 53 ms. After a brief distraction task (solving math problems), participants performed a surprise memory task in which they saw the four types of previously read words, intermixed with new words (all words printed in green), and made recognition judgments on a four-point scale (Sure New, Maybe New, Maybe Old, Sure Old). As expected, production improved memory by 21%. Critically, signal detection theoretic analyses revealed an interaction such that the memory improvement from reading aloud was smaller for predictable words (discriminability: silent 0.43, aloud 0.89) than for unpredictable words (silent 0.45, aloud 1.07), $F = 5.025$, $p = .028$. This is consistent with the idea that reading predictable words can involve covert production.

Experiment 2 employed a source memory task in which participants identified which words they had read aloud and which ones they had read silently. If reading predictable words involves covert production, then for predictable words it should be relatively more difficult to remember whether they had been read aloud or silently; that is, production and predictability would be confusable in memory. Seventy new participants read 120 sentences according to the same procedure as in Experiment 1. Halfway through the sentence reading phase, the assignment of production condition (aloud/silent) to sentence color (blue/red) was switched around; thus, memory for the color of the sentence context was not by itself diagnostic of whether a word had been produced or read silently. Predictability shortened naming latencies by 49 ms. Overall, words read aloud received more "Sure Aloud" and "Maybe Aloud" judgments (58%) than did words read silently (24%), demonstrating reliable source memory. Critically, there was a small reduction in aloud/silent discriminability for predictable words (0.86) compared with unpredictable words (0.93), $F = 3.643$, $p = .061$, suggesting that predictability made it harder to tell the difference. In summary, predictable words seem to leave production-like traces in memory, which supports the hypothesized relationship between prediction and production.

Chang, F., Dell, G. S., & Bock, K. (2006). *Psychological Review*, 113(2), 234.

Federmeier, K. D. (2007). *Psychophysiology*, 44(4), 491-505.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 671.

Pickering, M. J., & Garrod, S. (2007). *Trends in Cognitive Sciences*, 11(3), 105-110.

DIFFICULTIES TRACKING ROLE-REFERENT SWITCHES CAN HELP TO EXPLAIN THE SUBJECT/OBJECT RELATIVE CLAUSE ASYMMETRY

Andrew Jessop (University of Liverpool; ESRC LuCiD Centre) and Franklin Chang (Kobe City University for Foreign Studies)

a.a.jessop@liverpool.ac.uk

Subject/object-extracted relative clauses (SRC/ORC) have different role-referent bindings that may contribute to the asymmetry in RC processing. We examined this using a multiple object tracking task where two separate push events occurred, followed by periods of random motion in a display of nine indistinguishable white circles (Figure 1). In the SRC-target conditions, the same circle was the agent in both pushes (double agent). In the ORC-target conditions, the patient of the first push was the agent of the second (patient-agent). After the random motion, four of the circles were given different colors and the participants described how these circles interacted using an RC: "the red that pushed blue pushed green" (SRC) or "the red that blue pushed pushed green" (ORC). Experiment one found that description accuracy was significantly higher for the SRC-target trials ($M = 74.39\%$) than the ORC-target trials ($M = 52.58\%$). This was replicated in experiment two (SRC-targets = 78%; ORC-targets = 55%), where the participants were instructed to use passive ORCs (e.g. "the red that was pushed by blue pushed green"), as these can be easier in English (Gennari & MacDonald, 2008). In the third study, participants described the same stimuli with active transitive sentences (e.g., "red pushed blue") and we found significantly higher accuracy in the SRC ($M = 76.83\%$) than ORC-target trials ($M = 68.17\%$). Since the SRC bias was observed without animacy differences and even when non-RC structures were used, the results suggest that tracking the patient-agent switch for the ORC main clause subject imposed a cost that created a subject/object RC asymmetry in production. We tested this theory in three additional experiments where the participants produced transitive sentences to describe scenes in which the target agent (e.g., "red") or patient (e.g., "blue") played the same role in both pushes (double agent/double patient) or switched to the alternative role between the two events (agent-patient/patient-agent). These studies found significantly higher description accuracy for scenes where the thematic role of the targets was consistent. This was observed for active sentences with an overlapping agent (exp. 4: double agent = 87.92%, agent-patient = 70.00%) or an overlapping patient (exp. 5: double patient = 88.75%, patient-agent = 83.33%), as well as passive sentences with an overlapping patient (exp. 6: double patient = 82.50%, patient-agent = 73.54%). Collectively, the present findings suggest that thematic role consistency contributes to RC processing. The increased difficulty in tracking the role switches in ORC-type events could weaken the link between meaning and ORC forms (Fitz, Chang, and Christiansen 2011). We suggest that these biases in production may shape the frequency/structural biases implicated in the RC asymmetry in comprehension (MacDonald, 2013).

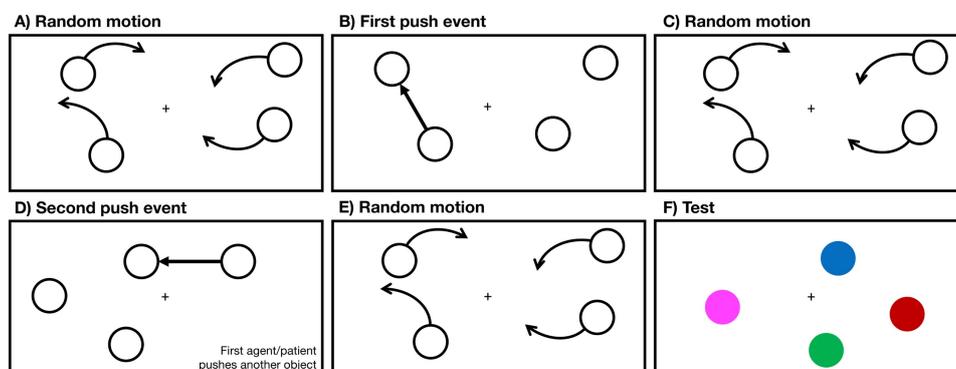


Figure 1. A diagram of the multiple push tracking task (the actual stimuli contained 9 circles). If the pusher in B is also the pusher in D, then it is a double agent. If the target pushed in B becomes the pusher in D, then it is a patient-agent. In exp. 1-2, two of the circles highlighted at test (F) appeared in the same color, which created an ambiguity that required a relative clause to disambiguate.

LONG-TERM EFFECTS OF STRUCTURAL REPETITION ON SENTENCE PLANNING?

Zsofia Stefan (University of Aberdeen) & Agnieszka Konopka (University of Aberdeen)
zsofia.stefan@gmail.com

Speakers tend to reuse recently heard or produced sentence structures, even across multiple unrelated sentences (i.e., over shorter and longer time intervals^[1]). Structural priming is a form of implicit learning whereby experience with a given structure results in long-lasting changes in structure-building procedures that facilitate re-selection of that structure. At short intervals, structural priming also influences the time-course of message and sentence planning^[2]: when using a primed structure, speakers begin message and sentence planning by encoding a conceptual framework for the entire sentence (hierarchically incremental planning) rather than planning one concept/word at a time (linear planning). The present study is the first to test whether priming effects at the level of sentence planning are as robust and long-lasting as the repetition of structure. If they are, then the benefits of implicit learning of structure may include persistent changes in the way speakers map preverbal information onto language. But if repetition of structure persists without an accompanying change in planning, then the expansion of planning scope due to priming may be only a short-lived practice effect.

Experiment: Forty-eight eye-tracked native English speakers described a series of pictures (40 prime-target pairs and 200 intransitive fillers). On prime trials, participants heard and repeated active or passive recorded descriptions (e.g. *The cat is catching the mouse* or *The mouse is being caught by the cat*). Target events were presented immediately after the primes (no-lag condition) or after two intervening fillers (lag condition). Target events uniformly contained actional verbs (e.g. “catch”, “hold” as opposed to non-actional such as “remember”, “see” etc), mostly animate agents (37/40 items), and half animate and half inanimate patients.

Structure choice: Speakers produced more active sentences after active than passive primes, but priming was stronger in the *no-lag* than *lag* condition (7% vs 2%; $z = -2.16$, $p < .05$). Structure choice also showed the typical effects of patient animacy^[3]: speakers produced more passive sentences to describe events with animate than inanimate patients ($z = -6.59$, $p < .001$). Character animacy did not interact with prime structure.

Sentence planning (active sentences): Speakers showed a preference for the agent over the patient in the 0-400 ms time window (early conceptual and linguistic planning), and then fixated the agent with priority between 400 ms and speech onset (primarily linguistic encoding of the agent). Importantly, in the *no-lag condition*, speakers showed a smaller early preference for the agent (0-400 ms) after hearing active than passive primes ($t = -1.79$, $p = .07$). This indicates a shift away from early linear planning and towards hierarchical planning when a sentence structure is easy to assemble^[2]. After 400 ms, speakers also directed their gaze to the agent more quickly after active than passive primes, showing strong top-down guidance from a conceptual framework during linguistic encoding of the agent. In contrast, there were no changes in eye movements due to priming in the *lag condition*. Speakers fixated the agent rapidly and preferentially before 400 ms following both active and passive primes, showing clear linear planning. They also fixated the agent preferentially after 400 ms, but the distribution of fixations was more flat than in the *no-lag condition*: this suggests less support from a sentence framework in deploying attention to the sentence-initial character. To assess item sensitivity to priming, we also compared planning in structurally flexible events (20 events with animate patients eliciting both actives and passives) and nonflexible events (20 events with inanimate patients eliciting mainly actives). In flexible events, there was a strong shift towards hierarchical planning in the no-lag condition and a similar but weaker effect in the lag condition, confirming that the experiment had sufficient power to detect persistent shifts in planning strategies. Structurally nonflexible events did not show priming-related changes in planning strategies.

Conclusion: Structural priming was observed at the level of *structure choice* in both lag conditions, but priming effects at the level of *sentence planning* were only observed in the no-lag condition in structurally flexible items. Thus, changes in sentence planning are more likely to be influenced by recent experience with a given structure than to reflect long-lasting changes in structural representations or in message-to-language mapping strategies.

[1] Pickering & Ferreira, 2008; [2] Konopka & Meyer, 2014; [3] Ferreira, 1994.

LINEAR VS. STRUCTURAL INCREMENTALITY IN THE FACE OF SENTENCE PRODUCTION IN CONTEXT

Xiaogang Wu & Johannes Gerwien (Heidelberg University)
xiaogang.wu@hotmail.com

In three picture description experiments with German native speakers, we manipulated the discourse status of the agent and patient in transitive scenes by adding a mini lead-in discourse before Ss were to produce the target responses (Prat-Sala & Branigan 2000), as well as the sequence of participant's visual attention directed to the event referents by means of an attention capture technique (Gleitman et al. 2007). Our aim was to determine (1) whether the discourse status of the referents as a factor for syntactic choices outranks thematic role information and visual salience; (2) how contextual information is mapped from the conceptual representation to linear positions of a sentence; and (3) whether sentence production in context is a linear or structural incremental formulation process.

In three experiments, Ss first saw an introduction picture that showed two animate referents which appeared again as the participants of a transitive scene in a subsequent target picture. While being exposed to the introduction picture, Ss heard a mini discourse in which the referents' names were introduced by a first sentence ("In this picture, you see a king and a soldier."), and a specific response about the target picture was evoked by a second sentence. In Exp 1 (N=16), the second sentence was „Please tell me what happens in the next picture“. In Exp 2 and 3 (N=32/32), the format of the sentence was „In the next picture you will see the [Referent 1] again, please tell me what happens to [Referent 2]“. In half of the trials, [Referent 1] was the agent in the target picture and [Referent 2] was the patient, in the other half of the trials this was reversed. This manipulation allowed us to determine whether the agent or the patient in the target pictures received the role of the discourse topic. After the introduction phase, Ss saw a fixation cross (500 ms) and a blank screen (200 ms) and then a visual cue (80 ms). In Exp 1, the cue appeared at the position of the following agent or patient. In Exp 2, the cue always appeared at the position of the following patient. In Exp 3, the cue always appeared at the position of the following agent. The cue was immediately followed by the target picture. Filler items followed the general format of the critical items, but with a task of judging the relative positions of the two figures in the target pictures. Ss were instructed to respond as quickly as possible. Eye movements were recorded.

Ss were unaware of the visual cue in all three experiments. In Exp. 1, the analysis of first fixation locations indicated that without the manipulation of discourse status, Ss' visual attention was drawn to the cued entity in more than 69% of all trials. However, regardless of the first fixation location, Ss directed their second saccade, which can be considered the first 'voluntary' saccade, in more than 85% of all cases to the agent. Furthermore, almost 100% of the produced sentences mentioned the agent as the first referent. Thus, the manipulation of visual attention did not predict syntactic choices in German participants, which is not in line with previous findings by Gleitman et al. (2007). With the introduction of topicality in Exp. 2 and 3, we found a modulating effect of discourse status on visual cueing. Ss' visual attention was drawn to the cued entity in more than 79% of all trials when this entity carried topic status, whereas the proportion dropped to 37% when the cued entity was not topical. With respect to syntactic choices, we found that topical referents, regardless of their semantic role, always appeared in the sentences' initial position. Moreover, we found that producing an active sentence (with agent as topic) was significantly faster than producing a passive sentence (with patient as topic). We conclude that (1) information structure specifications influence sentence production from early on by directing processing attention to topical entities, (2) they outrank thematic role information, as well as the visual salience of referents, and (3) they are mapped onto the semantic level before grammatical encoding (function assignment/linearization) begins. We expand the notion of structural incrementality to the level of information structure, showing that information structure, together with the semantic structure influences syntactic encoding. These results indicate that German sentence production in context is mainly a structural incremental formulation process.

EXPECTATION MANAGEMENT IN ONLINE DIALOGUE COMPREHENSION: AN ERP INVESTIGATION OF DUTCH *INDERDAAD* ‘INDEED’ AND *EIGENLIJK* ‘ACTUALLY’

Marlou Rasenberg (Radboud University/MPI), Joost Rommers (Donders Institute for Brain, Cognition and Behaviour) & Geertje van Bergen (Max Planck Institute for Psycholinguistics) geertje.vanbergen@mpi.nl

In conversational interaction, language provides us with explicit cues to acknowledge our interlocutor’s expectations. For instance, the Dutch discourse markers (DMs) *inderdaad* ‘indeed’ and *eigenlijk* ‘actually’ encode (mis)alignment between what is said and what is expected on the basis of the (extra)linguistic context. The question we address here is to what extent such expectation-managing cues affect online dialogue comprehension. More specifically, we investigate how their presence affects processing of subsequent linguistic input. If language comprehenders immediately use DMs to affect incremental processing, it is hypothesized that *inderdaad* enhances, and *eigenlijk* reduces predictability effects on subsequent words, as measured by modulations of the N400 and post-N400 positivities.

In two ERP reading experiments, Dutch participants (N=40 per experiment) read 216 short dialogues, consisting of an introductory context sentence followed by a question-answer pair (see Table). We manipulated (a) whether a critical word (CW) in the answer confirmed or disconfirmed a likely prediction (measured by cloze probabilities), and (b) whether the CW in the answer followed an adverb (e.g., *yesterday*, *very*) or a DM (*inderdaad* in Exp. 1, *eigenlijk* in Exp. 2).

Context	Question	Answer	Adverb/DM		CW	
<i>Ondanks haar angst voor dieren is Irene naar het circus geweest.</i>	Jan vraagt: <i>Je vond de slotact vast doodeng?</i>	Irene zegt: <i>Ik schrok</i>	<i>ontzettend nderdaad eigenlijk</i>	<i>van de rennende</i>	<i>olifant clown</i>	<i>aan het eind.</i>
Despite her fear for animals, Irene went to the circus.	Jan asks: You must have been terrified by the final act?	Irene says: I was scared	<i>very indeed actually</i>	<i>by the running</i>	<i>elephant clown</i>	<i>at the end.</i>

Analyses of ERPs time-locked to the CW show that predictable CWs elicited reduced N400 amplitudes compared with unpredictable CWs (300-500 ms after stimulus onset), replicating findings from numerous previous experiments and extending these to conversational contexts. This Predictability effect was not significantly modulated by the preceding DM. Strikingly, the Predictability effect on CWs differed significantly between experiments, such that N400 amplitude for unpredictable CWs was reduced in the *inderdaad*-experiment relative to the *eigenlijk*-experiment. The presence of *inderdaad* (signaling confirmation of one’s discourse expectations) in the experimental context may have encouraged the system to operate in “top-down verification mode”, at the expense of thoroughly processing the bottom-up input (e.g. van Berkum, 2010).

In the 500-800 ms time window, we found that unpredictable CWs elicited more positive amplitudes than predictable CWs over anterior channels, previously linked to additional processing when lexical predictions are disconfirmed (e.g., Federmeier, 2007; Van Petten & Luka 2012; DeLong et al., 2014). Again, preceding DMs did not significantly modulate this Predictability effect. Moreover, there was a main effect of DM that interacted with Experiment: CWs following *eigenlijk* elicited more positive amplitudes over posterior channels than CWs following an adverb, irrespective of Predictability; no such effect was found for *inderdaad*. We take this finding to show that encountering *eigenlijk* (signaling an upcoming conflict with one’s expectations) increases integration costs of subsequent input with the (pragmatically more complex) discourse.

Despite a lack of evidence that expectation-managing DMs modulate semantic pre-activation of subsequent words online, our findings show that *inderdaad* and *eigenlijk* differentially affect incremental integration of subsequent linguistic input with the wider discourse model. Moreover, our findings suggest that expectation-managing DMs modulate the utility of prediction in the experimental context. As such, DMs provide useful tools to further investigate the nature of predictions in incremental language comprehension.

PARTICIPANT ASSIGNMENT TO THEMATIC ROLES IN TZELTAL: EYE-TRACKING EVIDENCE FROM SENTENCE COMPREHENSION IN A VERB-INITIAL LANGUAGE

Gabriela Garrido Rodriguez (Max Planck Institute for Psycholinguistics), Elisabeth Norcliffe (Max Planck Institute for Psycholinguistics), Falk Huettig (Max Planck Institute for Psycholinguistics), Penelope Brown (Max Planck Institute for Psycholinguistics) & Stephen C. Levinson (Max Planck Institute for Psycholinguistics)
gabriela.garrido@mpi.nl

Studies using eye-movements to investigate predictive processing during sentence comprehension have mostly focused on subject-initial languages¹⁻² and have found that thematic role knowledge is used to restrict the set of possible interpretations quickly³⁻⁵. In subject-initial languages, it is not clear what type of information comprehenders actually use to anticipate upcoming event participants because one argument (often the Agent) is always available from the outset, but usually needs the verb to be interpreted semantically. Verb-initial languages provide the opportunity to investigate the early interpretation of events and the type of information that can drive anticipatory processing.

In a visual world eye tracking experiment on Tzeltal (Mayan), we investigated how information provided by verbs is used to predict thematic role assignments. Basic word order in Tzeltal is Verb-Patient-Agent (actives) or Verb-Agent-Patient (passives). Thus, the verb is always encountered first, making argument structure and syntactic information available at the outset. By hypothesis, this could be used to anticipate the post-verbal arguments. Thus, Tzeltal allows us to test whether anticipatory eye movements to Agents and Patients are driven by, 1) voice marking and word order (active: VPA or passive: VAP), or 2) if listeners follow a (potentially universal) Agent preference⁶⁻⁷.

Ninety-two Tzeltal speakers listened to verb-initial sentences such as "La **xchuk/sjoyin** bel woje jelek' te jchukawale" (*The policeman **arrested/accompanied** the thief yesterday*) and "**Chuk-ot/Joyin-ot** woje yu'un jchukawal te jelek'e" (*The thief **was arrested/was accompanied** by the policeman yesterday*) while seeing a display showing two potential referents (e.g., policeman, thief) and two distractors (e.g., girl, nun). We manipulated verb type (predictive: verb has typical agents/patients, e.g. "arrest"; non-predictive: verb does not select typical agents/patients, e.g. "accompany") and voice marking (active vs. passive).

We found differences in anticipatory processing between active and passive sentences in the predictive verb condition. After hearing a sentence-initial passive verb, listeners were more likely to first fixate the Agent after verb offset and then turned their gaze to the sentence-final Patient before it was mentioned. In contrast, after hearing predictive active verbs, participants did not show increased anticipatory looks to the argument immediately following the verb (the Patient), but having heard both the verb and the Patient, they turned their gaze to the sentence-final Agent before it was mentioned.

These results suggest that, 1) Tzeltal listeners use voice marking and verbal semantics to anticipate the Agent referent in passive sentences and 2) there is no sufficient evidence for an early Agent fixation preference during sentence comprehension in Tzeltal, challenging the notion of a universal Agent preference⁷. Our results provide cross-linguistic evidence of on-line incremental thematic role assignment during sentence comprehension in a verb-initial language and the use of passives as a linguistic resource to disambiguate sentences with two third-person arguments in Tzeltal⁸.

References

¹Kamide, Altmann, Haywood. (J Mem Lang, 2003;49).²Kamide, Scheepers, Altmann (J Psycholinguist Research, 2003;32).³Altmann (J Mem Lang, 1999;41).⁴Knoeferle, Crocker, Scheepers, Pickering. (Cognition, 2005;95).⁵Kukona, Fang, Aicher, Chen, Magnuson. (Cognition, 2011;119).⁶Bickel, Witzlack-Makaverich, Choudhary, Schlesewsky, Bornkessel-Schlesewsky. (PLOS ONE, 2015;10).⁷Sauppe. (Front. Psychol, 2016;7).⁸Polian. (CIESAS, 2013).

ASSESSING THE EFFECT OF ADDRESSEE'S LINGUISTIC COMMUNITY ON SPEAKERS' LEXICAL CHOICES

Anita Tobar (University of Edinburgh), Hugh Rabagliati (University of Edinburgh), Holly Branigan (University of Edinburgh)
anita.tobar@ed.ac.uk

Speakers tend to imitate each other's lexical choices to refer to objects that have several names (*umbrella vs broolly*) (lexical entrainment). Lexical entrainment has been described as a consequence of recently processed labels being more accessible than others (Pickering & Garrod, 2004), but also as a behaviour in response to the addressee's needs (Clark & Brennan, 1991) and to the speaker's social affiliations (communication accommodation theory, CAT) (Giles, 2012). Addressees' needs and speakers' social affiliations are important when interacting with partners from a different linguistic community, for they may not use the same labels as the speaker, e.g., while the European Spanish (ES) word for potato is *patata*, some speakers of Latin American Spanish (LS) would use the label *papa*. An interesting question is whether speakers generalise from one partner's lexical choices to the lexical preferences of that partner's community and, if so, whether this tendency is mediated by the speakers' tendency to lexically entrain with that partner.

We investigated the effect of addressee's linguistic community on speakers' tendency to (1) entrain with a partner and (2) reuse their lexical choices across partners. 143 **native speakers of ES** engaged in two sessions of an online picture-matching and -naming task. Participants and a 'partner' (in fact, a computer programme) alternated turns selecting and naming a target. Experimental items comprised a target that could be named with both a favoured and a disfavoured label (*papa vs patata*). The 'partner' named the experimental targets using the disfavoured names exclusively. Participants named the experimental targets after the 'partner' in Session 1 and before the 'partner' in Session 2. We manipulated participants' beliefs about their addressee's linguistic community so, in each session, they believed they played with a native speaker from either their own linguistic community (ES) or another one (LS). We manipulated Session 1 partner's community (ES vs LS) and Session 2 partner's community (ES vs LS), thus generating 4 conditions: ES-ES, ES-LS, LS-ES, LS-LS; between-participants. We measured (1) whether participants used the same word as their partner in Session 1 (lexical entrainment); and (2) whether they reused their choice from Session 1 when interacting with their new partner in Session 2 (reuse of lexical choice).

In Session 1, participants' entrained **48%** of the time, and their use of disfavoured labels was above chance (**V=0, p<.0001**). There was no effect of partner's community on participants' tendency to entrain (**LS: 50%, ES: 42%, p>.05**). In Session 2, participants were more likely to change their lexical choice if they had entrained in Session 1 (E) than if they had not (NE) (**E: 32%, NE: 16%, p<.0001**). After entraining, participants in LS-LS were more likely to reuse their lexical choice in Session 2 relative to participants in LS-ES (**LS-LS: 70%, LS-ES: 53%, p=.01**). There was no difference between ES-ES and ES-LS (**ES-ES: 74%, ES-LS: 78%, p>.05**). Our results do not suggest that the tendency to entrain is affected by partner's community. However, they do suggest speakers generalise from one partner's lexical choices to the preferences of that partner's community and this tendency is mediated by the speakers' tendency to entrain to that partner. Although, this occurred only when the 'partner' did not belong to the speaker's community: participants reused their lexical choices to the same extent with LS and ES partners after entraining with an ES partner but not after entraining with a LS partner. For this, we offer an explanation in terms of CAT. In Session 1, participants assumed the unusual word choice to be acceptable in the partner's community and kept using it as long as it did not pose a social cost. If the disfavoured label was associated with LS, changing the lexical choice with an ES speaker was convenient for two potential reasons: (1) decreasing social distance with other in-group members or (2) avoiding social rejection from a partner due to the use of lexical choices from a non-standard linguistic variety (i.e. LS). If the disfavoured label was associated with ES, participants could reuse it with both ES and LS partners without any potential social cost. We are currently conducting further studies to validate these hypotheses.

HIERARCHICAL STRUCTURE PRIMING FROM MATHEMATICAL EQUATIONS TO TWO- AND THREE-SITE RELATIVE CLAUSE ATTACHMENTS IN RUSSIAN

Andriy Myachykov (Northumbria University, Higher School of Economics), Christoph Scheepers (University of Glasgow), Anastasia Galkina (Higher School of Economics) & Yury Shtyrov (Aarhus University, Higher School of Economics)
andriy.myachykov@northumbria.ac.uk

A number of recent studies demonstrated cross-domain structural priming effects between mathematics, music, and language, suggesting the existence of shared structural representations across different cognitive domains. For instance, Scheepers et al. (2011) showed that English speakers' choices of relative clause (RC) attachments in partial sentences like *The tourist guide mentioned the bells of the church that ...* can be influenced by the structure of previously solved prime equations such as $80 - (9 + 1) \times 5$ (making *high* RC-attachments more likely) versus $80 - 9 + 1 \times 5$ (making *low* RC-attachments more likely). While this finding had been successfully replicated (e.g., Scheepers & Sturt, 2014; Van de Cavey & Hartsuiker, 2016), its generalizability, as well as the exact nature of the underlying structural representations, remain unclear. Experiment 1 (see Figure 1) of the present study is a close replication of Scheepers et al. (2011) in Russian, a morphologically rich language. More interestingly, Experiment 2 extended this finding to more complex *three-site* attachment configurations like *I always liked the jam of the grandma of my friend that...* and found that, relative to a structurally neutral baseline prime condition, N1-, N2-, and N3-attachments of RCs in Russian were equally susceptible to structural priming from mathematical equations such as $18 + (7 + (3 + 11)) \times 2$ [EQ1] versus $18 + 7 + (3 + 11) \times 2$ [EQ2] versus $18 + 7 + 3 + 11 \times 2$ [EQ3], respectively (see Figure 2). Taken together, our findings suggest (1) that cross-domain structural priming from mathematical equations to RC-attachment generalizes to languages other than English, and more importantly (2) that such cross-domain structural priming effects must rely on domain-general hierarchical structure representations whose level of detail goes beyond mere 'local vs. non-local' integration of constituents.

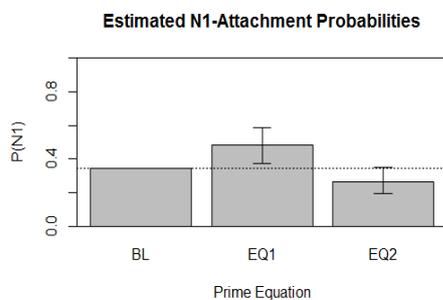


Figure 1 (above). Probabilities of N1-attachments in the target trials of Experiment 1 broken down by levels of Prime Equation Type: BL (e.g., $5 + 12$), EQ1 (e.g., $80 - (9 + 1) \times 5$), EQ2 (e.g., $80 - 9 + 1 \times 5$).

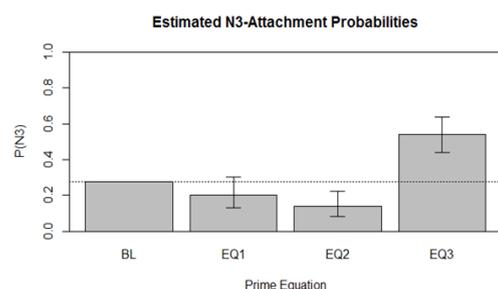
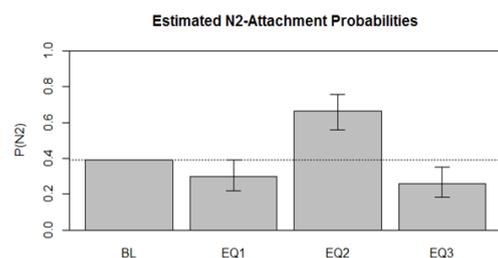
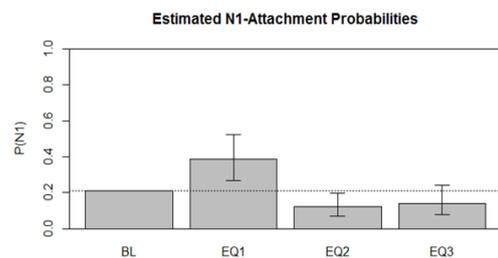


Figure 2 (right). Probabilities of N1-attachments (top panel), N2-attachments (middle panel), and N3-attachments (bottom panel) in the target trials of Experiment 2 broken down by levels of Prime Equation Type: BL (e.g., $5 + 12$), EQ1 (e.g., $18 + (7 + (3 + 11)) \times 2$), EQ2 (e.g., $18 + 7 + (3 + 11) \times 2$), and EQ3 (e.g., $18 + 7 + 3 + 11 \times 2$).

Keynote: Beyond language for visual objects

Asifa Majid

Center for Language Studies Radboud University Nijmegen
Research Fellow, Donders Institute for Brain, Cognition and Behaviour
The Netherlands

The field of psycholinguistics has made significant discoveries about the architecture and mechanisms of language processing by investigating language in relation to the visual world. In comparison, we know relatively little about how language interfaces with other perceptual modalities. Moving beyond vision is critical, however, in order to understand the broader cognitive architecture that supports language use, and that enabled the evolution of language in the first place. In this talk, I explore the coding of the perceptual senses of hearing, touch, taste, and smell - alongside vision - by describing experimental and corpus-based studies of a diverse sample of the world's languages. Overall, the data suggest that there are asymmetries in language use across the senses. Across the globe, vision is talked about far more frequently than the other senses; followed closely by audition. Moreover, while visual and auditory language appears to be 'embodied' (i.e., word meanings involving these senses recruit representations and processes involved in perceiving the real-world referent also); olfactory language appears not to be simulated in the same way. These data suggest there are cognitive architectural constraints on how language interfaces with perception. At the same time, the experimental data also show considerable cultural variation in the codability of the senses (how easy it is to express a particular sensory quality or experience), suggesting that a single language snapshot purporting to reflect the Faculty of Language seriously underestimates both its representational and processing capacity.

Keynote: Event knowledge and semantic processing

Ken McRae

Department of Psychology, and Brain & Mind Institute
University of Western Ontario
London, Ontario, Canada

People constantly use concepts and word meaning to recognize entities and objects in their environment, to anticipate how entities will behave and interact with one another, to know how objects should be used, and to understand language. Over the years, a number of theories have been presented regarding how concepts are organized and structured in semantic memory. For example, various theories stress that concepts (or lexical items) are linked by undifferentiated associations. Other theories stress hierarchical categorical (taxonomic) structure, whereas others focus on conceptual similarity spaces. In this talk, I will present evidence that people's knowledge of real-world events and situations is an important factor underlying the structure and (contextually-determined) usage of concepts in semantic memory. I will present experiments spanning word, picture, and sentence processing. Evidence for the importance of event-based knowledge will cover a number of types of concepts, including verbs, nouns denoting living and nonliving things, and abstract concepts. I conclude that semantic memory is structured in the mind so that the computation and use of knowledge of real-world events and situations is both rapid and fundamental. In other words, event knowledge is an important force that shapes the dynamics of real-time, context-sensitive, semantic computations.

CONTRASTING FACILITATION PROFILES FOR AGREEMENT AND REFLEXIVES REVISITED: A LARGE-SCALE EMPIRICAL EVALUATION OF THE CUE-BASED RETRIEVAL MODEL

Lena A. Jäger (University of Potsdam), Daniela Merten (University of Potsdam), Julie Van Dyke (Haskins Laboratories), & Shravan Vasishth (University of Potsdam)

lena.jaeger@uni-potsdam.de

Even though both (1a) and (1b) are ungrammatical, reading studies consistently report faster reading times at the auxiliary verb *were* in (1b) vs. (1a).

1a. *The bodybuilder who worked with the personal trainer were competitive.

1b. *The bodybuilder who worked with the personal trainers were competitive.

The Lewis & Vasishth 2005 cue-based retrieval model of sentence processing (LV05) [1] correctly predicts faster reading times at the auxiliary verb *were* in (1b) vs. (1a): The plural marking on the auxiliary verb *were* triggers a retrieval of a plural-marked subject, but occasionally the plural-marked distractor noun *trainers* is misretrieved in (1b) vs. (1a). The LV05 model predicts a facilitation effect of approximately -26 ms; constrained variation of parameters can lead to mean predicted effects ranging from -10 ms to -57 ms (see Fig. 2 in [4]).

In an eyetracking study, Dillon et al. [2] showed that faster total reading times are seen at the auxiliary in (1b) vs. (1a), as predicted by the LV05 model (see figure below). A reanalysis of [2]’s data using a maximal Bayesian linear mixed model shows that the estimated mean facilitation in their data is -60 ms, with a 95% probability that the facilitation effect lies between -112 and -5 ms (this is the so-called 95% credible interval).

Interestingly, Dillon and colleagues also showed that a similar configuration, antecedent-reflexive dependencies, for which the LV05 model predicts similar facilitation effects as for subject-verb agreement, shows no facilitation effects at all at the reflexive *themselves*: our Bayesian linear mixed model reanalysis showed a mean total reading time of -18 ms, 95% credible interval $[-72, 36]$.

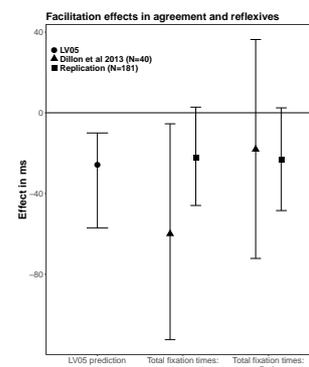
2a. *The bodybuilder who worked with the personal trainer injured themselves.

2b. *The bodybuilder who worked with the personal trainers injured themselves.

Dillon et al. argue that reflexives are immune to misretrieval effects because binding theory’s Principle A acts as a filter, allowing misretrieval-free and deterministic access to the antecedent. However, [2] had a relatively small sample size ($N=40$). Using the largest LV05-predicted effect size (-57 ms) and the standard error estimate from [2], the probability of detecting an effect correctly with 40 participants is 30%. When power is this low, many null results will be found and any statistically significant estimate (e.g., the facilitation in (1b) vs. (1a)) will *always* be exaggerated [3,5]. This is because the standard error is so large that under repeated sampling, the effect estimates will fluctuate, hence any estimate close to the true mean will not cross the significance threshold [3]. Crucially, both the agreement and reflexive effects in [2] have such wide uncertainty intervals that the LV05 model’s predictions are fully compatible with them (see figure).

Accurate estimates with narrower credible intervals can only be obtained with larger sample studies [3,5]. We therefore conducted a direct replication of [2]’s eyetracking study, but with a larger participant sample size ($N=181$); for a predicted effect of -57 ms, power is now 88% (we also had grammatical controls in the experiment, as in [2], but these are not discussed here due to space constraints). Here, both the agreement (1a,b) and reflexives (2a,b) show similar facilitation effect estimates in total reading times, closer to the magnitude predicted by the LV05 model: agreement: -22 ms $[-46, 3]$; reflexives -23 ms $[-48, 3]$.

In sum, both agreement and reflexive dependencies seem to show similar facilitation profiles, consistent with the predictions of the LV05 model. More generally, this work demonstrates the importance of conducting larger-sample studies in order to obtain more precise estimates for evaluating predictions of quantitative models. **References:** [1] Lewis and Vasishth, 2005, Cog Sci. [2] Dillon et al, 2013, JML. [3] Gelman and Carlin, 2014, PPS. [4] Engelmann et al. 2018, <https://osf.io/b56qv/> [5] Vasishth et al, 2018, <https://osf.io/p9baz/>



HOW THE INPUT SHAPES THE ACQUISITION OF INFLECTIONAL MORPHOLOGY: COMPUTATIONAL MODELLING ACROSS THREE HIGHLY INFLECTED LANGUAGES

Felix Engemann (University of Manchester), Sonia Granlund (University of Liverpool), Joanna Kolak (University of Manchester), Virve Vihman (University of Tartu), Ben Ambridge (University of Liverpool), Julian Pine (University of Liverpool), Anna Theakston (University of Manchester) & Elena Lieven (University of Manchester)
felix.engemann@manchester.ac.uk

In order to account for children's error patterns in the acquisition of complex inflection paradigms, usage-based (e.g., Bybee, 1995, *Lang and Cogn Proc*, 10) and even some traditionally rule-based accounts of morphology acquisition (e.g., Pinker & Ullman, 2002, *Trends in Cogn Sci*, 6:11) incorporate input-based learning mechanisms, such as **rote-storage** and **phonological analogy**. According to usage-based accounts, a learner stores highly frequent words as complete forms in memory and produces low frequent inflections by analogy with phonologically similar forms. Artificial neural networks have famously been used to argue against the need for symbolic rules because of their incremental learning with graded sensitivity to input similarities, while applying a single, integrated mechanism to learn regular and irregular forms (see McClelland & Patterson, 2002, *Trends in Cogn Sci*, 6:11). But to what extent can children's acquisition of complex morphology be explained by a similar mechanism? The majority of previous experimental and computational investigations has focused on simple systems like English or investigated only a small part of a paradigm. We conducted a large-scale investigation of the acquisition of the noun and verb inflection paradigms of three morphologically rich languages by combining experiments and neural network modelling.

Method We carried out elicited-production studies with children between the ages of three and five on singular noun case marking in Polish, Finnish and Estonian, and present-tense verb person/number marking in Polish and Finnish. The results were compared with simulations with three-layer feed-forward networks that were trained on natural, child-directed speech data. The input to the models consisted of phonemes representing nominative noun forms (noun models) or verb stems (verb models) and a code for the target case or person/number context, respectively. The models were trained to output the correct phoneme representation of the target form. Inputs were presented probabilistically according to their token frequencies in child-directed speech corpora.

Results All models acquired mastery of the system after maximally three million training trials (presenting one form per trial) and could generalise (i.e., produce the correct target for untrained items) to 78–88% (except PL nouns with 60%) of the test items used in the experiments. The key phenomena predicted by usage-based theories were observed in both the experiments and the simulations: effects of token frequency and phonological neighbourhood density (inflectional class size) of the target form, and a general error pattern that involved the replacement of low-frequency targets by higher-frequency forms of the same lemma, or forms with the correct case or person/number, but with a suffix from an inappropriate inflection class. The models furthermore showed an interaction that is predicted theoretically but wasn't found experimentally, namely that the effect of phonological neighbourhood was smaller for items of higher token frequency, suggesting that analogy is used mainly for low frequent forms. Finally, hierarchical clustering of the models' internal representations revealed that lemmas were grouped on the basis of phonological similarities that included items from different inflection classes. Errors could therefore be better predicted when defining phonological neighbourhood with a computational measure of similarity instead of a class-based one.

Our findings demonstrate that acquisition of highly complex systems of inflectional morphology can be accounted for by rote storage and phonological analogy, as opposed to formal symbolic rules. The fact that this process can be modelled theoretically by a simple feed-forward network speaks in favour of a single, incremental and domain-independent learning mechanism that operates on graded phonological similarities in the input.

**N400 AMPLITUDES REFLECT CHANGE IN A PROBABILISTIC REPRESENTATION OF MEANING, EVEN IF INDUCED BY DIFFERENCES IN WORD FORMS:
A NEURAL NETWORK MODEL**

Milena Rabovsky (Freie Universität Berlin, Germany)

milena.rabovsky@gmail.com

The N400 component of the event-related brain potential (ERP) is the most widely used component in research on language and meaning, but its functional basis remains incompletely understood. We present a computationally explicit account of the N400's functional basis, simulating N400s as the stimulus-induced change in a probabilistic representation of meaning. Specifically, we use the Sentence Gestalt (SG) model (St. John & McClelland, 1990), which uses incoming words as cues to constrain the activation state at a hidden layer (called the SG layer), which implicitly and probabilistically represents all aspects of meaning of the event described by a sentence. The model's N400 correlate is the magnitude of change in SG layer activation induced by the new word, i.e. $\text{Model N400} = |\text{SG}_n - \text{SG}_{n-1}|$. Using this approach, we simulate sixteen empirical N400 effects (Rabovsky et al., accepted). Here, we focus on a new simulation of a finding, which recently triggered a debate on probabilistic prediction in language comprehension, due to a partially failed replication attempt (Nieuwland et al., 2018; but see Yan et al., 2017). Specifically, DeLong et al. (2005) exploited the fact that English indefinite articles are adjusted to fit the words they are preceding such that "an" is used prior to vowels while "a" is used prior to consonants. In sentences such as "The day was breezy so the boy went outside to fly...", N400s were smaller to "a" (compatible with the high cloze continuation "kite") compared to "an" (to be followed by a lower cloze continuation). Because "a" and "an" do not differ in meaning, this N400 effect was taken to indicate probabilistic pre-activation of word forms. To simulate these results, we trained the model on an artificial environment where specific sentence beginnings such as e.g., "At breakfast, the woman eats..." continue to describe certain events with high probability, e.g., "... an egg", and other events with low probability, e.g., "...a mango". Even though the articles themselves are not associated with any specific meaning in the model environment (they just reliably precede specific continuations), the model's N400 correlate was smaller at the article in high probability as compared to low probability sentences ($p < .01$), in line with the original findings. The simulation suggests that even though the statistics of natural language may not always support the effects because of unreliable relationships between articles and nouns (e.g. "...an old kite"), the effects should be observable in principle when the articles constitute reliable cues. Furthermore, the simulation demonstrates that even if the effects are observed, they do not necessarily indicate pre-activation at the level of word form, but instead may reflect a change in a predictive representation of meaning, which is cued by the encountered word forms.

References

- DeLong, K.A., Urbach, T.P., & Kutas, M (2005). *Nature Neuroscience*, 8(8), 117-121.
Nieuwland, M. S., et al. (2018). *eLife*, 7: e33468.
Rabovsky, M., Hansen, S.S., & McClelland, J.L. (accepted). *Nature Human Behaviour*. Preprint on bioRxiv doi: 10.1101/138149.
St. John, M.F. & McClelland, J.L. (1990). *Artificial Intelligence*, 46, 217-257.
Yan, S., Jaeger, T. F., & Kuperberg, G. R. (2017). Preprint on bioRxiv doi: 10.1101/143750.

**THE MIRRORING EFFECTS OF INFORMATION THEORY BASED
AND DISCRIMINATION LEARNING BASED QUANTIFICATIONS
OF INFLECTED ADJECTIVES**

Dušica Filipović Đurđević (Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia) & Petar Milin (Department of Journalism Studies, University of Sheffield, Sheffield, United Kingdom)
dusica.djurdjevic@f.bg.ac.rs

A growing body of research has demonstrated that information theory based descriptions of language are successful at predicting processing latencies, eye-fixations, as well as neural responses to various linguistic stimuli. It has been argued that measures derived from information theory reflected immense sensitivity of the cognitive system to probabilistic features of the environment in general, and language in particular. In this research we aim to further investigate this sensitivity by relating information theory based quantifications to the fundamental process of discrimination learning introduced by Rescorla and Wagner (1972).

On information theory part, we focused on the effect of paradigmatic relative entropy (Kullback-Leibler divergence) between the frequency distribution of the inflectional paradigm (all inflected forms of the given word) and the frequency distribution of the inflectional class (all words that inflect in the same way). Simultaneously, we investigated syntagmatic relative entropy derived from prepositional phrases. Relative entropy, which reflects the degree of dissimilarity of the two frequency distributions, has been shown to inhibit processing (Baayen, Milin, Filipović Đurđević, Hendrix, & Marreli, 2011; Filipović Đurđević & Gatarić, in press; Hendrix & Baayen, 2014; Kuperman et al., 2010; Linzen, Marantz, & Pytkänen, 2013; Milin, Filipović Đurđević, & Moscoso del Prado Martin, 2009).

In parallel, based on naïve discrimination learning, we trained a simple, two-layer network mapping trigraph cues from the orthographic input to the output layer consisting of pointers to locations in the semantic space (Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017). During the course of mapping cues to outcomes, the network was learning to discriminate cues that are good predictors of the given outcome from the unpredictable cues and preserving this information in the values of the association weights. Based on these weights, we derived several discrimination learning based quantifications that were previously attested as good predictors of processing cost.

In visual lexical decision task native speakers were presented with inflected forms of Serbian adjectives for which we derived the two sets of quantifications (lexical-distributional and information theory based predictors on the one hand, and discrimination learning based predictors on the other hand).

In two separate analyses we fitted the predictors derived from information theory framework, and the predictors derived from discrimination learning framework to processing latencies. Both analyses revealed a complex pattern of effects. Within information theory based set, there was a three-way interaction which revealed a fine interplay of inhibitory effects of both paradigmatic and syntagmatic relative entropy, which was further modulated by lemma frequency. Similarly, significant three-way interaction among the learning-based predictors revealed how the effects of the diversity of cues and paradigm typicality are fine-tuned by entrenchment of the word, i.e. the level of previous experience with a given word. Most importantly, in addition to observing significant effects of our critical predictors, we observed a striking similarity in the pattern of the effects. This similarity was further confirmed by the finding that predictions derived from the two statistical models share 85% of variance (after partialling out random effects).

This way we have demonstrated how two different approaches can jointly help elucidate the complex morphological phenomena. Most importantly, we have shown that complex morphological structures can appear as the consequence of the simple learning principle. We believe that the same principle could be applied to other language phenomena, as well.

CATCHING YOUR EYE: LOW-LEVEL PERCEPTUAL CUES INFLUENCE PRESCHOOLERS' SENTENCE FORMULATION

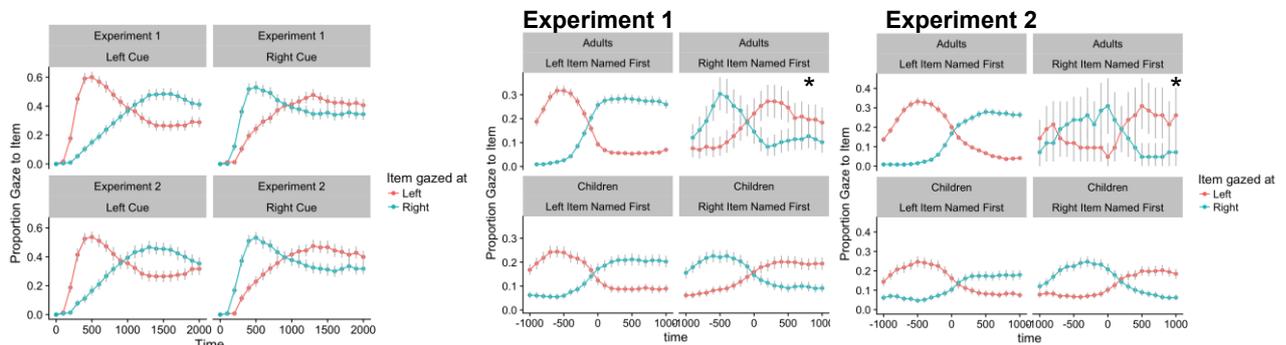
Laura Lindsay (University of Edinburgh), Hugh Rabagliati (University of Edinburgh) & Holly Branigan (University of Edinburgh)
L.M.Lindsay@sms.ed.ac.uk

When formulating a simple utterance, such as “*the cat and the dog*,” adults tend to look at the referents they are describing in the order of mention (e.g. Griffin & Bock, 2000), suggesting that their utterance planning is incremental (i.e. planned in small chunks). However, speakers must make a choice about their ‘starting point’ (MacWhinney, 1977), i.e., between talking about *the cat* or *the dog* first. In adults, the starting point can be influenced by both linguistic and non-linguistic factors, such as the speaker’s visual attention (Gleitman et al., 2007). Is the same true for children?

In two eyetracking experiments, we investigated whether 3-4 year olds’ referential productions were similarly influenced by low-level perceptual cues. To do this, we tracked participants’ eye movements as they named two pictures, one of which was preceded by a subliminal cue that was presented on-screen for 75 ms before the picture onset. In Exp 1, 3-4 year-olds (N=30) and adults (N=30) produced NP conjunctions (e.g. *the cat and the dog*). In Exp 2 (N=30 3-4 year-olds; 30 adults), participants produced complete sentences expressing the relative location of the objects (e.g. *the cat is next to the dog*), thus requiring participants to determine grammatical relations and generate a more complex constituent structure.

For both experiments, the cue influenced children’s first fixations (those within the first 500ms of picture onset; Fig 1). Critically, the cue influenced children’s order of mention: When the cue appeared on the left, children were more likely to name the left picture first than the right picture (E1: .65 vs .42, $p < .001$, E2: .57 vs .43, $p < .001$); when the cue appeared on the right, they were more likely to name the right picture first than the left picture. Furthermore, children fixated the referents in the order that they mentioned them. These results suggest that children’s utterance planning for very simple utterances is incremental, like adults (Fig 2). Low-level cues did not affect adult’s fixation patterns, perhaps because they showed an overwhelming bias to fixate and describe the left object first.

Overall, this suggests that, in an impoverished task whereby children produce constrained utterances with minimal semantics, children formulate their utterances incrementally, as adults do; moreover, their choice of ‘starting point’, including choice of sentence subject, can be influenced by low-level perceptual cues.



*N.B. Estimates are noisy as adults rarely named the right item first.

Fig 2: Proportion of gaze to each item as a function of order of mention across the different age groups in Exp 1 (left) and Exp 2 (right). Time 0 corresponds to speech onset.

children’s gaze to each item from picture onset as a function of which side the cue appeared on across each Experiment.

ERPS DO NOT SHOW THAT LEXICAL ACCESS DURING WORD PRODUCTION BEGINS WITHIN 200 MS

Daniel Kleinman & Kara Federmeier (University of Illinois at Urbana-Champaign)
 kleinman@gmail.com

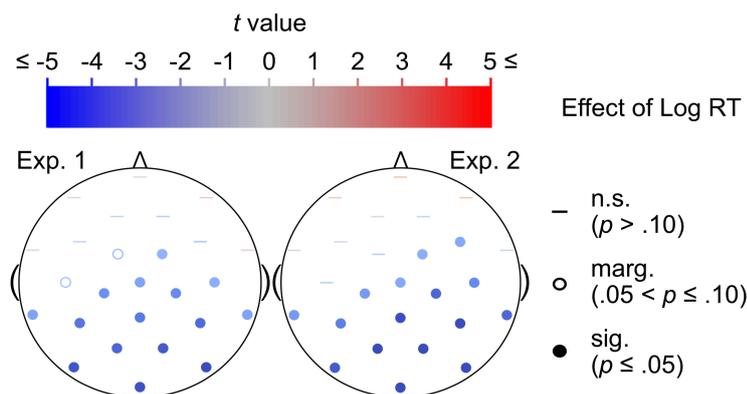
How long does it take to retrieve a word’s representation in long-term semantic memory when preparing to speak? EEG studies using language production tasks have found that the posterior P200 component has a greater amplitude when lexical selection is more difficult: Larger P200s are evoked when pictures are named in a speaker’s second (vs. first) language, or when they have lower- (vs. higher-)frequency names or have recently-named semantic competitors (Costa et al., 2009; Strijkers et al., 2010). Researchers have suggested this “production P200” component indexes how much speakers engage top-down attentional resources to boost the activation of hard-to-retrieve lexical items (Strijkers et al., 2011). Crucially, this means speakers must initiate lexical access within 200 ms – but ERP studies from other domains suggest it takes at least 300 ms to access long-term semantic memory (cf. Federmeier et al., 2015). However, production P200 studies have exclusively used picture naming tasks, which afford a single correct response that must be produced without a broader context – factors that may limit their generalizability. Here, we investigated the relationship between P200 amplitude and lexical selection difficulty using a sentence completion task, which affords many acceptable responses and a semantically rich context.

In two experiments, we recorded subjects’ EEG (Exp. 1: $n=40$; Exp. 2: $n=28$ so far) as they read 240 RSVP sentences that varied in constraint (how strongly the context predicted the final word). On 50% (Exp. 1) or 100% (Exp. 2) of the trials, the last word of the sentence was omitted and subjects instead saw a blank, prompting them to overtly produce a completion. (Subjects performed a different task with the remaining sentences in Exp. 1.) Response latencies were manually time-stamped; trials with RTs < 300 ms were excluded.

To determine the relationship between word selection difficulty and ERP waveforms, we selected our time window (190 to 240 ms after onset of the blank) to match Costa et al.’s (2009) and computed, for each trial, the average amplitude during that window for all 26 scalp electrodes. Then we assessed the relationship between each electrode’s mean amplitude and the subject’s (log) RT on each trial, using mixed-effects models with maximal random effects for subjects and items. t values for each electrode are shown below. We found a significant relationship ($p < .05$) between RT and P200 amplitude at central and posterior sites in both experiments, **but in the opposite direction from that previously observed**: A larger P200 was associated with *faster* RTs. (There was no effect of sentence constraint.) As every critical stimulus was a blank, this P200 effect cannot reflect visual properties of the stimulus. Furthermore, as results from posterior regions were the same when a 500 ms response cutoff was used, key findings do not simply reflect muscle artifacts.

These data show that the relationship between posterior P200 amplitude and lexical selection difficulty is task-dependent. We agree that the P200 indexes the recruitment of attentional resources during word production, but our results suggest that more attention may be associated either with increased difficulty (leading subjects to recruit extra cognitive resources when a picture looks like it will be difficult to name or they know they will have to

name it in their second language) or with better preparation (when subjects make better use of context to formulate a response). Importantly, under this account, the posterior P200 is not sensitive to the activation of specific lexical items. Thus, it is premature to conclude, on the basis of data from production P200 experiments, that speakers initiate lexical access within 200 ms.



GESTURE INCONGRUITY EFFECTS PRESERVED WITH VERBAL BUT NOT VISUOSPATIAL WM LOAD: AN ERP STUDY

Seana Coulson & Jacob Momsen (UC San Diego)

The present study investigated the roles of verbal and visuospatial working memory (WM) during the comprehension of speech accompanied by iconic gestures. In previous experiments, participants engaged in discourse comprehension tasks using naturalistic videos while concurrently performing secondary WM tasks to tax either verbal or visuospatial WM resources [1]. These studies suggest a relatively significant role of visuospatial WM in integrating information in co-speech gestures with concurrent speech. However, [1] relied on behavioral responses that occurred after the multimodal discourse. Here we utilize EEG to compare the impact of verbal versus visuospatial memory load on real-time processing of speech in multimodal discourse.

EEG was recorded as healthy adults performed a verbal (n=14) or a visuospatial (n=14) WM task. In the verbal condition, participants encoded a series of either one (low load) or four (high load) digits; in the visuospatial condition, they encoded a series of either one or four dot locations on a grid. During the rehearsal period of this WM task, participants observed a video of a man describing objects followed by a picture probe that showed the referent of his discourse. Gestures matched the speech in half of the videos, and mismatched the speech in the other half. ERPs were time-locked to the onset of the final item encoded during the memory task, the first content word in each video, and the onset of the picture probes.

ERPs time locked to the final item in the memory encoding task were measured 200-500ms post stimulus onset. Visuospatial load resulted in a widely distributed positivity, whereas verbal load produced a negativity (Dots: $F(1,13)=18.1$, $p < 0.001$; Digits: $F(1,13)=52.7$, $p < 0.001$). Differences in the load effects confirm that partially non-overlapping brain regions were recruited for each memory task.

N400 effects for the first content words in the videos were measured 200-400ms post stimulus onset. Repeated measures ANOVA in the digits condition with factors Load (low/high), Gestures (match/mismatch), and ROI revealed an interaction of Gestures by ROI ($F(6,78)=3.54$, $p < 0.05$). An identical ANOVA revealed no significant effects in the dots condition. The N400 effect in the digits condition resembled that reported in [2], and suggests participants' sensitivity to gestural information was preserved under the imposition of a verbal load. Its concomitant absence in the dots condition suggests co-speech gesture comprehension was compromised by the load on visuospatial WM.

The mean amplitude of ERPs to pictures was measured 200-500ms post stimulus onset to index the N300/N400. Pictures elicited no gesture effects in the digits condition. In the dots task, analysis revealed an interaction between Load and Gestures, reflecting gesture effects only in high load trials ($F(1,13)=6.39$, $p < 0.05$). Associated with poor task performance, we hypothesize that gesture effects on pictures emerge when participants were unable to suppress irrelevant gestural information during the videos.

Speech-gesture incongruity effects emerged on the speech under verbal load, and on picture probes presented afterwards under visuospatial load. Observed differences of the timing of speech-gesture incongruity effects thus support a dissociation in the contribution of verbal and visuospatial WM to multimodal discourse comprehension.

[1] Wu, Y. C., & Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta psychologica*, 153, 39-50.

[2] Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of cognitive neuroscience*, 19(7), 1175-1192.

EFFECT OF DISCOURSE AND ACTION ON VISUAL ATTENTION IN COMPREHENSION

Ruth Maddeaux (Toronto), Margaret Grant (Humboldt), Daphna Heller (Toronto)
 ruth.maddeaux@mail.utoronto.ca

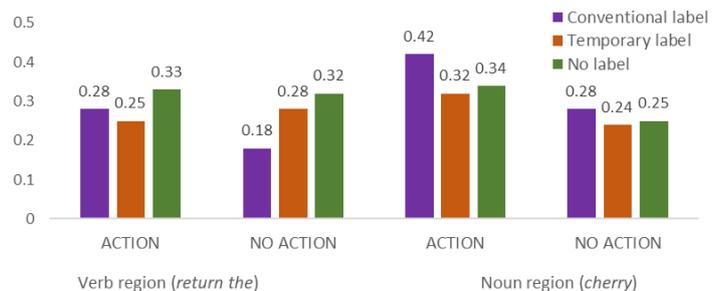
This paper examines the link between language and gaze in the visual world paradigm. We compare gaze when interpreting language that is directly about objects in the visual scene, with language that depends on discourse information encoded in a mental model: in an instruction like *return the cherry*: the verb *return* causes listeners to gaze at an object that has moved, which requires consulting a mental model of prior discourse [1], whereas upon hearing the noun *cherry*, listeners gaze at an object that directly fits the noun information [2].

Method. Two factors were manipulated (3x2 design). Each discourse had three sentences: (i) the movement of the target object, where labelling of the object and the action were manipulated (3 levels); (ii) an unrelated clicking instruction; (iii) an instruction to return or not return the target object (2 levels). First, we hypothesize that the labeling of the object and the movement will affect the encoding of the object in discourse, and may subsequently affect the salience of that object later in the discourse. Second, the action manipulation allows us to ask whether language and gaze are directly related – in which case both conditions should exhibit similar gaze patterns – or whether they are indirectly related via the need for action.

(i) labeling	(ii) clicking	(iii) return
Conventional label <i>Move the cherry to square 5</i>	<i>Now click on the chair</i> 	Action <i>Now return the cherry to square 1</i>
Temporary label <i>Move the object in the blue square to square 5</i>		No action <i>Now don't return the cherry to square 1</i>
No label [<i>cherry moves on its own</i>]		

Results (n=37). We analyze the likelihood of saccades to the target (we report only important aspects). **During the verb**, listeners made more saccades when an action was required ($z=2.19, p=.03$), indicating that the link between language and gaze is (at least in part) mediated by the need to plan a motor movement. Since the interaction was significant, we examine separately the action conditions, where labeling conditions did not differ ($ps>.35$), and the no-action conditions: listeners launched fewer saccades after conventional labels than after temporary labels ($z=2.01, p=.04$) or no label ($z=2.78, p=.01$). This is surprising: the conventional label should, if anything, render listeners *more* certain of the upcoming object. Instead, we follow [3] in proposing that, when the need for an action is removed, listeners look at the visual scene *to facilitate comprehension*, rather than as a direct reflection of the comprehension process [as in 4].

During the noun, listeners also launched more saccades with a required action ($z=2.50, p=.01$). However, here the interaction between labeling and action was not significant ($ps>.60$); the only effect was marginally more saccades after conventional labels than after temporary labels ($z= -1.89, p=.059$); this might be because after conventional label, the label *cherry* is being reused. As in the verb region, listeners looked less at the target with no action, but here this was not modulated by prior labeling.



Conclusion. When an action is required, listeners devote more visual attention to the relevant objects, independent of the relationship of language and the visual scene. Importantly, when interpretation requires an abstract discourse model, then the better information is encoded in the model, the *less* likely listeners are to attend to it in the absence of a goal: this is because when information is most salient in the mental model, listeners can easily consult this abstract representation and do not need visual support. This conclusion has important methodological implications for designing visual-world experiments.

References. [1] Chambers & San Juan (2008). *Cognition*; [2] Allopenna et al. (1998). *JML*; [3] Yee et al. (2009). CUNY poster; [4] Altmann & Kamide (2007). *JML*.

VISUAL CUES AND THE GRADED REDUCTION OF REFERENTIAL UNCERTAINTY

Mirjana Sekicki & Maria Staudte (Saarland University)
 mirjana@coli.uni-saarland.de

Previous research has shown that a speaker's referential gaze towards an object she is meaning to mention reduces the uncertainty about the target. This visual cue allows the listener to anticipate the likely resolution of the reference and to process the actual referent noun more easily. However, little is known about if and how a visual cue itself might reflect the reduction of uncertainty about the upcoming reference. The present study manipulated the degree to which cue specificity reduces uncertainty about the referent, aiming to shed more light on how visual cues are included in the interpretation of the linguistic material.

30 German native speakers took part in an eye-tracking experiment in the VWP which assessed their eye-movements and measured cognitive load from their pupillary activity. In order to inspire intuitive cue-following, we employed a gaze-like visual cue resembling typical behavior of the human speaker. By manipulating cue specificity (size of the cued group of objects) we also manipulated the uncertainty about which object will be mentioned, i.e. the referential entropy. Linguistic stimuli were identical among the three visual conditions of an item: *The man spills before the meal the juice* (original: *Der Mann verschüttet vor dem Essen den Saft*). As illustrated by Fig. 1, we presented visual displays of eleven objects, all fitting the verb selectional preferences. Gaze appeared upon the verb, cuing either a single object (*CueToOne*), a group of three (*CueToThree*), or of five objects (*CueToFive*). Referential entropy was highest prior to the visual cue. Gaze reduced it either totally (*CueToOne*), or to a certain degree. Crucially, when a group of five objects was cued, the cue reduced the referential entropy to a smaller degree and the actual referent noun became more surprising due to a bigger number of competitors (probability of 0.2 for each of them to be mentioned). When a group of three was cued, in contrast, each object in that group was more likely (0.33) to become the target referent. Finally, the referent noun disambiguated the sentence.

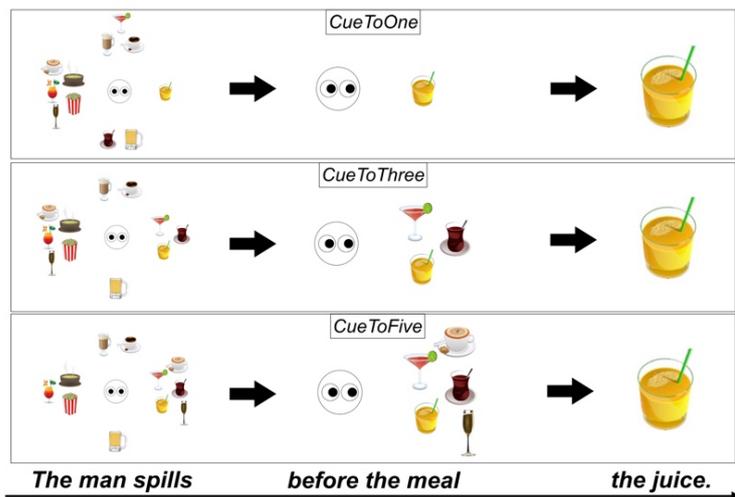


Fig. 1: Steps of uncertainty reduction in the three conditions (zooming in on the relevant objects for illustration purposes)

Our pupillary data show that the specificity of the visual cue has a quantitative effect on processing the linguistic reference. The bigger the target group was, the higher was the uncertainty about the target object, and consequently, we measured higher cognitive load on the noun: $CueToOne < CueToThree/Five$ ($p = 0.004$), $CueToThree < CueToFive$ (n.s.). The cognitive load measured at *CueToThree* was numerically positioned between the other two conditions but the differences did not reach significance. Interestingly, even though the visual cue played a crucial role in reducing the referential entropy, no differences in cognitive load were measured on the cue itself.

Prior to the appearance of the visual cue, no concrete anticipation was created about the referent so that the cue itself was not surprising. Upon presenting the cue, however, more information was acquired about the potential referent, which inspired a more concrete anticipation. Hence, the load induced by the referent noun reflected the noun's surprisal given the linguistic and visual context. We conclude that by eliciting a shift in the listener's visual attention, visual cues increase the probability of an object to become the next referent. Hence, a graded effect of cue specificity is measured on the cognitive load required for processing the linguistic reference.

PROCESSING OF AD HOC METONYMY: EVIDENCE FROM CO-REGISTRATION OF EYE MOVEMENTS AND ERPS

Petra B. Schumacher (Univ. of Cologne), Jenny Knowles (Univ. of Birmingham), Andrea Krott (Univ. of Birmingham), Steven Frisson (Univ. of Birmingham)
petra.schumacher@uni-koeln.de

Research on metonymic expressions like “reading Dickens” suggest that understanding a metonymic sense can be as *straightforward* as understanding the corresponding basic sense (e.g., “meeting Dickens”), at least as long as the different senses are already known to the reader (e.g., Frisson & Pickering 1999). In contrast, certain metonymic interpretations must be inferred ad hoc, illustrated by “the espresso” in (3), which is used by a restaurant supervisor to refer to a customer associated with this item (cf. Nunberg 1979). These types of meaning extensions have been reported to exert *processing costs* (e.g., Schumacher 2011). Here we ask whether this difference is due to the type of expression tested – i.e. conventionalized extensions vs. ad hoc extensions – or is tied to the methods applied – i.e. eye tracking in the former case and electrophysiology in the latter case.

The present research sought to align findings from these separate methodologies through the co-registration of eye movements and event-related potentials (ERPs). **Experiment 1** (N=19) considered the processing of novel metonymies such as “the espresso” in (1). Compared to non-metonymic control sentences (2), metonymic expressions elicited early eye movement disruptions and an N400 effect in the ERP data (250-450ms post-onset of first fixations). Later eye movement disruptions and a later right frontal negativity in ERP records (750-850ms post-onset) were also seen. **Experiment 2** (N=19) examined the role of context in the processing of metonymic expressions (3) and their controls (4). When metonymies were contextually-licensed, both early eye movement disruptions and N400 effects disappeared, but later eye movement effects remained. Additionally, a sustained right frontal negativity (250-850ms post-onset) emerged in the ERP data.

- (1) **Neutral context, metonymic expression:** *Claire asks Liam who it was who left without paying. Liam replies that **the espresso** left without paying.*
- (2) **Neutral context, control expression:** *Emma asks Jim what it is that's new on the menu. Jim replies that **the espresso** is new on the menu.*
- (3) **Supportive context, metonymic expression:** *The attendant asks her supervisor who it was who left without paying. Her supervisor replies that **the espresso** left without paying.*
- (4) **Supportive context, control expression:** *The waitress asks her boss what it is that's new on the menu. Her boss replies that **the espresso** is new on the menu.*

In light of these results, we argue that early effects observed in eye-tracking and ERP studies may represent the same underlying mechanism: an expectation-based prediction error that is modulated by the contextual manipulation (Exp.1 vs. Exp.2). The presence of supportive context is sufficient to eliminate early processing costs. This is in line with reports of context sensitivity of novel metonymy (Frisson & Pickering 2007, Schumacher 2014).

By contrast, our results do not confirm context-independent findings of later effects during the processing of this type of metonymy (Late Positivity in Schumacher 2011, 2014). This suggests that later processing costs seen in response to ad hoc metonymies may depend on the task demands associated with stimulus presentation (RSVP presentation in Schumacher 2011, 2014 vs. availability of the entire text in the present study). The frontal negativities may be due to greater demands posed by metonymic expressions during natural reading.

This research complements previous work on metonymy by indicating that (i) ad hoc metonymies exert processing demands reflected in both eye tracking and ERP measures. (ii) They thus differ from more readily accessible meaning extensions (i.e. “espresso” vs. “Dickens”). And most crucially (iii) presentation mode influences the underlying mechanisms.

Frisson & Pickering 1999. *JEP:LMC* 25(6). **Frisson & Pickering 2007.** *LCP* 22(4). **Nunberg 1979.** *Linguistics and Philosophy* 3. **Schumacher 2011.** *Experimental pragmatics/ semantics* (ed. by Meibauer & Steinbach). **Schumacher 2014.** *Philosophical Studies* 168.

L2 LEARNERS PREDICT AT THE LEVEL OF THE DISCOURSE: EVIDENCE FROM ERP

José Alemán Bañón (Stockholm Univ.), Elena Fano (Uppsala Univ.) & Clara Martin (BCBL)
jose.aleman.banon@biling.su.se

An ongoing debate in L2 processing research concerns whether adult L2ers can generate predictions online. Grüter et al. (2016) argue that adult L2ers have Reduced Ability to Generate Expectations. In contrast, Kaan (2014) posits that prediction in the L2 is similar to the L1, but is impacted by individual differences in cognitive factors that modulate prediction in L1 speakers. We used ERP to investigate L2 prediction at the level of the discourse. We examine the assignment of *Focus*. Focus can be defined as a licit answer to a *wh*-question. In English, it can be encoded via the *it*-cleft (see 1). We examine whether L2ers can use the cue provided by the *it*-cleft to anticipate the Noun Phrase (NP) with Focus status.

(1) What should Ann buy, a book or a calculator? **It is** _{FOCUS}[a book] that Ann should buy.

Participants read contexts and *wh*-questions like (2). We then recorded their brain activity while they read the answer to the *wh*-question (RSVP: 450/300ms). In (2a-b) the answer includes an *it*-cleft, which acts as a predictive cue for Focus assignment. Half of the times, Focus is assigned to an accessible NP (2a) and the other half, to the Topic NP (2b), thus violating information structure. In (2c-d) the answer does not involve the cleft, making Focus assignment less constrained/predictable. To better tap into prediction, we used the *a/an* alternation of the English indefinite article depending on whether the noun begins with a consonant (*a banker*) or a vowel (*an agent*). As shown in (2), the two NPs with Focus status (*an agent/an adviser*) and the Topic NP (*a banker*) select different articles (counterbalanced in the overall design). This allows us to examine prediction effects at the article, before semantic integration occurs (DeLong et al. 2005). The study involves 30 items per condition.

(2) Either an agent or an adviser could work for a banker. In your opinion, which of the two should a banker hire?

- a. In my opinion it is an agent that a banker should hire.
- b. *In my opinion it is a banker that should hire an agent.
- c. In my opinion an agent should be hired. d. In my opinion a banker should hire an agent.

L1-English speakers ($n=23$) show an N400 (250-400ms) for unexpected (2b) relative to expected articles (2a), but only in the conditions with the *it*-cleft (Cleft by Expectedness, $F(1,22)=7.52$, $p < .05$). This suggests that L1 speakers predict the phonological form of upcoming nouns (DeLong et al. 2005). Incorrectly focused nouns (2b, *banker*) yielded a P600 relative to correctly focused nouns (2a, *agent*) in the conditions with the *it*-cleft, suggesting that L1 speakers processed violations of information structure as structural errors. Without the cleft, Topic nouns (2d, *banker*) yielded an N400 relative to Focus nouns (2c, *agent*) suggesting a preference towards nouns that could value the *wh*-word.

L1-Spanish L2-English learners ($n=22$, intermediate/advanced) show an Anterior Positivity for unexpected (2b) relative to expected articles (2a). This effect, linked to prediction disconfirmation (DeLong et al. 2014) only emerged in the conditions with the *it*-cleft (Cleft by Expectedness, $F(1,21)=6.89$, $p < .05$). Incorrectly focused nouns (2b, *banker*) yielded an N400 effect relative to correctly focused ones (2a, *agent*), suggesting that the L2ers processed the information structure errors more lexically. Without the cleft, the N400 was larger for Topic (2d, *banker*) than Focus nouns (2c, *agent*), suggesting that L2ers also had a preference towards nouns that could value the *wh*-word. This effect was marginally smaller than in the conditions with the cleft (Cleft by Expectedness, $F(1,21)=3.33$, $p = .082$).

Finally, the size of the prediction effect on the article correlated with processing speed in both L1 and L2ers (Huettig & Janse, 2016). Our results are not fully consistent with either Grüter et al.'s or Kaan's proposals, but they show that L2ers can predict at the level of the discourse, although differently from L1 speakers. They also suggest that prediction in the L2 is impacted by similar cognitive factors as in L1 speakers (i.e. processing speed).

FORMING UNGRAMMATICAL STRUCTURES SHORT-SIGHTEDLY: LOCAL COHERENCE EFFECTS IN THE VISUAL WORLD PARADIGM AND READING

Kamide, Yuki & Kukona, Anuenue (Dundee University)

Previous research has shown that comprehenders form locally coherent yet globally ungrammatical syntactic parses during sentence processing. For example, Tabor et al. (2004) showed longer reading times for ‘tossed’ than ‘thrown’ in “The coach smiled at the player tossed/thrown a Frisbee”, wherein the ‘the player tossed a Frisbee’ also forms a local active parse that cannot be grammatically integrated with the global sentence constraints. Such results are incompatible with sentence processing theories that assume immediate constraint by global information. However, experimental evidence for local coherence effects is limited. In the current study, we ask: (i) whether local coherence effects persist outside of ambiguous stimuli (e.g., see the past / participle ambiguity of ‘tossed’); and (ii) how strongly do locally coherent parses compete with globally grammatical parses. In two eye-tracking experiments, we examined the effects of locally coherent structures on predictive processes (Exp 1) and reanalysis processes (Exp 2) using sentences such as (1a) and (1b):

- (1a) *Local Coherence (LC)*: **The girl who likes the man from London will ride the carousel.**
 (1b) *Non-Local Coherence (Non-LC)*: **The girl who likes the man very much will ride the carousel.**

Experiment 1: Visual world paradigm. Participants (N=32) viewed scenes with objects like a *young girl* (global agent), *man* (local agent), *carousel* (global object), *motorbike* (local object), *beer* (local distractor), and *sweets* (global distractor), while hearing sentences like (1a) or (1b) (or 2 counterbalanced versions with man/girl swapped; e.g., “The man who likes the girl... will ride the motorbike”).

Object	Local Coherence	Non-Local Coherence
	M (SD)	M (SD)
Local Object (motorbike)	16.56 (8.76)	11.93 (7.51)
Global Object (carousel)	13.84 (9.05)	16.30 (9.99)

The Global object was a predictable direct object of the sentences, while the Local object was a predictable direct object of the local coherence “The man... will ride the...” in (1a). The table shows proportions of looks to the Local and Global objects by condition during ‘ride the’. We found a reliable interaction between Object (Local / Global) and Sentence (LC / non-LC). Crucially, the Local

objects were looked at more in the LC vs. Non-LC condition. This pattern was not replicated amongst distractors, suggesting a structural effect. In addition, the Local Objects were not fixated reliably more than the Global objects in the LC condition (i.e., supported by the counterbalancing), suggesting that the locally coherent and globally grammatical parses were activated similarly.

Experiment 2: Reading eye-tracking. Participants (N=55) read modified versions of (1a) and (1b) (e.g., “**The little girl who likes the man from Rome / very much will ride the tricycle in the schoolyard.**”).

Region	Measure	Local Coherence	Non-Local Coherence
		M (SD)	M (SD)
Direct Object (‘the tricycle’)	Total	539 (162)	504 (144)
PP (‘in the schoolyard’)	First	234 (40)	225 (35)
	Go-past	1778 (1006)	1621 (840)

The table shows reading times (in msec) for regions with significant differences between conditions. We found longer total reading times for the direct object (‘the tricycle’) in the LC vs. Non-LC

condition. The same pattern was found in the subsequent region in first pass and go-past reading times, suggesting interference between the locally coherent parse and direct object, which required subsequent reanalysis.

Overall, our results provide support for local coherence effects in sentence processing among (e.g., lexically) unambiguous stimuli. Rather, locally coherent (yet ungrammatical) parses impacted on predictive processes (Exp 1) and reanalysis (Exp 2). We discuss implications of these results for sentence processing accounts that assume the processor always pursues grammatically well-formed constructions.

Why time and rhythm matter in speech/language comprehension

Sonja A. Kotz

Faculty of Psychology and Neuroscience
Dept. of Neuropsychology and Psychopharmacology,
Maastricht University, The Netherlands

The influence of time and rhythm in music is clearly recognized but there is less clear evidence on their impact in speech and language research (see Kotz & Schwartz, 2010). This is surprising as time and rhythm (i) play a significant role in speech and language learning, (ii) can compensate developmental and acquired speech and language disorders, and (iii) further our understanding of subcortical contributions to linguistic and non-linguistic functions. More specifically, recent neuroimaging and clinical evidence has confirmed the contributions of classical motor control areas (cerebellum (CE), basal ganglia (BG), supplementary motor area (SMA)) in timing, rhythm, music, and speech perception (Chen et al., 2008; Grahn et al., 2007; Geiser et al., 2009; Kotz et al., 2009; Kotz & Schwartz, 2011). We propose that serial order and temporal precision are shared mechanisms in simple and complex motor behavior (e.g. Salinas, 2009), but also in higher order cognitive functions such as speech and language (Kotz & Schwartz, 2010; 2015; 2016). I will present behavioral and neuroimaging evidence on the role of timing and rhythm in speech/language comprehension, and the compensation thereof in clinical populations. This empirical work will be discussed within a cortico-subcortical framework for speech/language processing.

The lateralization of expectations: Evidence from a divided visual field ERP study

Yoana Vergilova, Les Sikos & Matthew Crocker
Saarland University

To gain further insight into how expectation-based processing influences semantic access and subsequent message-level integration, we conducted two ERP studies exploiting a pseudoword manipulation. Expt 1 found that pseudowords (e.g., *geme*) which resemble contextually-supported words (e.g., *game*) are effectively treated as misspellings (cf. [1; 2]). In high-predictability contexts, semantic access (N400) of target (pseudo)words was facilitated and was followed by attempts to repair misspellings (posterior P600) and integrate plausible but unexpected targets into the context (anterior PNP).

Expt 2, which we focus on here, examines hemispheric contributions to the above findings by presenting target (pseudo)words to the left or right visual field (lvf, rvf), which selectively biases processing to the contralateral hemisphere (RH, LH respectively) [3]. Previous work suggests that contextual expectations have a greater influence on the processing of rvf/LH-presented stimuli than lvf/RH-presented stimuli [4; 5]. Therefore, if pseudowords are treated as misspellings of their corresponding words (cf. Expt 1), then in high-predictability conditions we expected greater attenuation of N400 effects for rvf/LH. We further predicted that the PNP effect associated with integrating low-predictability (pseudo)words (cf. Expt 1) would manifest in Expt 2 as a larger effect for rvf/LH than lvf/RH. While some prior evidence is consistent with this prediction [6], PNP results from hemispheric-differences studies have been mixed (cf. [5]).

Methods: 40 participants read 144 two-sentence stories in German (Potsdam Sentence Corpus; [7]). Context sentences manipulated the predictability (high, low) of target stimuli (word, pseudoword), which were embedded medially in otherwise identical final sentences. Pseudowords were created by replacing a medial letter in corresponding target words and resembled the target word more closely than any other word. Context sentences were presented in their entirety. Final sentences were presented word-by-word at a rate of 500ms, with target stimuli presented laterally (rvf/LH, lvf/RH).

Results: We applied repeated-measures ANOVAs with word predictability (high, low), word status (word, pseudoword), VF (left, right), and scalp topography as factors. **N400** (300-500ms): Predictability and word status elicited main effects for both rvf/LH and lvf/RH presentation, such that low-predictability conditions were more negative than high-predictability conditions, and pseudowords were more negative than words. No effects of lateral presentation were found. **Posterior P600** (800-1200ms): Unlike central presentation results, no main effects or interactions were found. **Anterior PNP** (800-1200ms): A significant VF * predictability * word status * anteriority interaction indicated that for rvf/LH-presented targets, low-predictability contexts elicited larger anterior PNP effects than high-predictability contexts, regardless of word status. No significant differences were found for lvf/RH.

Discussion: As in Expt 1, semantic access (N400) of targets was more difficult in low-predictability contexts and when orthographic cues were noisy (“misspelled”), and this was not modulated by visual field of presentation. This finding suggests, contra our predictions, that the hemispheres are similarly sensitive to top-down contextual information during semantic access. During the integration phase, we found no evidence of repair (P600) effects (possibly a consequence of reduced awareness of misspellings due to parafoveal presentation, cf. [8]). Crucially, however, low-predictability words and pseudowords elicited a PNP effect—typically associated with the integration of plausible but unexpected information into the message-level representation [9]—but only for rvf/LH-presented stimuli. Thus, while expectations influenced semantic access (N400) regardless of visual field, our observation that only rvf/LH-presented stimuli elicited expectation-driven PNP effects is consistent with the hypothesis that message-level integration relies primarily on LH mechanisms.

References: [1] Kim & Lai, 2012, *JoCN*; [2] van de Meerendonk et al., 2011, *NeuroImage*; [3] Bourne, 2006, *Laterality*; [4] Federmeier, 2007, *Psychophysiology*; [5] Wlotko & Federmeier, 2007, *Neuropsychologia*; [6] DeLong & Kutas, 2016, *Neuropsychologia*; [7] Dambacher et al., 2012, *Neuropsychologia*; [8] Sanford et al., 2011, *JoCN*; [9] Van Petten & Luka, 2012, *IJPsycho*.

CONTEXTUAL SPEECH RATE INFLUENCES MORPHOSYNTACTIC PREDICTION AND INTEGRATION

Greta Kaufeld¹, Wibke Naumann¹, Anna Ravenschlag¹, Andrea E. Martin^{1,2} & Hans Rutger Bosker^{1,2}; ¹Max Planck Institute for Psycholinguistics, Nijmegen, ²Donders Institute for Brain, Cognition and Behaviour, Nijmegen; *contact*: greta.kaufeld@mpi.nl

When processing and understanding spoken language, we are faced with a physical signal from which we – seemingly effortlessly – generate complex linguistic structures. This process requires the weighting and integration of a variety of relevant cues, but not much is known about which components of the speech signal the brain actually relies on to draw perceptual inferences and generate higher-level linguistic units [1].

One perceptual cue that has been studied extensively is contextual speech rate. Rate manipulations have been shown to substantially influence the perception of vowel length: For instance, the perception of a vowel ambiguous between short /a/ and long /a:/ in Dutch is biased towards /a:/ if preceded by a fast precursor [2]. It is unclear, however, whether this effect, known as rate normalization, bears wider implications for higher-level linguistic processing, such as prediction and subsequent cue integration. Here, we tested whether rate normalization can be observed for phonemes carrying morphosyntactic information, and whether it persists even in the presence of an earlier disambiguating syntactic cue.

Experiment 1 ($N=35$) investigated, using eye-tracking (visual world paradigm), the influence of contextual rate on the perception of the presence or absence of the morphosyntactic inflectional suffix /-ə/, marking gender on indefinite articles (feminine *eine* vs. neuter *ein*) in German. Spoken instructions to click on one of two pictures on screen included an ambiguous indefinite article *ein*[?] + adjectival phrase + target noun. We also introduced rate manipulations (PSOLA; slow vs. fast) in the precursor (*Now look at...*).

Pre-target eye fixations showed that participants were more likely to fixate the feminine object upon encountering *ein*[?] in a fast (vs. a slow) context. Moreover, after target onset, participants were slower to recognize the target word if the rate manipulation had biased them towards the distractor. These findings indicate that rate normalization can be observed for phonemes carrying morphosyntactic information, and as such influences prediction of upcoming referents. Moreover, it is strong enough to influence following target processing, thus affecting not only local perception, but also subsequent referential integration.

Experiment 2 ($N=36$) investigated the robustness of rate normalization, when combined with earlier morphosyntactic disambiguating target cues. Minimal object pairs, differing only in gender and vowel length (“*graf*_{NEU}” /*graf*/ *grave* – “*graaf*_{COM}” /*gra:f*/ *earl*), were presented with spoken instructions (e.g., “Kijk nu eens naar het_{NEU}/de_{COM} ADJ N”; *Now look at the_{COM}/the_{NEU} ADJ N*) including an early disambiguating definite article, a target noun ambiguous between /a/ and /a:/, and rate manipulations in the precursor.

Eye fixations showed an anticipatory target preference based on the gender cues on the article. Nonetheless, after ambiguous target vowel offset, we still found that participants were more likely to look at the object containing the long vowel /a:/ in fast (vs. slow) contexts, regardless of the preceding article. These results show that rate effects on vowel contrasts persist even when listeners are provided with a disambiguating cue well before the presentation of the target vowel.

Overall, our results indicate that contextual speech rates impact both morphosyntactic prediction and subsequent integration, implying that rate normalization may be a component of perceptual inference and a robust cue when perceiving language from speech. We show for the first time that rate-induced perceptual biases 1) are robust enough to impact morphosyntactic inference and cue integration further downstream, and 2) occur in the presence of early disambiguating information, thus impacting higher-level linguistic processing.

[1] Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in psychology*, 7.

[2] Bosker, H. R., & Reinisch, E. (2017). Foreign languages sound fast: evidence from implicit rate normalization. *Frontiers in Psychology: Language Sciences*, 8. Article 1063.

ATTENDING FAST AND SLOW 'COCKTAIL PARTIES': UNATTENDED SPEECH RATES INFLUENCE PERCEPTION OF AN ATTENDED TALKER

Hans Rutger Bosker (Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands),
Eva Reinisch (Institute of Phonetics and Speech Processing, Ludwig Maximilian University
Munich, Germany) & Matthias Sjerps (Donders Institute for Brain, Cognition and Behaviour,
Radboud University, Nijmegen, Netherlands)

HansRutger.Bosker@mpi.nl

Speech perception is affected by acoustic properties of surrounding contexts. For instance, durational cues in speech are perceived in relation to temporal cues in preceding speech. That is, the same sound, preceded by a fast context sentence, is perceived as longer as the same sound preceded by a slower context. This contrastive effect, known as rate normalization, has been argued to involve early general-auditory processes because the effect arises immediately as the target sound is being processed. It is generally thought to be beneficial for listening because it allows listeners to 'tune-in' to a given talker's speech rate.

However, often speech is heard in concurrence with *other* auditory signals (competing talkers, background music, busy traffic, etc.) that co-occur with the target speech stream. A well-known phenomenon is that selective attention can help listeners overcome the so-called 'cocktail party problem', allowing listeners to selectively enhance the processing of speech from a specific talker at the expense of the perception of unattended contextual signals. Attentional modulation of speech sound input occurs at early stages in perception.

However, it is largely unknown how perceptual contrast effects and selective attention interact during speech perception. This may seem surprising considering the ubiquity of situations in which the two effects would co-occur. In the present study we tested whether listeners who are presented with the speech of two speakers simultaneously would be able to selectively attend to one of these speakers and selectively use her and not the other speaker's speech rate to interpret a duration contrast. If rate normalization is a very early (general auditory) effect then no effect of attention would be expected. That is, we tested whether perceptual contrast enhancement functionally precedes selective attention.

In 6 experiments, (involving over 200 Dutch participants), participants categorized target words ($n=20$; produced by Talker A) that were durationally ambiguous between the presence/absence of the unstressed syllable /xə-/ (e.g., 'short' "geven" /'xevə/ *give* vs. 'long' "gegeven" /xə'xevə/ *given*). Target words were preceded by rate-manipulated context sentences of various lengths ($n=200$; made fast/slow using PSOLA in Praat).

Experiments 1 and 2 served as controls presenting only one speaker at a time. They showed that faster contexts biased participants' responses towards the 'long' target (Exp 1) and hearing the contexts in a different voice than the target did not reduce this effect (Exp 2).

Crucially, in Experiments 3-6, the same targets were preceded by *two simultaneous* context sentences from two different talkers, one in each ear (location counter-balanced). Rate was manipulated randomly within talkers, resulting in four different combinations: (rate-matching) fast+fast, slow+slow; (rate-mismatching) fast+slow, slow+fast. Half of the participants were instructed to attend to one talker, the other half to the other talker.

Participants' transcripts of the attended sentences demonstrated successful selective attention. Nonetheless, none of the four experiments provided evidence for attentional modulation of the rate effect. Specifically, attending a fast rate (and ignoring a slow rate) did not induce a greater proportion of 'long' responses, relative to attending a slow rate (and ignoring a fast rate); neither for talker-congruent attended sentences (Experiment 3), talker-incongruent sentences (Experiment 4 and 5), and not even when a video of the attended talker was provided (Experiment 6). Rather, fast+fast contexts consistently induced more 'long' responses compared to slow+slow contexts across the four experiments.

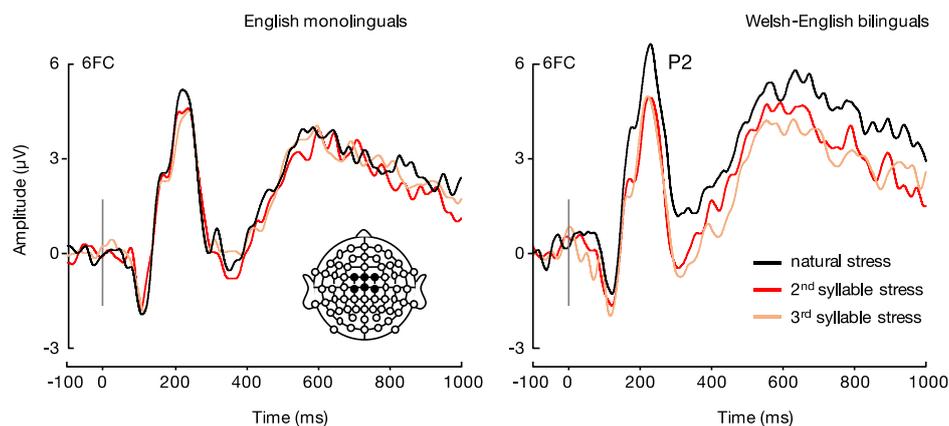
Together, these experiments indicate that normalization for speech rate is indeed characterized best as a duration-based perceptual contrast effect that is independent of selective attention: an unattended slow talker can 'cancel out' the biasing effect of an attended fast talker. Outcomes suggest that contrast effects may operate at a level in the auditory processing hierarchy that precedes attentional modulation.

DOES STRESS CLOSE THE LANGUAGE GATE? INHIBITION OF UNCONSCIOUS L1 ACTIVATION

Jennifer Lewendon (Bangor University) Anouschka Foltz (University of Graz) Guillaume Thierry (Bangor University)
J.Lewendon@bangor.ac.uk

Previous studies have shown spontaneous activation of the native language (L1) in late bilinguals tested in their second language (L2; Thierry & Wu, 2007; Wu & Thierry, 2010). On the other hand, prosodic information is known to influence lexical access in spoken word comprehension (van Donselaar *et al.*, 2005; Reinisch *et al.*, 2010). Therefore, although it has never been demonstrated, it is likely that the segmental and suprasegmental properties of spoken words modulate language non-selective access in bilinguals.

Here, we investigated whether lexical stress—a suprasegmental property—modulates unconscious native language access in fluent Welsh-English bilinguals. In an implicit priming paradigm, we presented correctly or incorrectly stressed English word primes followed by visual word targets. In critical trials, prime-target pairs concealed a sound repetition via Welsh translation. Participants were asked to perform an irrelevant semantic relatedness task on word pairs whilst undergoing 32-channel EEG recording. Word primes were either presented with their natural first-syllable stress, or incorrectly produced with stress on the second or third syllable. Two anomalous stress conditions were included to determine whether partial resemblance to Welsh penultimate stress might differentially affect lexical access, and ERPs were time-locked to presentation of the visual target.



We focussed on the P2 component of event-related potentials, which indexes word form repetition, cloze probability, and predictability. A repeated-measures ANOVA revealed a significant effect of L1 phonological overlap ($F_{1,11} = 7.39$, $p = 0.02$, $\eta^2 = 0.402$) and stress condition ($F_{1,11} = 7.07$, $p = 0.004$, $\eta^2 = 0.391$) on P2 mean amplitudes in Welsh-English participants. Critically, there was also a significant interaction between phonological overlap and stress ($F_{2,22} = 3.52$, $p = 0.047$, $\eta^2 = 0.243$), such that P2 mean amplitude elicited in incorrect stress conditions was significantly reduced as compared to the naturally-stressed condition only when L1 phonological overlap was present. None of these effects were found in monolingual English controls ($ps > .1$). We interpret the relative P2 mean amplitude increase as an index of phonological priming and an indication that bilingual participants spontaneously accessed Welsh translations only in the case of naturally stressed English primes. Thus, when processing speech in an L2 context, fluent bilinguals appear to only activate L1 representations when L2 words have natural stress, and possibly experience inhibition of native language access by anomalous suprasegmental information. More generally, this leads to the consideration that suprasegmental information is not only important in spoken language comprehension but even more so in a context of second language processing.

References. Thierry G and Wu YJ, 2007. Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104(30), pp.12530-12535. • Wu YJ and Thierry G, 2010. Chinese–english bilinguals reading english hear chinese. *Journal of Neuroscience*, 30(22), pp.7646-7651. • Van Donselaar, W., Koster, M. and Cutler, A., 2005. Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A*, 58(2), pp.251-273. • Reinisch E, Jesse A and McQueen JM, 2010. Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, 63(4), pp.772-783.

IT'S ALIGNMENT ALL THE WAY DOWN – BUT NOT ALL THE WAY UP

Rachel Ostrand (IBM Research)
ostrand.rachel@gmail.com

During conversation, speakers modulate characteristics of their production to match their interlocutors' characteristics, a process known as *alignment*. Speakers align at many linguistic levels, including syntactically (Branigan et al., 2000), lexically (Clark & Wilkes-Gibbs, 1986), temporally (Street, 1984), and phonetically (Pardo, 2013). An influential theory of alignment (Pickering & Garrod, 2004) proposes that communication is enhanced when partners align at multiple linguistic levels. We investigated this theory in two ways: by (1) testing for alignment at multiple linguistic levels within the same conversation; and (2) assessing whether speakers specifically matched their *current listener's* linguistic properties, rather than the overall linguistic statistics produced during the recent context.

Ninety-six subjects interacted with two experimenters (A and B), who alternately described sets of 6 dative pictures (24 pictures total from each experimenter). Then, the subject described 4 rounds of 12 pictures each, and A and B alternated as the listener for each round. Only one experimenter at a time was in the room with the subject, and every picture was unique. The degree to which subjects matched characteristics of the currently-listening experimenter's speech (as opposed to the non-listener's speech) was calculated.

To create strong pressures for partner-specific alignment, the experimenters had highly contrasting linguistic behavior. Syntactically, A produced only double object sentences (DO) and B produced only prepositional dative sentences (PD). Acoustically, one experimenter was a male native speaker of American English and the other was a female heavily-accented non-native speaker. Syntactic preference of the native and non-native experimenter was counterbalanced across subjects.

Subjects did *not* partner-specifically align in syntax: They were equally likely to produce PDs to the PD-producing (43.2%) as to the DO-producing experimenter (43.9%, $F < 1$). This was unaffected by the listener's native-ness, as subjects produced PDs at the same rate to both experimenters regardless of which experimenter produced PDs ($F < 1$).

Subjects *did* partner-specifically align on temporal characteristics: Subjects' speech rate was closer to the currently-listening experimenter's compared to the non-listening experimenter's (0.40 vs. 0.46 syllables/seconds apart; $p = .028$). Similarly, subjects' pause duration was closer to their current listener's pause duration compared to the non-listening experimenter's (1.15 vs. 1.36 seconds apart; $p = .031$).

Subjects did *not* partner-specifically align on several commonly-tested phonological measures: F0, vowel spectra energy, or the vowel space of /a/-/i/-/u/ (all $ts < 1$).

To explore a wider variety of features without excessive experimenter degrees of freedom (cf. Pardo, 2013), 129 acoustic (phonetic, phonological, and temporal) features were calculated on subjects' and experimenters' speech. For each subject, the difference between that subject's array of acoustic values and each experimenter's array of values was calculated. The difference scores (one for each feature) between a subject and their current listener is the "match" distance, and the difference scores between a subject and their non-listener is the "mismatch" distance. A machine learning binary classifier (using leave-one-subject-out cross-validation) was trained on the match and mismatch arrays for 95 subjects, and then predicted – for the 96th (left-out) subject – which difference score array was the match and mismatch. This process iterated, with each subject left-out. Thus the model measured the degree to which subjects sound more like A than B when addressing A, and more like B than A when addressing B. In contrast to some of the single-feature analyses, the collective acoustic features revealed partner-specific alignment, as the classifier performed significantly above chance (accuracy = 55.1%, $p = .002$). This indicates that speakers modulate some, but not all, acoustic properties to match their current listener, and using only one or two acoustic features may miss the presence of alignment.

Interlocutors can partner-specifically align at some linguistic levels (here, temporal and phonological), and not align at others (here, syntactic) in the same dialogue. Thus, communicative success may not require partner-specific alignment at all levels in tandem.

WHEN ATTENTION DISTRACTION HELPS RULE INDUCTION. AN ENTROPY MODEL

Silvia Rădulescu (Utrecht University), Mridhula Murali (Utrecht University), Sergey Avrutin (Utrecht University), & Frank Wijnen (Utrecht University)
S.Radulescu@uu.nl

What triggers the inductive leap from memorizing items and statistical regularities to inferring abstract rules? We propose an innovative information-theoretic model for both learning statistical regularities and generalizing to new input. Our entropy model predicts that *rule induction is an encoding mechanism triggered by the interaction between input complexity (entropy) and the limited encoding power of the human brain (channel capacity)*.

While traditional cognitive psychology claimed that rule learning relies on encoding linguistic items as abstract categories (Marcus et al, 1999), as opposed to learning statistical regularities between specific items (Safran et al., 1996), recent views converge on the hypothesis that one mechanism – *statistical learning* – underlies both item-bound learning and rule induction (Aslin & Newport, 2012; 2014; Frost & Monaghan, 2016). However, it is still not clear how a single mechanism outputs two qualitatively different forms of encoding – item-bound and category-based generalization, and what triggers the leap from one to the other.

In our model, less *input complexity (entropy)* facilitates finding regularities between specific items, i.e. item-bound generalization, while a higher complexity exceeding *channel capacity* drives category-based generalization.

In two artificial grammar experiments, we exposed adults to a 3-syllable XXY artificial grammar to probe the *effect of input complexity* on rule induction. We designed six experimental conditions with different degrees of input complexity and we used entropy to measure the complexity. Results showed that when input complexity increases, the tendency to infer abstract rules increases gradually (Fig.1)

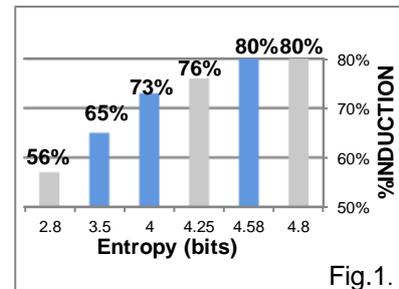


Fig.1.

Next, we aimed to capture the effect of overloading *channel capacity* in the process of rule induction. More specifically, we hypothesized selective attention and working memory modulate channel

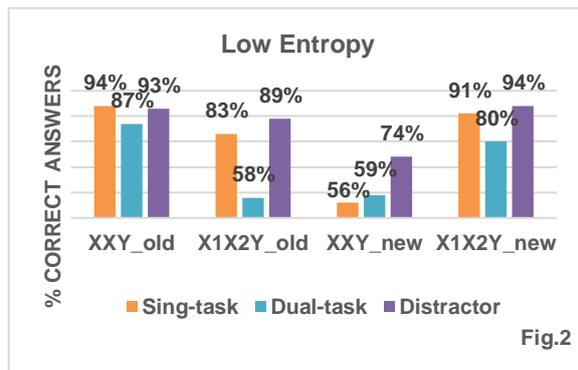


Fig.2

capacity. Thus, in a new experiment, adults' channel capacity was overloaded by simultaneously playing a stream of digits while exposing them to the lowest entropy grammar (2.8 bits). In one condition, participants had to pay attention to the digits and recall certain digits afterwards. In the second condition, they were asked to ignore the digits. Participants gave grammaticality judgements on: trained XXY strings (correct), new XXY strings (correct), ungrammatical X1X2Y (3 different trained syllables), and ungrammatical new X1X2Y strings. Results showed participants were more likely to generalize (i.e. accept new XXY) when their attention was divided between two active tasks which overloaded their working memory (Dual-task; Fig.2) compared to when no digits were played simultaneously (Single-task; Fig.2), despite the low entropy input. Moreover, distracting participants' attention passively by digits played simultaneously (Distractor; Fig.2) further increased their tendency to generalize (the difference in acceptance of XXY old vs XXY new decreased). In fact, we found the same pattern of responses (generalization) both by overloading participants' channel capacity in this dual task, and by increasing the input entropy from low to medium in the single task. Additionally, overloading participants' channel capacity with the irrelevant inflow of digits yielded the same tendency to generalize exhibited when increasing the input entropy from low to high in a single task. These findings support our main hypothesis that rule induction is triggered by the interaction between input entropy and channel capacity.

MODELLING OF MISMATCH NEGATIVITY RESPONSE AND NON-NATIVE STATISTICAL LEARNING

Jacolien van Rij¹, Jessie S. Nixon², Tomas O. Lentz³

¹University of Groningen, ²University of Tübingen, ³University of Amsterdam

Intro. Over the last several decades, a great deal of research has investigated the extent to which statistical learning can explain human listeners' linguistic knowledge (Maye, Weiss, & Aslin, 2008; McMurray, Aslin, & Toscano, 2009). Recent mismatch negativity (MMN) research suggests that statistical learning with the oddball paradigm can improve not only between-category discrimination, but also within-category acoustic cue perception (Nixon et al., 2018). To date, there have been very few quantitative models of MMN development during statistical learning of speech sounds. The present study uses Rescorla-Wagner discrimination learning to model trial-by-trial changes in EEG amplitude during statistical learning of non-native Cantonese lexical tones.

Method. EEG data from 19 native German speakers was recorded as they listened to stimuli. Stimuli were speech recordings manipulated into a 13-step pitch continuum from Cantonese mid-level to high-level tone with 0.14 semitones between steps. A sequence of four sounds was presented, sampled from one Gaussian (the standard, e.g. high tone) followed by one sampled from the other Gaussian (deviant, e.g. low tone) of a bimodal distribution. The input to the model was the auditory stimulus (continuum step) and the output was lexical tone (high or low). To model the continuous nature of the acoustic cues and sensitivity to pitch differences, three simulations used different cue-weight window sizes: 1) a single cue (one step), 2) one step either side or 3) two steps either side of the input cue.

Results. All simulations showed steeper learning for the standard in the first block. However, simulations predicted different learning effects depending on the window size of sensitivity to the pitch differences. With perfect discrimination of each stimulus (single input cue for each individual continuum step, i.e. window size = 1), connection weights followed a bimodal distribution after training. However, with larger window sizes, reflecting lower sensitivity, connection weights had a steep slope near the category boundary but then flattened out, reflecting 'categorical perception'. This pattern bears striking resemblance to offline data from two tasks: pitch estimation and discrimination. Pitch estimates were sensitive to single steps in the estimation task (high sensitivity) and developed a bimodal shape after training. Listeners seldom detected one-step differences in the discrimination task (low sensitivity), except near the category boundary. The distribution after training was almost flat, with a dip at the category boundary.

Discussion. Simulations showed steeper learning for cues sampled from the standard than the deviant distribution, as well as different trajectories for different continuum steps, reflecting the presentation frequency of the bimodal input distribution. Of particular interest is the effect of cue sensitivity, with bimodal representations emerging only when each continuum step could be discriminated (window size 1) and a relatively flat distribution with a dip at the boundary for lower sensitivity (window size 3). This predicts an effect of task on cue sensitivity, with lower sensitivity for category discrimination than for non-discrete estimation tasks. In light of the discrimination learning results, we will perform simulations using neural networks capable of representing past events as well as current input (namely LSTMs). Simulations of the two types of models will be compared for fit to behavioural and EEG data.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, 11(1).

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, 12(3), 369–378.

Nixon, J. S., Boll-Avetisyan, N., Lentz, T. O., van Ommen, S., Keij, B., Çöltekin, Ç., ... van Rij, J. (2018, June). Short-term exposure enhances perception of both between- and within-category acoustic information. In *Proceedings of the 9th International Conference on Speech Prosody 2018*.

YOUR EARS OR YOUR BRAIN? NOISE STRUCTURE CAN HIDE GRAMMATICAL PREFERENCES

Suhas Arehalli (Johns Hopkins University, UC San Diego) & Eva Wittenberg (UC San Diego)

sarehal1@jhu.edu

Listeners give speakers the benefit of the doubt – when there is noise in the signal, people seemingly effortlessly correct for it. This process has been described by the Noisy Channel model (Levy, 2008; Gibson, Bergen, & Piantadosi, 2013). We also know that people sensitize themselves to different kinds of errors, depending on the input: They pay more attention to word forms when proofreading for typos, and more to word meanings when reading texts with lexical substitution errors (Schotter, Bicknell, Howard, Levy, & Rayner, 2014), and the rate of error correction often depends on both grammar and semantic plausibility (Gibson et al., 2013; Poppels & Levy, 2016). But will people treat the same error differently depending on different contexts? If so, which kind of linguistic information can influence the correction process? In three experiments, we tested the influence of different kinds of noise in the filler structure on correction of misheard function words: We manipulate the ratio of randomly placed acoustic noise to noise that was placed over a function word (Exp.1 and 2); in Exp.3, we increase the ratio of noise on two types of function words to random noise. We predict that people’s noise correction process will be driven by an expectation that speakers generate grammatical sentences, and thus the more ungrammatical an input, the more people repair. We also predict that when noise is systematically placed, people will suppress this top-down, expectation-driven correction process and repair based on the auditory input alone.

Exp.1 replicates a study that investigated how people reconstruct the input when noise in a speech signal results in structures that are on a gradient of acceptability (Mack, Clifton, Frazier, & Taylor, 2012): Constructions with or without an expletive subject, such as “It/Ø seems to me like it’s going to rain.” Empty subject positions are usually ungrammatical in English, but ameliorated for present-tense, first-person utterances. Mack et al. (2012) constructed 24 of these kinds of sentences, manipulating two factors: personal immediacy (“seems to me” vs “seems to her”) and temporal immediacy (“seems” vs “seemed”). Each sentence recording was distorted at several points, one of them on the subject region. This distortion was designed to leave ambiguity as to whether there was an “it” under the noise. Subjects were asked to both repeat and type what they had heard for 24 critical items and 60 fillers. The dependent measure was the rate of “it” restoration.

We replicated this study (N=48) and found, exactly like Mack et al. (2012), that in past-tense sentences “seemed to me”, people restore “it” more ($F(1, 22.76)=8.71, p<.01$); the same trend was also found for less personal (“... to her”) sentences ($F(1, 23.34)=3.13, p<.10$). We take this as evidence that the more ungrammatical an input, the more people repair; and the two factors contributing to grammaticality of subjectless sentences are independent of each other. Crucially, this process was triggered by grammatical rules that do not affect plausibility; and what determined whether a repair was triggered was the degree of grammaticality, while holding the process and target of the repair constant: People only ever repaired an expletive “it.”

In Exp. 2 and 3, we ask whether people simply picked up on the experimental manipulation and developed repair strategies tailored to the overall structure of the experiment. Thus, we manipulated the overall ratio of noise on function words vs. random noise on the fillers (Exp.2). Exp.3 asks whether systematic noise on two different function words changes people’s strategy.

Exp.2 reduced the number of items to maximize the randomness of noise in the fillers: 24 “it”-repair items, and 24 fillers with noise overlaid at random points. Analyzing the spoken responses (N=48), we find again that present-tense sentences have lower restoration rates ($F(1, 21.87)=5.63, p<.05$), a marginal effect of personhood ($F(1, 20.51)=3.22, p<.10$) and a marginal interaction of tense and personal immediacy ($F(1, 22.11)=3.66, p<.10$).

Exp.3 used a subset of Mack et al.’s (2012) fillers that had overlaid noise on another function word: sentence-initial conjunction “that” (Ferreira, 1997), such as in “What did he say? That/Ø AMLaP is great”. Again analyzing the spoken responses (N=48), we only find the trend for present-tense sentences ($F(1, 46.97)=6.54, p<.05$). Comparing across experiments 2 and 3 yields a highly significant difference between restoration rates ($t(2301)=2.17, p<0.0016$).

Summary. We take this as initial evidence that people adjust their error correction process to the experimental environment in the following ways: When they cannot detect a pattern in the noise, they rely on their expectations for grammaticality (Exp.1 and 2). Thus, the messiness of the input brings forth their expectation about how speakers use grammatical rules. However, a discernible pattern in the noise leads to a reduction of the use of top-down cues in the correction process (Exp 3): People pay more attention to the actual acoustic input and use grammatical preferences (tense and personhood) less.

HOW SPEECH RATE NORMALIZATION AFFECTS LEXICAL ACCESS

Merel Maslowski (Max Planck Institute for Psycholinguistics), Antje S. Meyer (Max Planck Institute for Psycholinguistics), & Hans Rutger Bosker (Max Planck Institute for Psycholinguistics)
merel.maslowski@mpi.nl

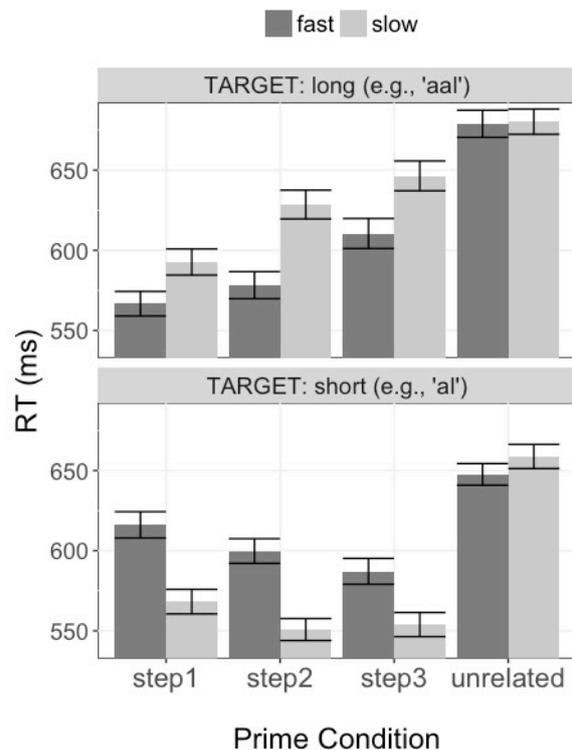
Speech can be produced at different speech rates, with considerable variation between talkers, within talkers, and even within utterances. Listeners have been suggested to cope with this temporal variation by normalizing segmental durations for surrounding speech rates. This has, for instance, been demonstrated in categorization experiments in which participants are presented with ambiguous speech sounds (e.g., between short /a/ and long /a:/ in Dutch) embedded in fast or slow precursor sentences. A fast speech rate typically biases target perception towards the long vowel and a slow speech rate to the short vowel.

This rate normalization process has been argued to involve low-level automatic perceptual processing, since effects arise immediately upon hearing the target vowel, are insensitive to talker voice changes between precursors and targets, and can even be induced by non-speech precursors (e.g., tones). However, rate normalization has been studied exclusively with a two-alternative forced choice (2AFC) paradigm, where participants' attention is directed to the vowel contrast. If rate-dependent speech perception involves early general auditory mechanisms, rate normalization should also influence linguistic processing when no overt categorization response is required. We investigated this hypothesis using a cross-modal repetition priming paradigm. This also allowed us to assess how normalization affects lexical access processes.

Experiment 1 ($N = 12$ native speakers of Dutch) established rate normalization in a 2AFC task. Three-step spectral continua (1: /a:/-like; 3: /a/-like) were created for 75 minimal target pairs (e.g., *al* /ɑl/ "already" - *aal* /a:l/ "eel"), and embedded in a rate-manipulated fixed precursor sentence (using PSOLA). As expected, categorization data revealed effects of the spectral continua and of the precursor, with fast precursors biasing perception towards /a:/.

Experiment 2 ($N = 80$) involved cross-modal repetition priming with a lexical decision task, using the sentences from Experiment 1 as auditory primes (plus control primes with unrelated words without the /ɑ-a:/ contrast). Targets were always the short or the long member of the /ɑ-a:/ minimal pairs, presented orthographically immediately after prime offset (plus an equal number of non-words with /ɑ/ or /a:/). The response times in each condition are illustrated on the right. Linear Mixed Models showed a significant interaction between Target Word (top panel: long; bottom panel: short) and Prime Condition (Spectral Step + unrelated controls), indicating shorter RTs for 'long' targets with more /a:/-like primes (and vice versa). Importantly, we also found an interaction between Target Word and Precursor Rate (dark grey: fast; light grey: slow). This interaction revealed shorter RTs for 'long' targets with fast primes, but longer RTs for the same targets with slow primes (and vice versa for 'short' targets).

These findings show that speech rate effects are induced even when no explicit attention is drawn to the temporally ambiguous word. This outcome is compatible with the view that rate normalization involves low-level perceptual processing, in turn affecting higher-level linguistic processes such as lexical access.



ON PREDICTION OF PHONOLOGICAL AND GRAMMATICAL GENDER INFORMATION

Aine Ito (Humboldt University of Berlin), Chiara Gambi (Cardiff University),
 Martin J. Pickering (University of Edinburgh), Kim Fuellenbach (University of Oxford) and
 E. Matthew Husband (University of Oxford)
 aine.ito@hu-berlin.de

When a sentence context is strongly predictive towards a specific word, people sometimes predict that the word is likely to occur in the sentence [1]. But do people predict when that word is likely to occur? To examine whether people predict that an expected word immediately follows a strongly predictive context, we created local and global mismatches between an article and a predictable noun. We used indefinite articles in Italian, since they agree with the noun both in phonology and gender if the article is adjacent to the noun but do not necessarily agree in phonology if the noun occurs later (i.e., their phonological form is conditioned by the next word, which may be an intervening adjective).

20 native Italian speakers read sentence contexts that were predictive towards a specific noun (cloze $M=84\%$; word-by-word 600ms SOA). These were followed by the expected noun (e.g., *un incidente*: 'accident') or another plausible but unexpected noun (cloze $M=2\%$). Unexpected nouns either began with a different class of sounds (consonant vs. vowel, e.g., *uno scontro*: 'collision') or had a different gender (e.g., *un'inondazione*: 'flooding'), thus required an article that mismatched the expected noun in phonology or gender. If people predict immediate occurrence of the expected noun, articles that mismatched the expected noun's phonology or gender should both disconfirm expectations and elicit differential ERPs relative to matching articles. However, if people do not predict when the expected noun will occur, just that it will occur, only gender-mismatch articles would disconfirm expectations, and mismatch effects should be found for gender but not phonology.

Figure 1 shows preliminary ERP data for articles at CP1 and for nouns at CP2 and AFz. Both mismatch articles elicited greater negativity than expected articles at posterior channels around 500-1000ms ($ps<.05$). Unexpected nouns showed a classic, posterior-distributed N400 effect ($ps<.001$) followed by anterior positivity ($ps<.001$) relative to expected nouns.

The late negativity at the articles may indicate detection of a mismatch between the encountered and expected information or reanalysis of expected information [2]. The N400 effects at the nouns may reflect the difficulty of integrating unexpected nouns into the context, and the anterior positivity may indicate suppression of the expected but unencountered noun [3]. The negativity at the gender mismatch articles may index revision of the prediction for a specific noun (because the expected noun cannot follow the article), but we suggest the negativity at the phonological mismatch articles indicates that comprehenders can revise predicted timing as well, as the expected noun may still occur later in the sentence.

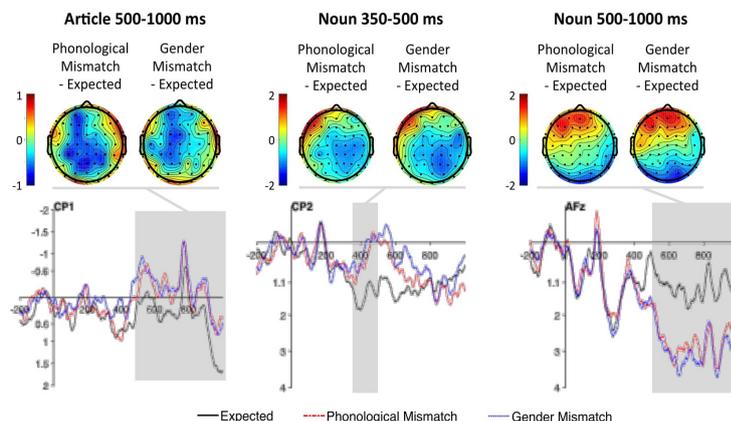


Figure 1. ERPs for the articles (left bottom) and nouns (middle & right bottom). The top panel shows scalp distributions of the ERP effects in 350-500ms (middle) and 500-1000ms (left & right) windows.

- References** [1] Van Petten & Luka (2012). *Int. J. Psychophysiol.*
 [2] Van de Meerendonk et al. (2009). *Linguist. Lang. Compass.*
 [3] DeLong et al. (2014). *Neuropsychologia.*

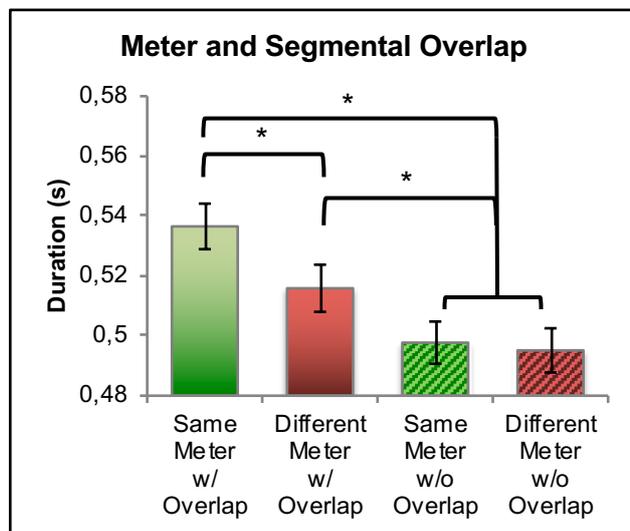
METER AND PHONOLOGICAL ENCODING DURING SPEECH PRODUCTION
 Brett R. Myers (Vanderbilt University) and Duane G. Watson (Vanderbilt University)
 brett.myers@vanderbilt.edu

Words are lengthened when they are more difficult to produce, which may occur based on a variety of explanations (e.g., Kahn & Arnold, 2012). Scene description experiments show that speakers lengthen target words when previously uttered words share initial phonological segments (e.g., *The canoe shrinks. The candle flashes*). The general claim is that lexical competition between similar sounding words creates interference that slows selection of phonemes in the second word (i.e., *candle*; Yiu & Watson, 2015). In two experiments, we investigate whether interference can also be induced by metrical similarity. Such an effect would suggest that phonological planning includes abstract representations for metrical structure. In Experiment 1, participants completed an event-description task in which a disyllabic target word shared segmental overlap with a prime word, and either had matching stress (e.g., *The butter shrinks. The button flashes*) or non-matching stress (e.g., *The baton shrinks. The button flashes*). Participants lengthened words more in trials with both segmental and metrical overlap, compared to trials with segmental overlap alone. This lengthening could either be the result of metrical interference or having uttered a prime with similar segmental realizations. To adjudicate between these possibilities, Experiment 2 included the same conditions as Experiment 1, as well as segmentally distinct word pairs with either matching stress (e.g., *The pickle shrinks. The button flashes*) or non-matching stress (e.g., *The cigar shrinks. The button flashes*). Participants again showed lengthening in trials with both segmental and metrical overlap, but no lengthening from metrical overlap alone. These data suggest that acoustic similarity between initial phones in prime and target words, characterized by segmental and metrical overlap, leads to lengthening. Results are discussed in the context of production models based on a speakers' auditory memory of the discourse (Jacobs et al., 2015). We posit that word duration may be influenced by the speaker's recent auditory experience, especially phonological content, which is emphasized by metrical structure.

Figure 1. Example display of the event description task where *casket* and *castle* form a critical prime-target pair, and *eggplant* and *moustache* are filler items.



Figure 2. Results from Experiment 2. Prime-target pairs with the same meter and segmental overlap (solid green) were significantly longer than trials with different meter (solid red). Pairs with no segmental overlap (striped green and red) were reduced.



12 Poster Presentations - Abstracts Thursday

Poster Abstracts Thursday, Sept 06, 2018

Thursday Poster.1

PERCEPTUAL ADAPTATION TO SEGMENTAL SOUNDS IN NON-NATIVE SPEAKERS

Hiroki Fujita (University of Reading), Ruri Ueda (Osaka Kyoiku University) & Ken-ichi Hashimoto (Osaka Kyoiku University)
h.fujita@pgr.reading.ac.uk

Previous research has shown that native speakers adapt to different regional dialects and foreign accents rapidly [2,3]. Some studies show that non-native (L2) speakers also adapt to foreign accents [6], but little is known about how they adapt to L2 segmental sounds that are difficult to perceive in the development of L2 phonetic acquisition. In this study, we explored whether and to what extent Japanese language learners of English ($n = 37$) adapt to English phonetic contrasts (/a/ - /æ/ and /s/ - /θ/) that they are known to have difficulty distinguishing [4,5], using a two-alternative forced choice task. In the task, they listened to a word followed by two visual probe words and judged which of the visually presented words matched the aurally presented word. Comprehension accuracy rates and response times adjusted to remove the adaptation effect to the task itself [see 2 for a similar method] were measured during the task, which was then divided into four blocks to see how these measures change over the course of the experiment. Target words in Blocks 1-3 were identical while Block 4 comprised different target words to see if adaptation effects differ between the same and different words. These words were matched for frequency, length and lexical decision speed according to the English Lexicon Project [1]. Block 1 was a pre-exposure phase, Block 2 was an exposure phase, and Blocks 3-4 were test phases. A reliable main effect of Block was observed with shorter response times to both phonetic contrasts (/a/ - /æ/ and /s/ - /θ/) in Block 3 compared with Block 1, showing phonetic adaptation. We also observed marginally significant shorter response times to only the consonant contrast (/s/ - /θ/) in Block 4 than Block 1. Accuracy did not change significantly across Blocks 1-4. These results suggest that L2 learners generally adapt to difficult L2 segmental sounds rapidly and that this rapid adaptation is more closely related to the reduction of processing cost rather than to the capacity to distinguish similar sounds accurately.

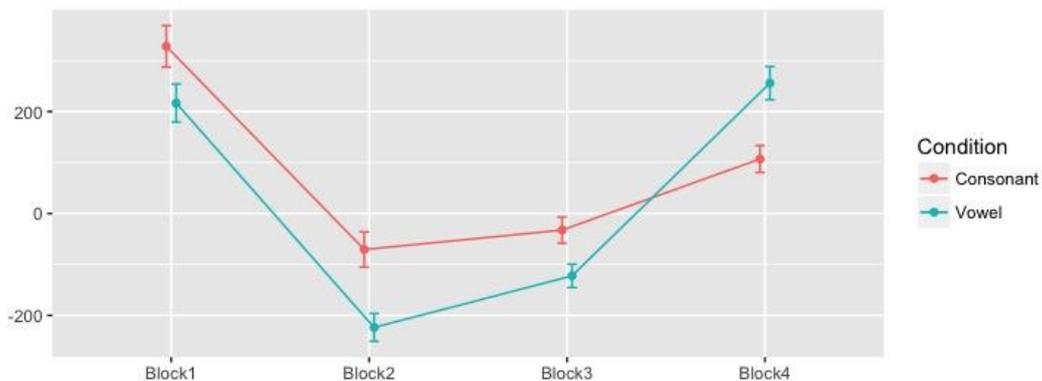


Figure 1. Adjusted response times during exposure and test phases

References

- [1] Balota et al. (2007), *BRM*, 39, 445-459; [2] Clarke & Garrett (2004), *JASA*, 116, 3647-3658; [3] Dahan et al. (2008), *C*, 108, 710-718; [4] Guion et al. (2000), *JASA*, 107, 2711-2724; [5] Strange et al. (1998), *JP*, 26, 311-344; [6] Weber et al. (2014), *JP*, 46, 34-51.

Thursday Poster.2

MAPPING SPEECH SEGMENTATION TO LINGUISTIC PROPERTIES – WHAT COUNTS?

Anna Mauranen and Alena Konina (University of Helsinki)
anna.mauranen@helsinki.fi

Humans make sense of speech by segmenting it into suitable chunks. In this, it can be likened to other kinds of perceptual segmenting, for example that for objects (e.g. Biederman 1987) or events (e.g. Zacks & Tversky 2001; Kurby & Zacks 2015). The fast flow of continuous speech places additional demands on working memory (Christiansen & Chater 2016), further enhancing the need to segment the speech stream into manageable units. Speech segmentation orients to understanding, and progresses by chunking up the input into processable units (Sinclair & Mauranen 2006). We hypothesize that chunking is driven by a search for meaning, which is facilitated by linguistic cues at various level of structure and sound. We further hypothesize that the perceptually most salient boundaries coincide with places where linguistic cues converge. We report an experiment that relates listeners' spontaneous chunking behaviour to linguistic properties of the input. The data is authentic (cf. Willems 2015) recorded and transcribed speech; participants (n= 48) chunk up 30-s speech samples (n=66) in which they intuitively mark boundaries using a tablet application (Vetchinnikova, Mauranen & Mikusova 2017) as they listen. That is, they hear the extract and manually mark the boundary in the transcript. After this behavioural experiment, the data is analysed in terms of Linear Unit Grammar, classic syntax, pauses and prosody. 78% of strong perceived boundaries is explained by LUG-predicted boundaries (with strong correlations between the two analyses in total ($\chi^2(4) = 376.059$, $p = 0.000$)). The way the majority of participants chunk extracts correlates well with clausal analysis of the data ($\chi^2 = 2579.740$, $p = 0.000$), yet only 56,5% of boundaries marked by the majority (1st quartile) of participants, aka strong boundaries, coincide with boundaries between clauses in terms of classic syntax (Biber et al., 1999). Intonation unit boundaries extracted from the data and the perceived boundaries are powerfully correlated as well ($\chi^2(1) = 1894.238$, $p = 0.000$). Our findings indicate that chunk boundaries where participants converge most are junctures where structural, prosodic and meaning patterns also converge.

Christiansen, M.H. & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *Behavioral and Brain Sciences*.1-19. doi:10.1017/S0140525X1500031X

Kurby, C. & Zacks, J. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences* 12 (2), 72-79.

Kurby, C. A., & Zacks, J. M. (2015). Situation models in naturalistic comprehension. In R. M. [Ed Willems (Ed.), *Cognitive neuroscience of natural language use* (p. 59–76, Chapter xiv, 265 Pages). Cambridge University Press (New York, NY, US).

Sinclair, J. & Mauranen, A. (2006). *Linear Unit Grammar*. Amsterdam: John Benjamins.

Vetchinnikova, S., Mauranen, A. & Mikusová, N. (2017). Investigating the relevant units of online speech processing. In *INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association, August 20-24, Stockholm, Sweden. Proceedings*: 811-812.

Willems, R. (Ed.) (2015). *Cognitive Neuroscience and Natural Language Use*. Cambridge: Cambridge University Press.

Thursday Poster.3

TWO MECHANISMS OF SCALAR IMPLICATURE IN COMPARATIVELY MODIFIED NUMERALS

Christoph Hesse & Anton Benz (Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin)
hesse@leibniz-zas.de

Our presentation is on scalar implicatures of numerals modified by the comparative quantifiers 'more than' and 'fewer than'. It is widely assumed in the literature that modified numerals do not generate scalar implicatures as the comparative quantifiers indicate that the speaker is ignorant of the exact number (Breheny, 2008). However, speakers routinely employ experiential world knowledge to draw scalar implicatures in the presence of comparative modifiers. For instance, in the context of 'an exceptionally hot summer', speakers will use their knowledge of average summer temperatures and record temperatures to construct a range of potential values. Other contexts, such as 'the number of signatures on a petition', are less informative about potential values and speakers may lack sufficient information to form expectations. We start from Cummins, Sauerland, & Solt (2012), who show that scalar implicatures are available from comparatively modified numerals, and who argue that modified numerals' roundness correlates with the range and likelihood of potential values. Contrary to Cummins et al. (2012), we argue that the numerical property of magnitude is more important in driving scalar implicature than roundness.

In a series of three experiments, 1,270 participants on Amazon Mechanical Turk were shown modified numerals of different degrees of roundness (93 = fine, 90 = medium, 100 = coarse), relative roundness differences (e.g., 100 vs. 110 and 1,000 vs. 1,100), absolute magnitudes (110 > 90, but both are similarly less round than 100), and orders of magnitude (<100, hundreds, thousands, tens of thousand). Experiment 1 looks at how speakers' world knowledge affects which range and likelihood they expect with particular modified numerals so that we can identify contexts for experimental items in experiment 2 and 3 where such expectations do not constrain scalar implicature. When speakers hold strong prior beliefs, the presence/absence of modified numerals has little influence on the likelihood distribution of potential values ($r = 0.89$); when these beliefs are weak, the presence of modified numerals does have an effect ($r = 0.18$). Experiment 2 looks at the effects of numerals' roundness and absolute magnitude in the same range as used by Cummins et al. (2012), i.e. 0 to 200, but with many more modified numerals. Experiment 3, in addition to looking at the effects of roundness and absolute magnitude, looks at the effect of order of magnitude by testing modified numerals in the thousands and tens of thousands. Items in experiment 2 and 3 were in the form of small conversations, e.g., "Brooke: I heard there's even a petition against the highway construction project. Oliver: [mod num] people already signed it. How many people do you think signed the petition? From ___ to ___, most likely ___." Linear mixed effects modeling (participant as random effect and roundness, absolute magnitude, and order of magnitude as fixed effects) suggests that numerals' magnitude affects range size and likelihood of potential values 12 to 14 times more than their roundness.

We conclude experimental participants employ one of two mechanisms to draw scalar implicatures depending on the informativity of context and the strength of prior beliefs: (i) In contexts with strong beliefs, speakers spell out their world knowledge. (ii) In contexts with weak beliefs, respondents estimate the likely true value is within a fixed distance from the modified numeral relative to its order of magnitude. This ratio is a Weber fraction (Fechner, 1860), and is in character and numerically similar to the acuity of the approximate number system (Dehaene, 2011), which acts as a bootstrap for scalar implicature in this case.

Breheny, R. (2008). A New Look at the Semantics and Pragmatics of Numerically Quantified Noun Phrases. *Journal of Semantics*, 25(2), 93-139.

Cummins, C., Sauerland, U., & Solt, S. (2012). Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy*, 35(2), 135-169.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. Oxford University Press, USA.

Fechner, G. T. (1860). *Elemente der Psychophysik*, vol. 2. Leipzig: Breitkopf und Härtel.

Thursday Poster.4

**CONNECTIVES AS PROCESSING CUES IN TRACKING INFORMATION SOURCE:
EVIDENCE FROM VISUAL WORLD PARADIGM**

Yipu Wei (Peking University), Pim Mak (Utrecht University),
Jacqueline Evers-Vermeul (Utrecht University), Ted Sanders (Utrecht University)
Contact email: weiyipu@pku.edu.cn

Background The processing of discourse involves not only parsing the linguistic input, but also constructing a mental representation, or situation model (Graesser, Millis, & Zwaan, 1997). In a situation model, language users keep track of the source of information. Information can be presented objectively, as originating from the real world, or subjectively, as originating from someone's mind. Causal relations can also be objective or subjective (Objective: *The factory has been polluting the water, so the local water supply is contaminated.* Subjective: *The factory has been polluting the water, so it has a very irresponsible owner.*) In establishing a subjective relation, the intentional mind involved in the reasoning is the *Subject of Consciousness* (SoC; Pander Maat & Sanders, 2001). Linguistic cues may encode subjectivity in the sense that they indicate whether an SoC is involved. For instance, certain connectives are prototypical for subjective relations (e.g. Chinese *kejian* 'so', Dutch *want* 'because' and *dus* 'so'), while others are specific for objective ones (e.g. Chinese *yin'er* 'as a result' and Dutch *daardoor* 'as a result').

In previous experiments, longer reading time was found for subjective relations compared to objective relations (Traxler, Sanford, Aked, & Moxey, 1997). Subjective connectives lead to an immediate processing delay (Canestrelli, Mak, & Sanders, 2013). Canestrelli et al. attribute the extra processing time to the fact that the mental representation of subjective information includes the SoC. This hypothesis is tested in this study.

Method We conducted a visual world paradigm eye-tracking experiment with the *EyeLink-1000*. Participants listened to sentences while they were presented with two scenes: one depicting the event being described in the first clause in the auditory input, and the other depicting a speaker (the SoC) talking about the event. We tested the effects of subjective and objective connectives in directing people's attention to the SoC in both Dutch and Chinese. In the Chinese experiment, we also included sentences with a connective that is underspecified for subjectivity (*suoyi* 'so', which can be used for both subjective and objective relations), as a baseline condition. Modal verbs as another kind of subjectivity markers were added to the second clause of subjective relations to test effects of subjective connectives in the later stages of processing.

Results & Conclusion In both languages, subjective connectives triggered an immediate increased attention to the SoC, compared to objective connectives before the rest of the second clause was given (Figure 1 & 2). Only when the subjectivity information was not expressed by the connective, modal verbs presented later in the sentence induced an increase in looks at the SoC. This focus on the SoC due to the linguistic cues can be explained as the tracking of the information source in the situation models, which continues throughout the sentence.

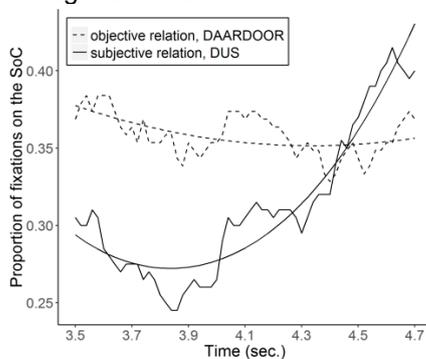


Figure 1. Proportion of fixations on the SoC during the connective time frame (Dutch experiment; 3.5s-4.7s)

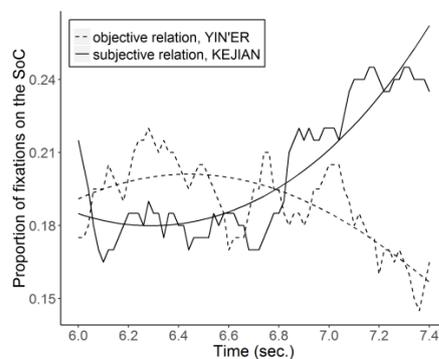


Figure 2. Proportion of fixations on the SoC during the connective time frame (Chinese experiment: 6.0s – 7.4s)

Thursday Poster.5

EXPRESSING SUBJECTIVITY IN DISCOURSE RELATIONS: EVIDENCE FROM COLLOCATIONAL ANALYSES

Yipu Wei (Peking University), Jacqueline Evers-Vermeul (Utrecht University),
Ted Sanders (Utrecht University)
Contact email: weiyipu@pku.edu.cn

Background & Research questions Connectives such as *because* and *so* are considered processing instructors in discourse (Britton, 1994; Mak & Sanders, 2013): they provide information on the type of coherence relation involved (e.g. temporal, adversative or causal), and in several languages they also code information on subjectivity, i.e. the involvement of a locutionary agent (Finegan, 1995). For example, the Dutch connectives *want* 'because' and *dus* 'so' (Spooren, Sanders, Huiskes, & Degand, 2010) and Mandarin Chinese *kejian* 'so' (Li, Evers-Vermeul, & Sanders, 2013) prototypically express subjective relations.

On-line processing studies suggest that the processing effects of connectives are interfered by the presence of perspective markers such as *John thinks*, *perhaps* and *according to Peter* (Canestrelli, Mak, & Sanders, 2013). These perspective markers all relate to the epistemic stance of the utterance, indicating – just like subjective connectives – that someone's mind is involved in the construction of the coherence relation. However, perspective marking is not restricted to epistemic stance; two other dimensions can be distinguished: attitudinal stance and style stance (Conrad & Biber, 2000).

Collocational analysis can advance our knowledge on the properties of a discourse marker on the basis of its contextual features. This corpus-based study investigates the following research questions: Do connectives marking different subjectivity degrees differ in the types of collocates? Specifically, do connectives differ in the types of perspective markers they co-occur with? Do the collocation patterns differ across various contexts and genres?

Method We focused on two Chinese causal connectives: the specific subjective *kejian* 'so', and a connective that is underspecified in terms of subjectivity and can be used in both objective and subjective relations: *suoyi* 'so' (Li et al., 2013). A distinctive-collocate analysis was performed by measuring the association strength between these connectives and other discourse elements. We retrieved data from the CCL corpus, a large, balanced Modern Chinese written corpus. Association scores (G^2 and Delta-P) were calculated based on contingency tables of observed and expected frequencies (Evert, 2008; Gries, 2013). The collocation patterns were investigated and compared in different contexts (global vs local; preceding vs following) and genres (narrative vs non-narrative).

Results & Conclusion The collocational analyses revealed an interpretable pattern in connective use in combination with perspective markers in discourse. The general connective *suoyi* prefers contexts with perspective markers expressing epistemic stance: cognition verbs (*think*, *know*), communication verbs (*say*, *tell*) and modal verbs (*should*, *may*). *Kejian*, the specific subjective connective, co-occurred more often with perspective markers related to attitudinal stance, such as markers of expectedness (*surprisingly*, *unexpectedly*) and importance (*important*). The collocation patterns were consistent across different contexts and genres with small variations. The findings show the efficiency of language use in terms of expressing subjectivity in discourse: when a specific subjective connective is used instead of a general one, it is less likely to have perspective markers of epistemic stance in the context. Moreover, the collocation patterns found in the corpus study echo to the processing effects of connectives and perspective markers expressing subjectivity.

Thursday Poster.6

**WORD ORDER BEATS MORPHOSYNTAX: INCREMENTAL THEMATIC ROLE
ASSIGNMENT IN TAGALOG L1 ACQUISITION**

Rowena Garcia (IDEALAB, University of Potsdam), Jens Roeser (Nottingham Trent University) & Barbara Höhle (University of Potsdam)
rgarcia@uni-potsdam.de

Previous studies have demonstrated that children tend to assign the first noun in the sentence as the agent, and they have difficulties revising this initial interpretation once a passive marker is encountered (Abbot-Smith, Chang, Rowland, & Pine, 2017; Huang, Zheng, Meng, & Snedeker, 2013). However, it has not yet been investigated whether children would still use a word order strategy if the morphosyntactic markers are given early in the sentence, such as in a language like Tagalog. In this language, the canonical verb-initial order and voice-marking system disambiguate thematic roles even before the nouns are given. The agent voice verb infix *-um-* assigns the noun phrase marked by *ang* the agent role, while the patient voice verb infix *-in-* designates the *ang*-phrase as the patient.

We first looked at the input frequency of voice and word order in Tagalog child-directed speech (Study 1) in order to determine the basis of a possible word order strategy. Based on Marzan's (2013) corpus of child-directed speech to 3 Tagalog-speaking children (ages 2;4-2;7), 53% of utterances with causative transitive verbs and at least one noun had verbs which were inflected for the patient voice, while 21% were in the agent voice. Utterances in both voices were predominantly agent-initial or contained only an agent. We also gave a sentence completion task to five- (5-yo) and seven-year-old (7-yo) children and adults (Study 2). They were asked to describe pictures (e.g., a cow pulling a pig) by completing sentences which start with voice-marked verbs. Adults showed an agent-initial preference only in the patient voice (agent voice: 53% agent-initial productions, patient voice: 98%). However, children coded the agent as the first noun phrase regardless of voice (five-year-olds – agent voice: 88%, patient voice: 81%; seven-year-old – both voices: 91%). Based on these results, we predicted that children's word order strategy would be to interpret the first noun as the agent.

In Study 3, we gave Tagalog-speaking adults and children (5-yo, 7-yo) a picture selection task while their looks to the computer screen were tracked. They had to choose which of two pictures (e.g., a cow pulling a pig, and a pig pulling a cow) matched the sentence they heard. We crossed voice (agent voice, patient voice) and word order (agent-initial, patient-initial) to create the stimuli sentences. Results showed that in the patient-initial condition, children were more accurate in the patient voice than in the agent voice. Moreover, in the agent voice patient-initial condition, 5-yo scored below chance level, implying that they interpreted the first mentioned noun as the agent. The eye-tracking data revealed that after the first noun phrase, adults looked to the target in all conditions; the 7-yo showed more looks to the target in agent-initial than patient-initial; while the 5-yo did not show any preference. After the second noun phrase, children showed looks to the target except in the agent voice patient-initial condition. The results indicate that children needed more time to interpret patient-initial sentences. Furthermore, they did not always use the morphosyntactic markers in the agent voice.

In conclusion, Tagalog-speaking children rely on word order heuristics even if the morphosyntactic markers are provided early on in the sentence. Moreover, the eye-tracking data showed that the need to overcome a first-noun-as-agent bias resulted to a processing delay in children's interpretation of patient-initial sentences. Aside from an agent-initial preference, we also observed a patient voice advantage which can be attributed to the higher frequency of patient voice in the input, providing children with more exposure to the mapping of the verbal marker *-in-* to the nominal *ang*-phrase. Overall, children's processing strategies for thematic role assignment seem to be largely dependent on the frequency patterns in their input.

Thursday Poster.7

THE AMBIGUITY DISADVANTAGE EFFECT IN WORD COMPREHENSION: RT AND EEG EVIDENCE FOR SEMANTIC COMPETITION

Greg Maciejewski (University of Leeds) & Ekaterini Klepousniotou (University of Leeds)
psgm@leeds.ac.uk

Research has shown that semantic ambiguity slows word comprehension (Duffy et al., 1988). However, it is not clear whether this ambiguity disadvantage effect is due to competition between multiple meanings during semantic activation (e.g., Rodd et al., 2004) or task-specific decision making during response selection (e.g., Pexman et al., 2004). Here, we examined these accounts in two studies exploring the effects of homonymy (i.e., the ambiguity between unrelated word meanings). Participants made semantic relatedness decisions to homonyms with balanced (e.g., “fan”) or unbalanced meaning frequencies (e.g., “pen”) and non-homonyms that were followed by related or unrelated targets.

Experiment 1 revealed a large disadvantage effect for balanced but not unbalanced homonymy, suggesting that the effect might be semantic in nature as it is sensitive to meaning frequency and its impact on semantic activation (Simpson & Burgess, 1985). Furthermore, the effect appeared on related trials (e.g., “fan-rotate”) where the different meanings triggered conflicting responses to the target as well as on unrelated trials (e.g., “fan-snake”) where the meanings triggered a single response. Crucially, these findings are inconsistent with decision-making accounts (Pexman et al., 2004) which suggest that ambiguity slows comprehension only when the meanings of ambiguous words create a conflict during response selection.

Consistent with this interpretation, Experiment 2, involving EEG recording, showed that the processing of balanced but not unbalanced homonyms was associated with larger N400 amplitudes indicating slower and more effortful semantic processing (Kutas & Federmeier, 2011). This further clarifies the locus of the disadvantage effect in relatedness decision tasks. The effect arises due to semantic competition, rather than decision making during the presentation of the target (e.g., response-conflict resolution or checking each meaning of the ambiguous word against the target).

Taken together, the findings provide strong support for models of word processing that predict a processing cost ensuing from multiple form-to-meaning mappings (e.g., Rodd et al., 2004). Ambiguous words slow comprehension (in isolation or neutral context) because their multiple semantic representations compete for full activation, unless meaning frequency facilitates rapid activation of one of the word meanings. In contrast, our findings present a serious challenge for alternative accounts (e.g., Pexman et al., 2004) which claim that the ambiguity disadvantage effect is due to decision making.

References

- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27(4), 429-446.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.
- Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1252-1270.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, 28(1), 89-104.
- Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 28-39.

Thursday Poster.8

LANGUAGE PLANNING IN WRITING RESEMBLES PLANNING IN SPEECH

Jens Roeser, Mark Torrance & Thom Baguley (Nottingham Trent University)

jens.roeser@ntu.ac.uk

Even for short sentences speakers typically mentally prepare less than the full sentence before starting to speak. The extent of pre-planning, and particularly whether or not it extends beyond a sentence-initial noun is determined by content and structure of some minimal linguistic unit (Bock & Ferreira, 2014). The scope of advance planning, however, is also subject to non-linguistic factors (Wagner et al., 2010). One possible source of variation is output modality. Existing research on sentence planning is solely based on spoken data. It is possible, therefore, that planning beyond the sentence-initial noun might result from the need to satisfy modality-specific constraints. Speech and writing differ in fluency demands – long pauses/frequent corrections are permitted in writing but not speech; written output is typically slower than speech suggesting increased demands for buffering planned items prior to output. These factors suggest a push towards shorter planning scope in written production, and less overlap between subsequent planning units. It is possible, therefore, that existing findings relating to planning in sentence production are influenced by modality-specific factors and do not solely reflect fundamental features of the language system.

We address two questions about advance planning of short sentences in both writing and speech. Specifically (1) can advance grammatical encoding of coordinated NPs (*A and the B*), as frequently reported for speech (see Martin et al., 2014), be explained by modality-specific requirements, and (2) for both spoken and written production is the extent of advance planning determined by syntactic or conceptual processing of the utterance/text that is to be produced.

In two series of experiments we elicited short sentences in keyboard typing and speech in response to arrays of simple line drawings shown on the computer screen. We recorded the time required to initiate the sentence and eye movements to the depicted referents of the utterance.

In the first series of 3 experiments ($N_s=32$), we manipulated structure of elicited sentences: Target sentences started with either a coordinated NP (e.g. *Peter and the cake moved above the hat*) or a simple noun (e.g. *Peter moved above the cake and the hat*). Both in writing and speech we found fixations on the referent of the second noun before onset in coordinated NPs associated with longer onset latencies.

In the second series of 2 experiments ($N=32, 64$), we tested the hypothesis that advance planning scope is determined at a conceptual level, before grammatical encoding. We manipulated the semantic structure of the underlying message of the target phrase but kept syntactic and lexical properties constant. Specifically in target phrases such as *The painting with the man* we manipulated the stimulus array such that either the first or the second noun was contrastive (e.g. *The painting with the man* [not the dog with the man], vs. *The painting with the man* [not the painting with the child]). In both writing and speech we found earlier fixations to the referent image of the second noun if it served a contrasting function.

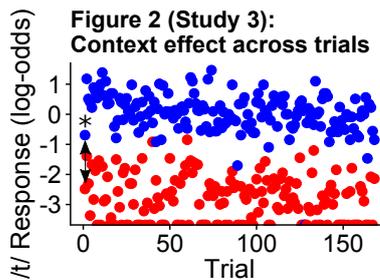
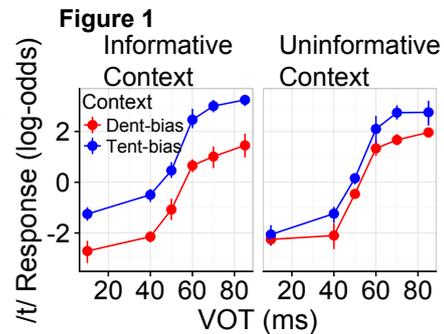
The most parsimonious interpretation of these results suggest two main conclusions: (1) Previous findings on advance planning of simple sentences generalise to writing and so are likely to result from a general feature of the language system and not modality-specific requirements of spoken production. (2) Conceptual contrasts within the message of a target phrase affect planning scope even when surface features of the phrase are held constant. This suggests that planning scope is (in at least some contexts) dependent on semantic and not grammatical features of the target. Again this holds true for both spoken and written output.

EXPECTED UTILITY OF LATER CONTEXT MEDIATES MAINTENANCE OF SUBCATEGORICAL INFORMATION

Wednesday Bushong (wbushong@ur.rochester.edu) & T. Florian Jaeger (fjaeger@ur.rochester.edu)

Spoken language understanding involves the integration of auditory cues with lexical context, including *later* context. Optimal integration with later context requires maintaining relevant information about auditory cues. Indeed, such integration has been found in previous work [1,2]. But, memory limits prevent listeners from maintaining such information indefinitely. We hypothesize that the degree to which listeners maintain subcategorical information is guided by its *expected utility*: the more listeners expect the interpretation of subcategorical information to benefit from later context, the more likely they are to maintain that information in memory. Here, we test i) whether listeners can maintain subcategorical information, ii) if maintenance is the default, and iii) the informativity of later context in recent experience changes the degree to which listeners maintain subcategorical information.

Methods. 4 web-based studies (MTurk, N=39-117 subjects/study) have participants listen to sentences (e.g., Table 1, top) and respond whether they heard *tent* or *dent*. We manipulate both voice-onset time (VOT) to cover a continuum of /t/-/d/, and later context to bias towards either “tent” (e.g., “campground”) or “dent” (e.g., “fender”; top of Table 1). **Studies 1 and 2** reduce the massive amount of repetition and unnaturally high levels of cue conflict common in previous work [1,2]. These changes avoid inducing experiment-specific strategies. With these problems removed, we actually find significantly *larger* effects of subcategorical maintenance than previous work. This is expected if listeners typically maintain subcategorical information in memory during everyday language processing. **Study 3** tests this interpretation further. Using data from Studies 1-2 and an additional replication, we analyze whether the maintenance effect is present from the start of the experiment or only emerges over time. In all cases, we see a strong effect on the very first trial (e.g. Fig 2). This further suggests that subcategorical maintenance is typical of language use.



We hypothesize that this is due to typically high informativity of later context in natural language. **Study 4** tests this, asking whether maintenance of subcategorical information can be reduced if listeners expect upcoming context to be uninformative about the interpretation of earlier speech input. Subjects are divided into two exposure groups (Table 1). For the *Informative Context Group*, later context is *always* informative. For the *Uninformative Context Group*, it *never* is. In the later test phase, all subjects hear sentences with informative later context. We find that both groups maintained information in the test phase ($p < 0.001$).

Critically, the effect is reduced in the Uninformative Context group ($p = 0.008$; Fig 1). **Conclusions.** We find that listeners maintain subcategorical information about the speech input in memory for integration with later context. This seems to be a default strategy, in line with the typically high informativity of later context in natural language use. When this informativity is removed, listeners reduce information maintenance. Together, these results suggest that listeners use the statistics of the input to guide strategies for memory allocation during real-time processing. [1] Brown-Schmidt & Toscano (2017) *LCN* [2] Connine et al. (1991)

Group	Exposure Phase	Test Phase
Informative	...[t/d]ent was noticed in the campground/fender...	...[t/d]ent in the forest...
Uninformative	...[t/d]ent was noticed...	...[t/d]ent in the forest...

WHAT IS “GOOD-ENOUGH” ABOUT THE PROCESSING OF GARDEN-PATH SENTENCES IN CZECH

Jan Chromý (Charles University, Prague)
jan.chromy@ff.cuni.cz

The paper is concerned with the idea of Good-Enough processing of garden-path sentences (GPs) which was first presented by Christianson et al. (2001). In that study, participants were shown GPs like *While Anna dressed the baby that was small and cute spit up on the bed* and were asked questions like *Did Anna dress the baby?* 65.6% participants answered the question incorrectly whereas only 12.5% participants gave the wrong answer on the same question after a control sentence with a different clause-order (*The baby that was small and cute spit up on the bed while Anna dressed*). The authors argue that the participants create a “good-enough” representation of these sentences.

The paper aims to analyze what is “good-enough” on these representations. I ran three self-paced reading experiments using GPs in Czech (1a) and control non-GPs (1b).

(1a) Kluci honili psa a kočku v podkroví
Kid-NOM.M.PL chase-3PL.M.PST dog-ACC.M.SG and cat-ACC.F.SG in attic
znepokojovali šediví hlodavci.
worry-3PL.M.PST grey-NOM.M.PL rodents-NOM.M.PL

‘Kids chased a dog and a cat in the attic was worried by grey rodents.’

(1b) Kluci honili psa a kočku v podkroví
Kid-NOM.M.PL chase-3PL.M.PST dog-ACC.M.SG and cat-NOM.F.SG in attic
znepokojovala šediví hlodavci.
worry-3SG.F.PST grey-ACC.M.PL rodents-ACC.M.PL

‘Kids chased a dog, and a cat in the attic worried grey rodents.’

These sentences differed so that a garden-path effect was possible in (1a) (noun *kočku* could have been analyzed as an object of verb *honili* at first) but not in (1b) (noun *kočka* is a nominative and hence it cannot be an object in Czech – it is a subject of the second clause).

In Exp1, I used the moving window (word-by-word) technique. After reading each sentence, I asked either a question (2a) *Did the kids chase a cat?* or (2b) *Did the rodents worry the cat?* (2a) asked whether the original GP interpretation lingered and (2b) asked whether participants formed a correct interpretation of the second clause. The analysis of RTs showed a clear GP effect in (1a) since the RTs were significantly longer on the last three segments in (1a) than in (1b). Similarly to Christianson et al. (2001), participants answered questions (2a) incorrectly significantly more after reading sentences (1a) than after reading (1b) (34.6% vs 7.7%). However, participants responded questions (2b) incorrectly more often after the GPs than after control sentences (21.9% vs 10.9%). The analysis showed no difference in RTs between correctly and incorrectly answered GPs; there is thus no evidence for heuristics (“fast and frugal processing”) which is a key feature of Good-Enough processing.

Exp2 replicated all the findings of Exp1 using a self-paced reading with sentences presented as a whole. Exp3 was a replication of Exp1, but two more questions were used: (2c) *Did the kids chase a dog?* (asking on the correct analysis of the main clause) and (2d) *Did the rodents worry the dog?* (asking on an analysis which should not emerge during reading of the sentence). Exp3 replicated the findings of Exp1 concerning RTs and answers on questions (2a) and (2b). Moreover, questions (2d) were answered incorrectly significantly more after GPs than after non-GPs (26.9% vs 7.9%) which suggest a more general problem of processing GPs.

The results indicate that the processing of GPs is often incoherent rather than Good-Enough. We found no evidence for heuristics in RTs. Also, higher rate of incorrect answers on (2b) and (2d) after GPs suggests that the participants often did not form a coherent representation, and that they just tried to answer the question based on the scarce information they actually retrieved.

References: Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42(4), 368-407.

LEXICAL ACCESS ON BEHALF OF TASK PARTNER: ELECTROPHYSIOLOGICAL INSIGHTS FROM JOINT PICTURE NAMING

Anna K. Kuhlen (Humboldt Universität zu Berlin) & Rasha Abdel Rahman (Humboldt Universität zu Berlin)
anna.kuhlen@hu-berlin.de

When naming a sequence of pictures, naming latencies increase with each new picture of a given semantic category, so-called cumulative semantic interference (e.g., Howard et al., 2006). Recently, it has been demonstrated that naming latencies not only increase in response to speakers' own prior naming of pictures, but also in response to their task partner naming pictures (Hoedemaker, Ernst, Meyer, & Belke, 2017; Kuhlen & Abdel Rahman, 2017). Here we investigate the electrophysiological underpinnings of this effect. EEG was recorded from 30 participants who believed to be naming pictures together with a remotely located task partner. We observed an increased posterior positivity around 250-400ms, which corresponds to an increase in naming latencies and has been taken to reflect lexical selection (Costa et al., 2009). Unlike previous studies, only a subset of subjects showed partner-elicited interference. Crucially, this group of subjects showed a stronger increase in posterior positivity when semantic categories were co-named with the partner (vs. named alone), and only these subjects showed a similar posterior positivity when the partner named a picture (vs. when nobody named it). This suggests that these subjects simulated lexical access on behalf of their partner. In conclusion, our study connects partner-elicited cumulative semantic interference to electrophysiological underpinnings, yielding promising insights into the processes of language production in social settings.

References

- Costa, A., Strijkers, K., Martin, Cl. & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *PNAS*, *106*, 21442–21446.
- Hoedemaker, R. S., Ernst, J., Meyer, A. S., & Belke, E. (2017). Language production in a shared task: Cumulative Semantic Interference from self- and other-produced context words. *Acta Psychologica*, *172*, 55–63. <https://doi.org/10.1016/j.actpsy.2016.11.007>
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*, 464–482. doi.org/10.1016/j.cognition.2005.02.006
- Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, *166*, 94-106. [doi: 10.1016/j.cognition.2017.05.024](https://doi.org/10.1016/j.cognition.2017.05.024).

PRIOR LEARNING OF ACOUSTIC CUES BLOCKS LEARNING OF NEW CUES IN NON-NATIVE SPEECH ACQUISITION

Jessie S. Nixon

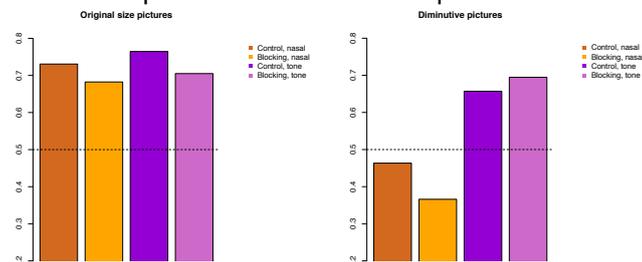
Quantitative Linguistics, University of Tübingen

How do listeners learn the sounds of their language(s)? Half a century ago, Kamin (1968) demonstrated that, if an animal learns that a cue (e.g. light) predicts an outcome (electric shock), then is later presented with a double cue (light+tone), learning of the second cue (tone) can be diminished, or 'blocked', by the first. Cues are learned not simply due to co-occurrence, but through predicting important outcomes (Rescorla, 1988) and through prediction error and unlearning (Ramscar, Dye, & McCauley, 2013). Most (statistical) models of speech acquisition and perception do not account for effects of blocking and cue competition. The present study investigates whether learning non-native speech cues can be blocked by prior knowledge of acoustic cues.

In an online game, 191 native English speakers heard Southern Min words and saw original size and diminutive pictures of 'alien' objects. Words were mid-level or high-level tone with a nasal or oral vowel. (Baseline: mid-tone oral; original size). The experiment had three phases: pre-training, training and test. Participants either received *blocking pre-training* or *control pre-training*. In the blocking pre-training, participants heard a single cue (e.g. high-tone words). In the control pre-training, participants heard unrelated cues, not presented in training or test. Training was the same for all participants: two cues, high-tone+nasal, signalled the diminutive. During test, the cue that was *not* pre-trained (e.g. nasal words) was tested. Accuracy was expected to be lower after blocking, compared to control pre-training.

Results are shown in Figure 1. A glmer model showed accuracy was higher for baseline, original-size pictures than diminutives. There was no effect of condition for original-size pictures, probably because they were baseline, the default if the additional cues were not perceived. However, for diminutives, accuracy was significantly lower after blocking pre-training, compared to control pre-training, for the nasal cue. Because tone was sufficient for predicting the size outcome in pre-training, there was little uncertainty left to drive learning of the nasal cue. Learning of nasality was *blocked*. There was no effect of condition for the tone cue. Because the nasal cue was *not* learned completely during pre-training, there was sufficient uncertainty to drive further learning in training. Therefore, learning of the tone cue was not blocked.

The present study replicates the Kamin (1968) 'blocking effect' in speech acquisition. Learning new cues can be blocked by already-learned cues. These results have important implications for theories of speech acquisition. Firstly, acoustic knowledge may not always directly reflect statistical structure of language input, but will depend on the predictive structure of learning events. Secondly, these results may help explain transfer effects in second language acquisition. The results also have implications for order of acquisition in first language acquisition.



Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–31).

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mice in adult speech. *Language*, 89(4), 760–793.

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151.

NOVEL LABELS INCREASE CATEGORY COHERENCE, BUT ONLY IN COORDINATIVE CONTEXTS

Ellise Suffill, Psychology, PPLS, University of Edinburgh
Holly P. Branigan, Psychology, PPLS, University of Edinburgh
Martin J. Pickering, Psychology, PPLS, University of Edinburgh
Contact: e.suffill@ed.ac.uk

Labels not only reflect the way that we categorize objects, but also allow us to communicate and share those categories. Labels are thus closely tied to communication. People develop more similar categories when they use novel labels to categorize, than when they do not use labels, in paired categorization tasks [4,5]. Novel labels might affect categories by influencing people select fewer, more abstractable dimensions for categorization that tend to be coherent across individuals [2]. However, in [4] & [5] participants performed a categorization task with a partner. Hence we do not know whether the category coherence yielded by novel labels is a direct effect of labelling, or specifically an effect of the need for coordination. We therefore investigated the effects of novel labels on category coherence with and without a coordinative context.

In two experiments ($N=200$), participants individually sorted images of mountains into two groups over three rounds. Participants either sorted items into groups without labels, or by using two novel non-word labels (with each participant using different labels). If category coherence is a general consequence of using labels, then people who used labels should show greater category coherence than people who did not, regardless of context. But if the labelling effect is specific to situations in which people must coordinate, then it should occur only in coordinative contexts. We therefore manipulated the context in which participants carried out the sorting task: in Experiment 1, participants were instructed to sort groups of items in a way that made sense to them alone; in Experiment 2, participants were informed that they would be paired (but not interact) with a partner, to whose categories their own categories would be compared.

We used an adapted Cultural Consensus analysis [3] to examine group category coherence across all participants within a condition (i.e., labels by coordinative context). Experiment 1 showed lower category coherence for people who sorted with labels, compared to those who sorted without labels ($\beta = -0.05$, $SE = 0.03$, $t = -2.12$). But Experiment 2 showed the reverse, with greater category coherence for participants who sorted using labels, than those who sorted without labels ($\beta = 0.18$, $SE = 0.03$, $t = 6.60$). These results provide further evidence that novel labels serve to direct people's attention to more abstractable dimensions of objects, thus increasing the potential for their categories to overlap (i.e., be coherent across individuals) [2]. But crucially they also suggest that novel labels do so specifically in contexts where labelling is coordinative and the need for coordination is foregrounded. We suggest that in coordinative contexts specifically, labels serve as a device for the coordination of people's categories [1].

References

- [1] Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- [2] Lupyan, G. (2008). From chair to "chair": A representational shift account of object labeling effects on memory. *J.Exp.Psychol: General*, 137, 348-396.
- [3] Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *J.Mem.Lang.*, 40, 230-262.
- [4] Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *J.Exp.Psychol: General*, 127, 331.
- [5] Suffill, E., Branigan, H. P., and Pickering, M. J. (2016). *When the words don't matter: arbitrary labels improve categorical alignment through the anchoring of categories*. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.). Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Thursday Poster.14

THE REALITY OF HIERARCHICAL MORPHOLOGICAL STRUCTURE IN MULTIMORPHEMIC WORDS

Yoonsang Song (Chinese University of Hong Kong), Youngah Do (University of Hong Kong), Jongbong Lee (Michigan State University), Arthur Thompson (University of Hong Kong) & Eileen Waegemaekers (University of Hong Kong)
youngah@hku.hk

This study is one of the first to empirically test the long-held assumption that individual morphemes of multimorphemic words are represented in hierarchical structure. Traditionally, it has been assumed, in generative morphology, that individual morphemes of multimorphemic words are arranged in a hierarchical fashion which exhibits binary branching, rather than linear sequence (e.g., Lieber, 1980), as depicted in Figure 1.



Figure 1. Hierarchical structure of the English derived words *unkindness* (left-branching) and *unreadable* (right-branching).

The current literature, however, provides no strong empirical evidence for the psychological reality of this hierarchical morphological structure.

In order to address this issue, we conducted a cross-modal (audio-visual) priming experiment with 29 native speakers of English. This study has two experimental conditions, namely, (1) left-branching and (2) right-branching condition. In the left-branching and right-branching condition, the targets are left-branching and right-branching trimorphemic words in English, respectively (e.g., [[re-sell]-able] and [un-[read-able]]). Each left-branching target has three types of primes: (a) constituent, (b) nonconstituent, and (c) unrelated prime. The constituent prime is the first bimorphemic substring of the target, which serves as a morphological constituent of the target (e.g., *resell*); the nonconstituent prime are the last bimorphemic substrings of the target (e.g., *sellable*), which does not serve as a constituent of the target. The unrelated prime is an affixed bimorphemic word unrelated to the target (e.g., *prolong*). Because no right-branching trimorphemic words possess a bimorphemic substring which constitutes a legitimate English word, only two types of primes were allocated to right-branching words: (a) constituent (e.g., *readable*) and (b) unrelated (e.g., *ticklish*). The right-branching condition was included to examine how independently of linear position constituent priming occurs. The different types of primes in each branching condition were matched in frequency and length. The targets in the two branching conditions were matched in frequency and length.

The results of our experiment support the idea that individual morphemes in multimorphemic words are nested within a hierarchical structure: Recognition of trimorphemic words (e.g., *resellable* or [[re-[sell]-able]]) was significantly facilitated by prior processing of their morphological constituents (e.g., *resell*), but not by that of their nonconstituents (e.g., *sellable*). In addition, morphological structural priming occurred independently of linear positions of the morphological constituents. Specifically, in left-branching condition, the participants' lexical decision on trimorphemic targets was significantly facilitated in terms of processing time (87 ms) and accuracy (12%) when the targets were preceded by the constituent prime, as compared with when preceded by the unrelated prime. In the right-branching condition, significant priming effects were observed in terms of processing time (77 ms), but not in terms of accuracy. Such priming effects were, however, not observed with the nonconstituent primes, although they are highly related to the targets formally, semantically, and morphologically. Taken together, the results of this study lend support to the idea that individual morphemes in multimorphemic words are nested with a hierarchical structure.

Thursday Poster.15

PEOPLE WITH SMALLER SOCIAL NETWORKS ARE BETTER AT TALKER IDENTIFICATION

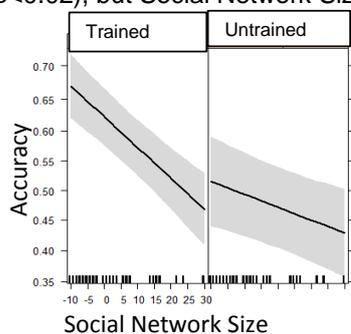
Shiri Lev-Ari (Royal Holloway, University of London)

Shiri.LevAri@rhul.ac.uk

Language processing is directed by our past experience such that we are more likely to attend to features that proved to be informative in the past than to ones that have not. Our social experience might similarly influence what type of information is useful by shaping our communicative needs. In particular, whereas people with smaller social networks might benefit from encoding the specifics of the speech of each person they interact with, such learning might be more demanding yet less useful for people with larger social networks, who interact with more speakers yet spend less time with each. The studies reported here therefore test whether individuals' social network size influences the identification of novel talkers (Exp. 1) and whether the effect is due to differences in encoding of acoustic or linguistic details (Exp. 2).

In Experiment 1 62 native Dutch speakers reported the people they regularly talk to in a week (i.e., network size). Then they performed a Talker Identification task, in which they were trained to identify 4 native Dutch speakers uttering each the same 6 monosyllabic words in a four-alternative forced choice task. Participants' training continued until they reached 80% accuracy or until 10 rounds have passed (240 training trials). They were then tested on their ability to identify the 4 talkers uttering the 6 trained words and 6 new monosyllabic words containing the same vowels as the trained words. Results showed that talkers were more easily identified when they uttered the trained words than the untrained words ($\beta=0.53$, $SE=0.13$, $z=4.13$, $p<0.001$). Having a larger social network, however, led to *worse* performance at the reference level of trained words ($\beta=-0.02$, $SE=0.01$, $z=-3.02$, $p<0.01$). While the interaction of Network Size and Word Type did not reach conventional level of significance ($\beta=0.012$, $SE=0.006$, $z=1.883$, $p<0.06$; See Figure), the effect of Social Network Size was smaller and n.s. for the untrained words ($\beta=-0.01$, $SE=0.01$, $z=-1.32$, $p=0.19$). Experiment 1 thus shows that people with smaller social networks are better at learning to identify new talkers, although the benefit might be confined to trained words, and thus might relate to non-generalizable aspects of production.

The results of Experiment 1 raise the question of what type of information individuals with smaller social networks rely on to achieve the better performance, and whether it is acoustic or linguistic in nature, as both types of information are used for talker identification (Perrachione, Del Tufo & Gabrieli, 2011). Therefore, in Experiment 2, 100 native English speakers who do not speak Dutch performed the Talker Identification task from Exp. 1. Results showed that talker identification was better for trained words than for untrained words ($\beta=0.26$, $SE=0.11$, $z=2.48$, $p<0.02$), but Social Network Size did not influence performance despite having sufficient



variation (range: 1-52, $M=14$, $SD=9$). These results suggest that the added benefit of having a small social network in Experiment 1 is due to additional encoding of speaker-specific *linguistic* information.

The results provide evidence that social experience can influence which type of information is encoded during language processing, and can thus influence later linguistic performance, such as identification of speakers. The studies thus show how language processing adjusts to our social experience and needs.

Semantic Bias and Topicality in Pronoun Resolution

Yvonne Portele & Markus Bader (Goethe University Frankfurt)

The interpretation of p(ersonal) pronouns is subject to structural and semantic biases. [3] showed how interpretation results can be predicted when the two types of biases are estimated from production data and then combined by Bayes rules. We present two experiments that provide a quantitative test of this model with regard to d(emonstrative) pronouns in German, which are subject to different structural biases like their counterparts in Finnish [2].

Experiment 1 investigated short contexts (see (1)), with an object experiencer verb in the last sentence. A continuation prompt contained a pronoun (*er* 'he'/p-pronoun or *der* 'he'/d-pronoun) and a discourse marker (*deshalb* 'therefore' or *nämlich* 'the reason was that').

- (1) a. Gestern Abend wurde eine Talkshow aufgezeichnet. In der Runde saß auch *ein angesehener Experte*. Ein vorlauter Studiogast hat *den Experten* mehrfach irritiert.
'Yesterday evening, a talkshow was recorded. There was a distinguished expert among the participants. A cheeky studio guest irritated the expert several times.'
- b. Continuation prompt: **Er/Der** _____ nämlich (cause) / deshalb (consequence) _____

In accordance with the literature, the p-pronoun showed a preference for the stimulus with a cause relation and a preference for the experiencer with a consequence relation (see Figure 1). Unlike most prior experiments, which found complementary preferences for p- and d-pronouns with SO sentences, the d-pronoun showed the very same preferences. Thus, semantic bias governs the interpretation of p- and d-pronouns in the same way ([1]).

Experiment 2 manipulated the position of the topic in the final context sentence by varying the referent introduced in the preceding sentence.

- (2) Gestern Abend wurde eine Talkshow aufgezeichnet. Im Publikum saß auch *ein vorlauter Studiogast*. *Der Studiogast* hat einen angesehenen Experten mehrfach irritiert.
'Yesterday evening, a talkshow was recorded. There was a cheeky studio guest in the audience. The studio guest irritated a distinguished expert several times.'

The continuation prompt always contained the causal discourse marker *nämlich*. For the p-pronoun, the results show a strong preference to refer to the subject/stimulus, independent of the topic's position (see Figure 1). The d-pronoun also preferred reference to the first NP, but the preference for the subject/stimulus was much stronger when it was not the topic, in agreement with the non-topic orientation of d-pronouns.

In sum, the interpretation of p-pronouns was affected by the coherence relation established by the discourse marker but not by topicality. Like p-pronouns, d-pronouns showed a preference for the semantically most expected antecedent but simultaneously showed an anti-topic effect. For both experiments we have also obtained production data frequencies (continuations without a prompt) in order to test the Bayesian theory of [3]. Overall, we found a good fit between data and theory. We propose a purely production-based notion of prominence to capture the difference between p- and d-pronouns with regard to structural bias.

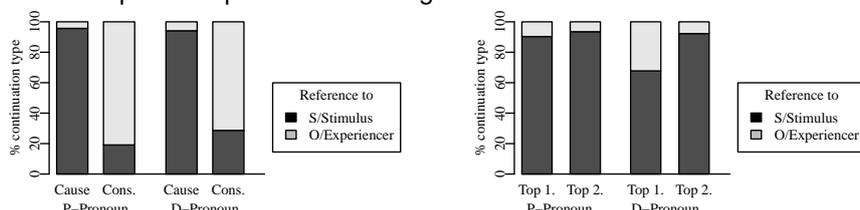


Figure 1: References to sub./stimulus and obj./experiencer for Experiments 1 (left) and 2 (right).

- [1] Juhani Järvi­kivi, Roger PG van Gompel, and Jukka Hyö­nä. The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of psycholinguistic research*, 46(3):525–550, 2017.
- [2] Elsi Kaiser and John C. Trueswell. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, 23(5):709–748, 2008.
- [3] Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L. Elman. Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44, 2008.

Thursday Poster.17

FEATURES MATTER IN COMPUTATIONAL MODELS OF WORD READING

Stéphan Tulkens, Dominiek Sandra & Walter Daelemans (CLiPS, University of Antwerp)
firstname.lastname@uantwerpen.be

Computational models of word reading require that their input words be turned into numbers for processing, which, in turn, requires a featurization procedure that turns strings into numerical vectors. The choice of featurization procedure can have a large impact on the practical and theoretical applicability of the model, as the way words are represented, be it internally or externally, restricts the inferences a model can make.

In Artificial Intelligence or Machine Learning methodology, selection of a feature set is part of the fitting process of a model, and is rarely done on the basis of *a priori* reasoning. In computational psycholinguistics this is usually not the case, and features are chosen using *a priori* reasoning or experimental evidence, without examining the effect this choice has on model fit. We suggest that adopting a machine learning methodology can improve findings from computational models. By testing the same model architecture with a variety of feature sets, the interaction between models and features can be partially disentangled, resulting in more robust estimates of model performance.

This is made all the more important because recent models of word reading have assigned greater importance to the “orthographic code”, which is sometimes seen as a feature set, and sometimes as a mechanism in the model.

To show that the choice of features matters, we conducted a neighborhood experiment. We implemented eight orthographic featurization techniques used in previous models of word reading, including slot-based, open bigram, sub-orthographic, and wickelgraph schemes. In analogy to OLD20 (Yarkoni et al. 2008), we then defined RD20 as the mean cosine distance to the 20 closest neighbors for a given feature set. We calculated RD20 for all eight feature sets on a large selection of words from 3 lexical databases: the Dutch and English Subtlex databases (Keuleers et al. 2010; Van Heuven et al. 2014), and Lexique (New et al. 2001). We then correlated these distances with mean RT measurements from the various lexicon projects in these languages. This allows us to directly establish whether there are differences between feature sets, without involving an explicit theory of the operations performed on these feature sets.

For each of the three languages, we bootstrapped differences between the feature sets and OLD20 over 10,000 samples to determine 95% Confidence Intervals. This caused 36 pairwise difference comparisons per corpus. After correcting for multiple comparisons using Bonferroni correction, we were able to reject the null hypothesis for 28, 32, and 29 of 36 feature sets for English, French, and Dutch, respectively, implying that the choice of featurization technique leads to a differences in processing time in most cases.

We argue that these differences show that features should be taken seriously, and that, regardless of the structure of a model, different possible featurizations should be taken into account.

References

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3), 643-650

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année psychologique*, 101(3), 447-462.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5), 971-979.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176-1190

Thursday Poster.18

FOCUS PARTICLE POSITION, AND ACCENTS, AFFECT ATTACHMENT

David Potter (Morehead State University) & Katy Carlson (Morehead State University)
d.potter@moreheadstate.edu

Syntactic attachment is affected by the location and presence of pitch accents, with accents on heads drawing the attachment of modifiers (Schafer et al. 1996; Lee & Watson 2011; Carlson & Tyler 2017). But why do accents draw attachment: does a focus-marked syntactic head attract modifiers to this important information (the Focus Attraction hypothesis)? or does the salience of a pitch-accented head attract attachment (the Salience hypothesis)? To test these hypotheses, we conducted two auditory experiments which varied the presence and position of two types of focus marking: L+H* pitch accents and the focus particle *only*, in ambiguous attachment sentences. In both experiments, we found *only* to draw attachment, as predicted by the Focus Attraction hypothesis. Furthermore, pitch accents did not affect attachment when *only* was present, though they did when *only* was absent.

In Experiment 1 (N = 52), 20 sentences like (1) were produced with *only* before Verb1 (*claimed*) or Verb2 (*lied*) and no contrastive accents (1a/1c), or with *only* and a contrastive accent on the verb immediately after *only* (1b/1d). Participants chose between paraphrases of the two meanings (claimed on Monday vs. lied on Monday). We found a significant main effect of the position of *only*, with *only* on V1 raising high attachments ($\beta = 1.17 \pm 0.22$, $\chi^2(1) = 22.53$, $p < .001$). The addition of accent numerically raised high attachments, but not significantly ($p = .06$). There was no interaction between the factors ($p > .85$). This strong effect of the position of *only* supports the Focus Attraction hypothesis, in which focus marking, of any type, is expected to attract attachment. The Salience hypothesis predicts that only salient focus marking, such as pitch-accenting, should attract attachment. Thus this hypothesis does not directly explain why the addition of *only* should have the same effect as making a word longer, higher in pitch, and louder. Nor does it explain why, in the presence of *only*, accenting the word after *only* had no reliable effect on attachment rates.

In Experiment 2 (N = 53), 20 sentences like (2) were produced with contrastive accents on either V1 (2a/2c) or V2 (2b/2d), and, in conditions a- b, the focus particle *only* preceding V1. Participants heard the recordings and chose between paraphrases. The accent alone conditions showed an accent effect, with V1 accent drawing a higher proportion of high attachment ($\beta = -0.45 \pm 0.21$, $\chi^2(1) = 4.44$, $p = .035$). However, the conditions with *only* before V1 showed no effect of accent position ($p = .96$) and had more high attachments overall ($\beta = -0.5 \pm 0.15$, $\chi^2(1) = 10.86$, $p < .001$). In short, when *only* was present, it attracted attachment, while the position of a pitch accent within the scope of *only* did not. This pattern of results is incompatible with the Salience theory, because the accented verbs were equally salient across all conditions, but their effect on attachment disappeared in the presence of the focus particle. These results are expected under Focus Attraction: the position of a focus particle unambiguously marks the scope of focus, and thus the location of the important information. Consequently, no matter where a focused word within the scope of *only* is located, it will be interpreted relative to the focus particle, where attachment is drawn.

Overall, then, the results of these two experiments show that multiple means of marking the position of focus affect syntactic attachment. The pattern of results supports the Focus Attraction Hypothesis. And the way that focus particle position interacts with accents sheds interesting light on the exact implementation of that theory.

- | | | |
|-------|-----------------------------------------------------------------------------|---------------------|
| 1. a. | Kathie <i>only</i> claimed that Bill had lied # _{ip} on Monday. | 63% high attachment |
| | b. Kathie <i>only</i> CLAIMED that Bill had lied # _{ip} on Monday. | 68% |
| | c. Kathie claimed that Bill had <i>only</i> lied # _{ip} on Monday. | 37% |
| | d. Kathie claimed that Bill had <i>only</i> LIED # _{ip} on Monday. | 42% |
| 2. a. | Kathie <i>only</i> CLAIMED that Bill had lied # _{ip} on Monday. | 49% high attachment |
| | b. Kathie <i>only</i> claimed that Bill had LIED # _{ip} on Monday. | 50% |
| | c. Kathie CLAIMED that Bill had lied # _{ip} on Monday. | 44% |
| | d. Kathie claimed that Bill had LIED # _{ip} on Monday. | 37% |

Thursday Poster.19

RETENTION OF SURFACE INFORMATION DURING L1 AND L2 READING: AN EYE-TRACKING STUDY

Denisa Bordag (University of Leipzig), Andreas Opitz (University of Leipzig), Max Polter (University of Leipzig) & Michael Meng (Hochschule Merseburg)

denisav@uni-leipzig.de

Research on memory for gist vs. surface linguistic information seems to converge on the assumption that linguistic information is not retained verbatim, but is converted to conceptual form which is then stored in long term memory. While conceptual information can be stored and retrieved also after long periods of time, verbatim surface information is assumed to decay rapidly and almost immediately after processing (Sachs, 1967, Just & Carpenter 1992, Rummer et al., 2013). At the same time, usage-based theories of grammar acquisition assume that grammatical knowledge is derived from a database of chunks that are stored verbatim in memory (Bybee 1985; Ellis 1996; Goldberg 2006; Langacker 1988; Tomasello 2003). Yet how can a database of memorized verbatim sequences emerge, if verbatim linguistic information is not retained?

The study explored whether L2 German learners (B2-C1 level, Slavic/Romance L1) overrepresent verbatim information in the mental text models they construct during reading compared to L1 German natives. We hypothesized that non-proficient readers are more likely to store verbatim information, since they might either need it more for acquisition purposes and/or compensate with it for e.g. more shallow representations without hierarchical structure organization (cf. Shallow Structure Hypothesis, SSH, Clahsen & Felser 2006, 2017).

Both groups of participants read six short texts (300-400 words) twice while their eye movements were tracked. There were eight ROIs in each text. Four ROIs involved nouns (lexical condition) and the other four ROIs involved two sentences in active voice and two in passive voice (syntactic condition). The second version (V2) of each text differed from the first version (V1) in that two of the four ROIs always stayed the same and two were changed. In the lexical condition, the noun was exchanged for its near-synonym, in the syntactic condition active was transformed into passive and vice versa. The rationale of the design was that if readers represent verbatim surface information in their mental text model, they should notice the changes in V2 and respond to them by longer fixation times. Thus, the critical comparisons were between fixation times at the same vs. changed ROIs in lexical and syntactic conditions during the second reading of the text, e.g. total fixation duration at the word "city" after "city" in V1, or after "town" in V1. After V1 was read, participants performed several mathematical tasks. After V2, they answered comprehension questions which justified the usefulness of the second reading (no difference between the groups).

The results revealed that L2 learners (N=20) fixated the changed ROIs significantly longer in both lexical and syntactic conditions, while there was only a small numerical difference between the critical conditions for L1 speakers (N=24). The findings support the initial hypothesis that less proficient readers retain more details regarding linguistic surface information during reading (recently also Gurevich, Johnson & Goldberg, 2010; Sampaio & Konopka 2013). The results will be discussed in the context of present cognitive (Fuzzy Trace Theory), acquisition (Declarative/Procedural Model) and processing (SSH) approaches.

The findings support the initial hypothesis that less proficient readers retain more details regarding linguistic surface information during reading (recently also Gurevich, Johnson & Goldberg, 2010; Sampaio & Konopka 2013). The results will be discussed in the context of present cognitive (Fuzzy Trace Theory), acquisition (Declarative/Procedural Model) and processing (SSH) approaches.

BEING NICE IS HARD: UNDERSPECIFICATION & COERCION IN COPULA SENTENCES

Anna Prysłowska (University of Tübingen)
anna.pryslowska@uni-tuebingen.de

The difference between (1a) and (1b) has been subject to much debate. (1a) describes Sophie's characteristic; (1b) carries an element of Sophie's control over her behavior, i.e. agentivity (Partee, 1977; Carlson, 1977). Some see the agentive interpretation as the result of an optimal re-interpretation (Maienborn, 2003). Others argue that the copula is underspecified relative to the situation argument it can take; the agentive interpretation arises through composition (Rothstein, 1999). The German copula has the same surface form for both stative and agentive interpretations. We present three studies that investigate the semantics of the German copula.

- (1) (a) Sophie is friendly. (b) Sophie is being friendly.
- (2) (a) Sophie | {war, verhielt sich} | freundlich, | und zwar | um
 Sophie {was, behaved herself} friendly and namely in.order.to
 | die Eltern | stolz auf sie zu machen.
 the parents proud of her to make
- (b) Sophie | {war, verhielt sich} | freundlich, | und zwar | weil |
 Sophie {was, behaved herself} friendly and namely because
 die Eltern | sie gut erzogen haben.
 the parents her well raised have

Experiment 1 investigated whether the agentive interpretation of copula predicate constructions involves a semantic re-interpretation of the copula. The eye-tracking during reading study had a 2×2 design: verb (*sein* copula, *sich verhalten* 'to behave'), conjunction (agentive *um*, neutral *weil*). *Sich verhalten* was used as a control condition. 40 participants were tested in 4 lists with 60 items; see (2), | indicate invisible IAs. Critical IAs were the conjunction *weillum*, one IA preceding and one following it. The results confirm that the agentive interpretation of the copula requires coercion: copula+*um* was more difficult than copula+*weil*. This was visible through interactions in: 1st fixation duration ($\beta=0$, SE=0, $t=3.8$, $p<0.01$), 1st pass RT ($\beta=0$, SE=0, $t=3.4$, $p<0.01$), and regression path duration ($\beta=0.4$, SE=0.1, $t=4$, $p<0.01$) on the conjunction.

Experiment 2 investigated whether agentive coercion is easier when the adjective is modified by the particle *so*. *So* introduces evaluation or focus, possibly facilitating agentive coercion (Wiese, 2011). Methods and analysis were as in Exp. 1. Materials had one crucial modification: in copula sentences, the adjective was preceded by *so*. There was an interaction between factors in regressions out of IA *und zwar* ($\beta=1.19$, SE=0.51, $t=2.36$, $p<0.05$). This effect was weaker than in Exp. 1 and present in only one measure. Otherwise, the participants favoured the agentive conjunction over the neutral one (1st fixation duration: $\beta=-0.20$, SE=0.08, $t=-2.6$, $p<0.01$; 1st pass RT : $\beta=-0.23$, SE=0.08, $t=3.05$, $p<0.005$). The addition of *so* facilitates the agentive re-interpretation of the stative VP in combination with the conjunction *um*.

Experiment 3 is an ongoing eye-tracking study, which will be completed by June '18. It explores whether the coercion effects from Exp. 1 and 2 persist when the neutral conjunctions is *da* 'since' instead of *weil*. *Da* removes the need to correct for word length differences compared to *um*. Replicating the results would strengthen the claim that the copula is semantically stative.

In sum, Exp. 1 and 2 found evidence of the copula's lexical stativity. The agentive interpretation requires a cognitively costly process of coercion, but the addition of *so* facilitates it.

G. N. Carlson. *Reference to Kinds in English*. PhD thesis, University of California, 1977.

C. Maienborn. Against a Davidsonian analysis of copula sentences. In *NELS 33 Proceedings*, pages 167–186, Amherst, 2003. GLSA.

B. Partee. John is easy to please. *Linguistic structures processing*, pages 281–312, 1977.

S. Rothstein. Fine-grained structure in the eventuality domain: The semantics of predicative adjective phrases and be. *Natural language semantics*, 7(4):347–420, 1999.

H. Wiese. So as a focus marker in German. *Linguistics*, 49(5):991–1039, 2011.

A VERBAL ILLUSION WITHOUT THE VERB: DERAILED COMPOSITIONAL INTERPRETATION IN SENTENCE COMPLETIONS

Dario Paape, Shravan Vasishth (University of Potsdam)
 paape@uni-potsdam.de

Consider the sentence *No head injury is too trivial to be ignored*. It exemplifies the so-called 'depth charge' illusion, a persistent misinterpretation in which the principle of compositionality is violated and sentence meaning is reversed [1,2,3]. The correct but absurd interpretation (*All head injuries should be ignored, even if they seem trivial enough to be treated*) is supplanted with a sensible, incorrect meaning (*Treat even seemingly trivial head injuries*). It has been hypothesized that the origin of the illusion is the verb *ignored*, which is misinterpreted in the presence of too many negations because compositional interpretation crashes [1]. This claim, however, has never been empirically investigated, and it is possible that compositional interpretation is already compromised before the verb appears. In order to find out whether the source of the illusion lies before the verb, we conducted a sentence completion study in German with 60 participants and 32 sentences. The stimuli were truncated after *to*, as shown below. The adjectival negation condition (b) serves as a control case for which no illusion is expected.

a. DOUBLE NEGATION	Keine	Kopfverletzung	ist zu	ungefährlich, um	___
	no	head injury	is too	un-dangerous to	
b. ADJ. NEGATION	Manch eine	Kopfverletzung	ist zu	ungefährlich, um	___
	some a	head injury	is too	un-dangerous to	

Under a compositional interpretation, ... *be treated* would be a sensible continuation for both sentences. However, if the depth charge illusion occurs, participants should instead supply a continuation like ... *be ignored* for the double negation sentence (a). If the illusion appears in this task, it is unlikely that the verb is the main source of the effect.

Participants' varied responses were grouped according to a binary criterion (illusion/no illusion) for analysis. Twenty coders judged the supplied continuations according to one of two coding schemes. The coders were blind to the experimental manipulation as the original preambles were never presented to them. Scheme A asked whether a combination such as *head injury – to see a doctor about* (when this was the supplied continuation) indicated that the head injury was considered to be 'of importance' or not. This scheme was used because potentially low importance of the subject (e.g., *ignore trivial head injuries*) typically accompanies the illusion. For Scheme B, coders judged whether the continuation supplied for the (a/b) sentence fit with a negation- and quantification-free sentence, as shown below.

c. NO NEGATION	Diese Kopfverletzung	ist zu gefährlich,	um [behandelt/ignoriert] zu werden.
	This head injury	is too dangerous	to [be treated/be ignored].

Under the illusion, the meanings of the double negation (a) and no negation (c) sentences are similar. This is evidenced by the fact that ... *be ignored*, the seemingly sensible but unlicensed continuation of (a), is also a good continuation of (c). The compositionally licensed continuation of (a), ... *be treated*, on the other hand, is not.

Inter-coder agreement was higher for Scheme A (Fleiss' kappa: 0.77) than for Scheme B (Fleiss' kappa: 0.49). Results showed that according to both coding schemes, participants produced illusion-signaling continuations like ... *be ignored* more often in the double negation than in the adjectival negation condition (Scheme A: $\hat{\beta}$ = 80%, CrI: [88%, 70%]; Scheme B: $\hat{\beta}$ = 64%, CrI: 76%, 49%). There was no evidence that the time taken to supply a continuation was affected by the manipulation.

Contrary to previous speculation, the result suggests that the origin of the depth charge illusion lies before the verb. We suggest that the prediction of the verb is disrupted because the implicit negation carried by *too* (*too X to Y* → not Y) induces overload in the presence of global negation, which illicitly scopes over the to-be-produced verb.

[1] Wason & Reich (1979). Q J Exp Psychol. [2] Natsopoulos (1985). J Psycholinguist Res. [3] Kizach, Christensen & Weed (2016). J Psycholinguist Res.

Thursday Poster.22

IS MEMORY OF CORRECTED INFORMATION SUPPRESSED IN TEXT PROCESSING?

Scott McKenzie (University of Dundee), Roger van Gompel (University of Dundee)

r.p.g.vangompel@dundee.ac.uk

Language is not always produced in a completely fluent form, therefore comprehenders need to deal with errors, hesitations and pauses so that these don't hinder understanding. Recent research has started to look at how this is done, particularly regarding structural ambiguity. It has been found that comprehenders often do not revise their initial interpretation of a sentence following the correction of a speech error (Corley, 2010; Lau & Ferreira, 2005; Slevc & Ferreira, 2013). However, it is unclear whether comprehenders also fail to revise their interpretation of larger parts of a corrected text involving propositional information.

This study investigated how the interpretation of a text is affected by corrections. Written stories consisting of 16 sentences (shortened example below) were presented either with (version a) or without a correction (b). Each text mentioned two characters (e.g., a painter and sculptor). In the version with a correction, a statement that was made earlier about one of the characters (e.g., the painter) was changed to a statement about the other character (e.g., the sculptor). In the version without correction, the earlier statement was repeated. Each text was followed by one question about information that was changed in the version with a correction. In both versions, there was one question condition requiring a YES answer and one requiring a NO answer. Thirty-two participants read the stories in a non-cumulative sentence-by-sentence format. They were told that the stories sounded as if they were spoken from memory. Reading times and question responses were recorded.

The painter and the sculptor met up in the morning and played chess. (...)

The (a) painter/(b) sculptor suddenly had to leave to complete a shift at work.

She had a part-time job at a pizza restaurant and had to make deliveries. (...)

The (a) painter/(b) sculptor was only paid five dollars an hour.

Eventually the (a) painter/(b) sculptor decided to tell jokes when she made deliveries so that she could get good tips.

(a) No, sorry, /(b) Yes, that's right, it was the sculptor who had a job at the pizza restaurant.

Later that day, the painter and the sculptor met up again and went to the cinema. (...)

Was it the painter who was paid five dollars an hour? NO

Was it the sculptor who was paid five dollars an hour? YES

The override hypothesis predicts that the correction (a) overrides the earlier incorrect statement. The initial incorrect interpretation is successfully revised, and therefore, participants should answer the question correctly as often as when the story contains no error (b). The lingering information hypothesis predicts that the earlier incorrect statement is not always revised and should therefore retain some activation. As a result, they should answer the question incorrectly more often after (a) than (b). The inhibition hypothesis predicts that the correction inhibits the earlier incorrect statement, therefore participants should answer the question correctly more often in (a) than (b).

Both the proportion correct answers and their answering times showed that participants had significantly ($p < .05$) greater difficulty answering questions about text that was corrected (a) than text that was not (b). This was despite the fact that reading times for the sentence containing the correction (*No, ...*) were longer ($p < .05$) than for the alternative sentence without correction (*Yes, ...*), suggesting that participants did sometimes realise they had to revise their interpretation. The findings support the lingering information hypothesis and indicate that, similar to syntactic analyses (Lau & Ferreira, 2005; Slevc & Ferreira, 2013), incorrect propositional information within larger text is not completely overwritten after a correction.

Thursday Poster.23

ACCESS OF DEMONSTRATIVES IN DISCOURSE: EVIDENCE FOR DIRECT ACCESS

Klaus von Heusinger (University of Cologne) & Andreas Brocher (University of Cologne)
kheusing@uni-koeln.de

Indefinite demonstrative noun phrases, such as *this guy* in *Then I met this guy*, are specific (Prince, 1981), cataphoric (Gernsbacher & Shroyer, 1989), express noteworthiness (Ionin, 2006), and are discourse prominent (von Heusinger, 2011). In the present study, we tested the claim that indefinite demonstratives introduce their discourse referent more directly than regular indefinites (*a guy*), a claim that is compatible with previous studies (Gernsbacher & Shroyer, 1989). The rationale is that when speakers establish reference in comprehension, they use the descriptive content of the noun phrase much more strongly when they encounter a regular indefinite than when they encounter a demonstrative. In other words, in contrast to demonstratives, indefinites (as well as regular definites) introduce their discourse referent by delimiting a set of potential referents associated with their descriptive content. Establishing reference, then, triggers a selection from that set. If this claim were correct, we should find differences in accessibility for indefinites with pre-activated, i.e. inferred concepts and indefinites with brand-new concepts. Similar accessibility levels should obtain for inferred and brand-new demonstratives, however.

We conducted a visual world, eye-tracking experiment, in which 45 participants (ps) listened to German short stories consisting of three sentences (40 expt; 80 fillers). The first sentence provided a context, the second sentence introduced two human referents, and the third sentence contained a personal pronoun that could be interpreted as the subject or object referent of the second sentence. Importantly, the critical object referent could either be inferred from context (inferred, *gym... a trainer vs. this trainer*) or not (brand-new, *theatre... a trainer vs. this trainer*) and was either introduced with a noun following a demonstrative (*this*) or an indefinite article (*a/an*). We measured ps' eye fixations starting at pronoun onset and analyzed which of the four pictures onscreen participants were fixating: picture of the object referent, picture of the subject referent, or one of the two unrelated filler pictures.

Starting at ambiguous pronoun onset, ps looked overall more to the picture of the object than to the picture of the subject referent. More interestingly, for object NPs with an indefinite article, ps looked more to the related picture when the referent could be inferred than when it was brand-new. For object NPs with a demonstrative article, fixation times were slightly longer for brand-new than inferred referents. Next, we fitted generalized mixed models with looks to the object picture vs. looks to all four pictures as dependent measure and included information status (inferred or brand-new) and article (demonstrative or indefinite) as predictors. For the bins 500 – 800 ms and 800 – 1100 ms post pronoun onset, we found a statistically reliable Information status x Article interaction, both z s > 5, both ps < .001. Follow-up models only including the indefinites revealed a marginal difference between inferred and brand-new referents for the 500 – 800 ms bin, $z = 1.74$, $p = .083$, and a reliable difference for the 800 – 1100 ms bin, $z = 4.74$, $p < .001$. The two respective models that only included the demonstratives showed no significant differences between inferred and brand-new referents, both z s < 1.1, both ps > .2.

We take our result as evidence that indefinite demonstratives introduce their discourse referent in a more direct way than regular indefinites, possibly because their access interacts much more weakly with the descriptive content of the associated NP.

Gernsbacher, M. A., & Shroyer, S. (1989). The cataphoric use of the indefinite *this* in spoken narratives. *Memory & Cognition*, 17, 536–540.

von Heusinger, K. (2011). Specificity, referentiality and discourse prominence: German indefinite demonstratives. In Ingo Reich (ed.), *Proceedings of Sinn und Bedeutung 15*, 9–30. Saarbrücken: Saarland University Press.

Ionin, T. (2006). This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics*, 14, 175–234.

Prince, E. F. (1981). On the inferencing of indefinite-*this* NPs. In Bonnie Webber, Aravind K. Joshi & Ivan Sag (eds.), *Elements of Discourse Understanding*, 231–250. Cambridge: Cambridge University Press.

A NEURAL NETWORK MODEL OF ADAPTATION IN READING

Marten van Schijndel (Johns Hopkins University) & Tal Linzen (Johns Hopkins University)
vanksy@jhu.edu

Humans rapidly adapt their lexical and syntactic expectations to match the statistics of the current linguistic context (e.g., Fine et al., 2013). Computational word prediction models (language models) that adapt to the current context make more accurate predictions (e.g., Kuhn & de Mori, 1990). Combining these two research traditions, we propose a simple adaptive neural language model, and show that adaptation improves our predictions of human reading times.

Our baseline model is a long short-term memory (LSTM) language model trained on 90 million words of English Wikipedia articles. For adaptation, at the end of each new sentence, we update the parameters of the model based on its errors in predicting that single sentence. We tested the model on the Natural Stories Corpus (Futrell et al., 2017), which has 10 narratives with self-paced reading times from 181 English speakers.

Linguistic accuracy: We first measured how well the model predicts upcoming words. We use the standard measure of perplexity; this measure is lower when the model assigns higher probabilities to the words that in fact occurred. Adaptation over the test corpus dramatically improved test perplexity compared to a non-adaptive version of the model (86.99 vs 141.49).

Fit to reading times: We next tested whether our adaptive language model is a better model of human expectations than a non-adaptive one. We adapted the model to each story independently¹ and used its surprisal at each word to predict the corresponding self-paced reading times. Adaptive surprisal was predictive of reading times over a linear mixed model baseline containing non-adaptive surprisal² ($p < 0.001$), and its presence caused non-adaptive surprisal to no longer be a significant predictor (Table 1). This result indicates that this model more closely represents human expectations than non-adaptive language models.

Does the model adapt its syntax? To test whether the model adapts its lexical predictions, its syntactic predictions, or both, we generated 200 pairs of dative sentences, each with a prepositional object (PO) variant (*The boy threw the ball to the dog*) and a double object (DO) variant (*The boy threw the dog the ball*). We shuffled 100 PO items into 1000 filler items from Wikipedia and adapted the model to these 1100 sentences. We then froze the weights of the adapted model and tested its predictions for two types of sentences: the PO counterparts of the DO sentences used during adaptation, and 100 sentences that had the same syntax as those used during adaptation (DO) but shared no content words with them. We then repeated the experiment with the role of DO and PO reversed. This process was repeated 10 times each for PO and DO with different critical items and filler sentences. We found that the model adapted more strongly to vocabulary choice than syntax but was sensitive to both (Figure 1).

Overall, adaptation greatly improved the language model's accuracy and RT predictions. This improvement was due not only to lexical but also syntactic adaptation.

	$\hat{\beta}$	$\hat{\sigma}$	t
Sentence position	0.29	0.53	0.5
Word length	6.42	1.00	6.4
Surprisal	-0.89	0.68	-1.3
Adaptive surprisal	8.77	0.68	13.0

Table 1: Fixed effect self-paced reading regression coefficients.

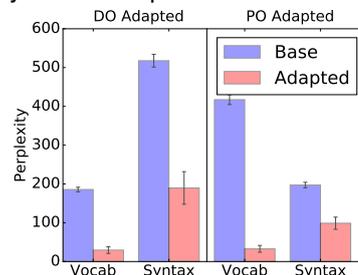


Figure 1: Lexical vs syntactic adaptation. (Note that the base model prefers PO.)

¹After each story, the model reverts to the initial language model and must restart adaptation on the next story.

²The baseline is as follows: $RT \sim \text{word_length} + \text{sentence_position} + \text{non-adaptive_surprisal} + (1|\text{word}) + (1 + \text{word_length} + \text{sentence_position} + \text{non-adaptive_surprisal} + \text{adaptive_surprisal} | \text{subject})$

ACCESSIBILITY AND AFFECTEDNESS OF BARE DIRECT OBJECTS

Elyesa Seidel (Universität zu Köln)

elyesa.seidel@uni-koeln.de

Turkish bare nouns in object position have typically been analyzed as pseudo-incorporated nouns (Öztürk, 2005; Kamali, 2015). They are assumed to exhibit the following properties: First, they evoke a number neutral interpretation; second, they cannot act as antecedents for anaphoric pronouns and, third, they obligatorily take narrow scope. The first two properties are variable properties across languages (cf. Hungarian, Persian, and Hindi), while narrow scope, is arguably not a sufficient test for pseudo-incorporation, as it also holds for different types of arguments (cf. narrow scope indefinites). We focused on the first two properties and conducted two acceptability judgment experiments.

The design of **Experiment 1** consisted of four conditions organized in a 2x2 factorial design. We manipulated the type of anaphoric expression (pronoun vs. definite description) and the number marking of the anaphoric expression (singular vs. plural). A total of 36 critical items (in addition to 12 fillers) were constructed. Items consisted of a context sentence (cf. (1)) and a target sentence (a - d). Materials were distributed across four lists such that each list contained only one condition of one set. Lists were distributed across all participants and items were presented in pseudo-randomized order. Eighty monolingual, native speakers of Turkish were asked to rate how naturally they thought the context and target sentences were linked to each other on a scale from 1 ("not natural at all") to 7 ("completely natural"). Results reveal a significant main effect of number marking, $b=-0.94$, $SE=0.22$, $t=-4.32$, and a main effect of anaphoric expression, $b=0.23$, $SE=0.10$, $t=2.22$. Together, our data indicate that bare direct objects can act as antecedents of pronouns, suggesting that these expressions do not behave like typical incorporated structures: They show that continuations with singular anaphors are more acceptable than continuations with plural anaphors, regardless of pronoun type.

The design of **Experiment 2** consisted of four conditions, organized in a 2x2 factorial design. We manipulated the presence (*bu* 'this' + N) vs. absence (null pronoun) of an anaphoric expression as well as verb type, comparing *verbs of use* (cf. (3)) with *verbs of creation* (cf. (2)). We constructed 48 critical items (in addition to 24 fillers). Each item again consisted of a context sentence (cf. (2) and (3)) and a target sentence (a - b). Verb type and pronoun type were crossed (12 item per condition for verbs of use and 12 items per condition for verbs of creation) and all items were distributed across four lists. The task was the same as in Experiment 1. Results for 160 participants show a significant main effect of verb type, $b=1.26$, $SE=0.18$, $t=6.86$, and a reliable effect of anaphoric expression, $b=0.40$, $SE=0.13$, $t=2.51$. We take these as evidence (i) that anaphoric reference to the pseudo-incorporated noun is more acceptable in contexts with creation verbs than in contexts with verbs of use and (ii) that the anaphoric link created by the definite description *bu+N* is more acceptable than the anaphoric link by the null pronoun.

In sum, our results suggest that, first, bare nouns in Turkish are anaphorically accessible (Exp. 1 & 2) and, second, the accessibility of inanimate bare nouns in Turkish is associated with a degree of affectedness (Exp. 2). Third, both animate bare nouns (Exp. 1) and inanimate bare nouns (Exp. 2) trigger singular interpretations. On a more theoretical level, our findings reveal that Turkish bare nouns are not discourse opaque. They rather show properties of discourse transparency and should therefore be analyzed as weak indefinites rather than pseudo-incorporated nominals.

(1) Samet dün Taksim meydanında hırsız yakaladı.
'Samet did **thief-catching** at the Taksim Square yesterday.'

(a) Onu rezil etti. (b) Onları rezil etti.

(c) Hırsız rezil etti. (d) Hırsızları rezil etti.
'He embarrassed **him/them/the thief/the thieves**.'

(2) Gönül geçen gün ofiste mektup okudu.

'Gönül did **letter-reading** at the office yesterday.'

(3) Sami geçen gün çalışma odasında mektup yazdı.

'Sami did **letter-writing** at the office yesterday.'

(a) *pro* Üç sayfalıydı. (b) Bu mektup üç sayfalıydı.
'**It/This letter** was three pages long.'

References: Kamali, Beste. 2015. Caseless direct objects in Turkish revisited. In André Meinunger (ed.), *Byproducts and side effects: Neben-produkte und Nebeneffekte*. ZAS Papers in Linguistics 58, 107–123. Berlin: ZAS. • Öztürk, Balkız. 2005. *Case, referentiality and phrase structure*. Amsterdam: John Benjamins.

IS QUOTATION ARTICULATED?

QUOTATIONAL CONSTRUCTIONS AND THEIR ACOUSTIC PROPERTIES

Marcel Schlechtweg (University of Kassel) & Holden Härtl (University of Kassel)
schlechtweg@uni-kassel.de

Name-mentioning constructions (NaMenC) involving predicates like *call* as in *Blood poisoning is also called "sepsis"* are instances of pure quotation, i.e., a meta-linguistic device used to point to linguistic shapes, see, e.g., Cappelen & Lepore (1997). NaMenC inform the addressee about the conventionalized name of a lexical concept, e.g., the name *sepsis*, as opposed to *The doctor diagnosed a sepsis*, where *sepsis* is used denotatively (DenoC). While the semantic and pragmatic properties of (pure) quotation are well explored, its phonetic realization is widely understudied, with only few studies examining the acoustic profiles of reported speech, see, e.g., Klewitz & Couper-Kuhlen (1999), and ironic utterances, see Anolli et al. (2000).

Objectives. Our paper aims at investigating (i) whether quotation in general is reflected acoustically and (ii) whether the articulator is sensitive to name-mentioning quotation. For this purpose, we compared the acoustic parameters of NaMenC (see 2a/b) vs. DenoC (see 2c/d) and of unquoted (see 2a/c) vs. quoted (see 2b/d) nouns. Consider the examples:

- (1) Context sentence (identical across the four conditions):
Diese korbähnliche Transporttasche für den Rücken erleichtert die Arbeit in der Landwirtschaft. 'This basket-like bag for the back facilitates the work in agriculture.'
- (2)
 - a. *Man nennt sie Kiepe unter Bauern. 'One calls it Kiepe among farmers.'*
 - b. *Man nennt sie „Kiepe“ unter Bauern. 'One calls it "Kiepe" among farmers.'*
 - c. *Man kennt die Kiepe unter Bauern. 'One knows the Kiepe among farmers.'*
 - d. *Man kennt die „Kiepe“ unter Bauern. 'One knows the "Kiepe" among farmers.'*

Materials and procedure. In a production study, eight native speakers of German were recorded while reading eight German nouns of low frequency (*Kapern* 'capers', *Pappel* 'poplar', *Kutte* 'robe', *Kippa* 'kiprah', *Koppel* 'paddock', *Kate* 'cottage', *Kiepe* 'pannier', *Puppe* 'cocoon') embedded in the four conditions (within-subject/item design). The data was analyzed with Praat and five dependent variables were measured: (A) Constriction duration of noun-initial plosive, (B) VOT of noun-initial plosive, (C) duration of first nominal vowel, (D) maximum intensity of this vowel, and (E) maximum F0 of this vowel.

Results. Quoted nouns (see 2b/d) were pronounced with significantly longer (A) and (B) than non-quoted nouns (see 2a/c). Further, NaMenC (see 2a/b) showed a significantly longer (B) and higher (D) and (E) than DenoC (see 2c/d). Moreover, we found a significant interaction between the two independent factors for (E). The post-hoc comparisons of this interaction indicated that NaMenC were produced with a significantly higher maximum F0 than DenC if both were not enclosed in quotation marks. If quotation marks were present, however, the effect disappeared. Based on the two main effects, we argue that NaMenC are articulated differently than DenoC and that the presence/absence of quotation marks has an influence on the acoustic realization of an item.

Outlook. In a follow-up study, the (i) syntactic differences between NaMenC and DenoC are eliminated and (ii) the four conditions are tested without context to balance out the givenness of the critical nouns. To conclude, we will discuss the implications of our results for theories of quotation as well as the interface between semantics and phonetics.

References

- Anolli, L. & R. Ciceri & M. G. Infantino. 2000. Irony as a Game of Implicitness: Acoustic Profiles of Ironic Communication. *Journal of Psycholinguistic Research* 29(3). 275-311.
- Cappelen, H. & E. Lepore. 1997. Varieties of Quotation. *Mind* 106(423). 429-450.
- Klewitz, G. & E. Couper-Kuhlen. 1999. QUOTE – UNQUOTE? The role of prosody in the contextualization of reported speech sequences (= Interaction and Linguistic Structures 12). Konstanz: University of Konstanz.

Thursday Poster.27

OROFACIAL EXPRESSIONS AND ACOUSTIC CUES IN WHISPERED AND NORMAL SPEECH

Marzena Żygis (Leibniz-ZAS), Susanne Fuchs (Leibniz-ZAS), Katarzyna Stoltmann (Leibniz-ZAS)

zygis@leibniz-zas.de

Communicative functions of prosody are not only executed by means of acoustic cues, but also by facial expressions and gestures. The extent to which the facial expressions are used differs across speakers and languages (see Srinivasan et al. (2003) for English, House (2002) for Swedish, and Borràs-Comes (2012) for Catalan). While eyebrows, lip aperture and head movements have already been investigated in the context of different sentence modalities (see e.g. Grice & Mücke 2014) it is entirely unclear whether and how speakers execute a difference between questions and statements when they whisper, i.e. when F0 is absent. The very few studies on whispered speech concentrated on the role of visual versus audio information for speech understanding (Dohen & Loevenbruck 2008, 2009). The aim of the present study is, however, to examine whether there are possible trade-offs between acoustic cues and oro-facial expressions due to the absence of F0 in whispered speech.

To this end, we probed the movements of eyebrows and lip aperture in questions and statements produced by ten native speakers of German (all females, mean age 25.7). We put seven markers on the informant's face positioned (i) above the upper lip, (ii) below the lower lip (iii) close to the left lip corner, (iv) close to the right lip corner, and three reference points.

Parallel to acoustic recordings, eye brows' and lips' movements records were collected by means of a motion capture system (OptiTrack, *Motive* Version 1.9.0) with 12 cameras (Prime 13). The participants were asked to read sentences displayed on a computer screen in whispered and normal speech. The sentences included questions and statements which differed only in the punctuation, i.e., questions ended in a question mark and statements in a full stop, e.g., *Er mag diese Piste*. 'He likes this slope.' *Er mag diese Piste?* 'Does he like this slope?' The sentence-final words always started with a bilabial stop followed by an unrounded vowel /a, ε, ɪ/, e.g. *Bitte* 'request' with stress falling on the first vowel. In total, we analysed 2377 items with respect to the lip aperture and 2314 items with respect to eyebrows.

Linear mixed effect models employed for the statistical analysis revealed that questions were produced with a larger lip aperture than statements ($t= 4.23$, $p<.001$). The difference was present in the production of all three vowels in both whispered and normal speech. The vowels [a] and [ε] were produced with a greater lip aperture in questions than in statements leading to a significant interaction of sentence modality and vowel type ($t= 3.69$, $p<.01$). Furthermore, the significant interaction of speech mode and vowel type was due to the fact that whispered [ε] and [ɪ] were produced with a greater lip aperture than their voiced counterparts ($t= 2.17$, $p<.05$). Furthermore, our results point to a raised right eyebrow in questions as opposed to statements but the effect was found only in a few speakers. The influence of sentence type on the left eyebrow was highly speaker dependent and not significant.

Our acoustic analysis shows that in whispered speech the sentence-final words are more prominent in questions than in statements: both sentence-final words and stressed vowels in these words are significantly longer in questions (words: $t=3.36$, $p<.01$; stressed vowels: $t=2.63$, $p<.01$). As expected, the RMS amplitude is higher in normal speech than in whispered speech ($t=9.34$, $p<.001$) but does not contribute to differences in questions and statements. Finally, F0 of sentence-final words also changes in the expected direction with F0 max and F0 mean being significantly higher in questions than statements.

Taking together, the study shows that greater prominence in questions is expressed by both acoustic and oro-facial expressions whereby selected cues (lip aperture, duration) are even more pronounced in whispered than in normal speech suggesting compensatory visual effects for the auditory F0 absence.

REPRESENTATION OF NUMBER IN AGREEMENT COMPREHENSION

Yingzhao Zhou, Rhea T. Eskew, Jr., and Neal Pearlmuter (Northeastern University)
zhou.yingz@husky.neu.edu

Language processing requires mechanisms for reliably encoding linguistic representations in working memory. We examine whether representations of number used during comprehension are discrete (e.g., binary-valued, as in singular vs. plural) versus continuous (taking on a range of possible values between singular and plural endpoints). While some agreement processing models are built upon discrete representations (e.g., Franck et al., 2002:*LCP*, 2006:*Cogn*) and others on continuous ones (e.g., Eberhard et al., 2005:*PsyRev*), the distinction has not previously been directly investigated.

306 participants (Ps) read 160 singular-head stimuli like *The key to the cabinet(s) was/*were rusty...*, which varied head-local noun match and grammaticality, mixed with 80 plural-head stimuli with the same conditions, across two sessions, using self-paced reading with yes/no comprehension questions. 255 Ps were analyzed (those with 90+% correct comprehension across all items). The experiment examined the nature of the subject noun phrase (NP) number representation by focusing on reading time patterns at the verb (*was* or *were*) and following word, for mismatch conditions (*key...cabinets*) compared to match conditions (*key...cabinet*). Patterns of means replicated some earlier experiments (e.g., Pearlmuter et al., 1999:*JML*; mismatch > match for grammatical cases; mismatch < match for ungrammatical), but both discrete and continuous representations predict this average difficulty pattern. We therefore analyzed reading time (RT) distributions from each condition, assuming that RT reflects the difficulty of reconciling verb and subject NP number.

Regardless of number representation, subject NP number in match conditions should be clearly singular, so reconciliation with the verb should be uniformly fast for grammatical trials or uniformly slow for ungrammatical; each of these conditions should yield its own single RT distribution (modeled as an ex-Gaussian). In the grammatical mismatch condition, however, the discrete model predicts a combination of two trial types: Either the subject NP is specified as singular, and verb reconciliation is identical to the grammatical match condition; or the subject NP is specified as plural (based on interference from the local noun *cabinets*), so that verb reconciliation is hard (comparable to the ungrammatical match condition). This predicts a mixture RT distribution in which the grammatical match condition's ex-Gaussian is mixed with the ungrammatical match's. The continuous model instead predicts grammatical mismatch trials should come from a single distribution (again modeled as an ex-Gaussian) with a slower mean than the grammatical match case, corresponding to the greater difficulty of reconciling the verb's singular number with the subject NP's value, which is shifted toward the plural end of the scale by the mismatching noun (*cabinets*). This same logic and comparable predictions apply to the ungrammatical mismatch condition.

For each P separately, we fit the continuous model (single ex-Gaussian) and discrete model (mixture of that P's two match condition ex-Gaussians) separately to the grammatical mismatch data and compared their χ^2 goodnesses-of-fit by *F* test, which accounts for differences in the number of model parameters. Per-P power was estimated for each model by Monte Carlo simulations, and mean (per-P) experiment power was used to estimate the number of Ps expected to show a significantly better fit for a given model versus the other. At the verb, using a per-P significance criterion of $\alpha = .2$ (identical patterns obtained for $\alpha = .1$ and $.05$), 8 Ps (out of 255) were predicted to show a significantly better fit for the discrete than the continuous model, but 11 Ps actually did. A binomial test thus fails to reject the discrete model as better than the continuous ($p = .91$). But a binomial test of the continuous model as better than the discrete finds far fewer Ps (15) than would be expected (38; $p < .01$). This same pattern held for the ungrammatical mismatch condition, and for the word following the verb.

Thus specific two-distribution mixture models of the mismatch conditions' verb regions fit significantly better than single-distribution models, supporting the idea that number is represented discretely rather than continuously during comprehension. We will compare discrete model variants further to investigate alternative models.

CONTRIBUTION OF GRAMMATICAL FUNCTIONS AND SEMANTIC ROLES TO DISCOURSE PROMINENCE IN TURKISH

Gökben Konuk (University of Cologne) & Klaus von Heusinger (University of Cologne)
goekben.konuk@uni-koeln.de

It is widely accepted that highly prominent discourse referents are picked up by reduced anaphoric forms, such as pronouns and null forms, whereas less accessible referents are picked up by more explicit forms, such as full noun phrases. The question, then, is what exact factors determine the accessibility of referents in discourse. There has been much empirical work on English and German but very little empirical work on Turkish. Previous more theoretical research has shown that anaphoric accessibility of discourse referents in Turkish depends on grammatical functions (Turan, 1998). It is assumed that subject referents are more accessible than object referents and are therefore re-mentioned with the most reduced anaphoric expression. However, there has often been a confound between subjecthood and agentivity, so that it is unclear whether grammatical function (subject vs. object), semantic role (agent vs. patient), or both determine referent accessibility in Turkish. Recent findings from German (Schumacher et al., 2016) suggest that multiple factors can affect referent accessibility in discourse and emphasize the relevance of agentivity next to subjecthood. To tease apart the effects of grammatical function and semantic role on discourse prominence in Turkish, we conducted a web-based language comprehension study with psych verbs. The proto-agent, i.e. the experiencer (Dowty, 1991, Primus, 1999) was either in subject (1a) or object position (1b). We expected more prominent referents to be mentioned again with more reduced forms than less prominent referents.

The 16 critical items consisted of a context sentence containing a psych verb and a target sentence containing an ambiguous subject pronoun, realized either as *pro* or a personal pronoun. We included eight subject-experiencer and eight object-experiencer verbs and each verb appeared in two pronoun conditions. The 20 filler items always contained verbs of transfer and the pronoun was resolved to the goal referent through world knowledge. Sixty monolingual native speakers of Turkish were asked to read one sentence at a time and to answer the question „kim?“ („who?“) by selecting the referent they thought the pronoun referred to. All experimental items had been tested in a pretest.

Our results show no main effect for pronoun type, $\beta = -0.47$, $SE = 0.51$, $z = -0.93$, $p = 0.355$, but a main effect for verb type, $\beta = -2.20$, $SE = 0.54$, $z = -4.1$, $p = 0.001$. In constructions with subject-experiencer verbs, pronouns were preferably interpreted as referring to the subject referent. In constructions with object-experiencer verbs, in contrast, both referents were roughly equally often selected as antecedent of the pronoun. These data indicate that both grammatical function and semantic role contribute to the prominence of referents in Turkish. As a follow-up, we are currently conducting a language production study using the same test materials as in the comprehension study discussed here.

- (1) a. [Gökhan]_{SE} dünkü kahvaltı daveti sonrasında Naz'ı büyüleyici gülüşünden dolayı gün boyunca düşledi. *pro/o* sekizde mail attı.
Gökhan dreamed after the breakfast yesterday all the time of Naz because of her charming smile. pro/o wrote at 8 pm an e-mail.
- b. Mete uyumsuz davranışlarıyla [Seher'i]_{OE} geçen haftaki grup çalışmasında çok kızdırdı. *pro/o* birden gruptan ayrıldı.
Mete angered during the group work last week Seher through his rude behaviour. pro/o left suddenly the group.

References: Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547-619.; Primus, Beatrice. 1999. *Cases and thematic roles: Ergative, accusative and active*. Berlin: De Gruyter.; Schumacher, Petra, Manuel Dangi & Elyesa Uzun. 2016. Thematic role as prominence cue during pronoun resolution in German. In Anke Holler & Katja Suckow (eds.), *Empirical perspectives on anaphora resolution*. Berlin: De Gruyter. 213-239.; Turan, Ümit Deniz. 1998. Ranking forward looking centers in Turkish: Universal and language specific properties. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering theory in discourse*. Oxford: Oxford University Press. 139-160.

LINGUISTIC PREDICTION AND VISUAL ATTENTION DO NOT INTERACT IN READING

Suzanne Jongman (Max Planck Institute for Psycholinguistics), Marta De Pedis (Università di Bologna) & Ashley Lewis (Haskins Laboratories)
 suzanne.jongman@mpi.nl

Both attention and prediction are thought to facilitate language comprehension, such that attended and predicted words are processed more easily than unattended or unpredicted items. This may lead to the conclusion that attention and prediction constitute functionally similar phenomena, and yet attention and prediction have opposing effects on related neural responses (the N400 for language stimuli): high attention increases neural responses [1] whereas high predictability decreases neural responses [2]. These two factors have been shown to interact in a visual perception task: without attention the standard predictability effect (reduced neural response) was found but with attention the prediction effect was reversed [3]. It was suggested that attention boosts the precision of predictions, leading to a heightened weighting of sensory evidence, and enhanced neural activity. In the present study the question was whether neural indices of attention and prediction interact similarly in sentence reading. During reading, visual attention is redirected constantly as new words are fixated. It's known that this visual attention can be guided by predictability [4]. We tested the converse relationship to see if attention can enhance neural responses related to prediction.

A sentence reading task was combined with a visuospatial cueing paradigm. Dutch sentences were presented with the final word omitted, then a cue appeared signaling participants where to direct their attention (without overt eye movement) to one of two locations. The final word appeared for 250ms either in the validly or invalidly cued location (50/50, visual attention manipulation similar to [3]). Half the final words were predictable (high cloze) and half were unpredictable (low cloze, with cloze conditions achieved by manipulating sentence constraint). While participants (N = 28) performed the task, EEG (64ch) was collected. The average amplitude in the time-window 300 to 500 ms was compared in a 2 x 2 x 5 design: high vs low cloze x valid cue vs invalid cue x 5 regions (with 9 electrodes each, left anterior, right anterior, left posterior, right posterior, vertical midline). We found main effects of predictability and attention ($F(1,27) = 22.22, p < .001$; $F(1,27) = 9.58, p < .01$), but importantly no interaction between the two factors nor a three-way interaction with region (both $F < 1$). Using Bayesian analysis, by comparing all models with the interaction of interest to all models without this interaction, we found evidence against the interactions (BF below 1 provides evidence for the null, two-way BF = 0.14, three-way BF = 0.02). All in all, these results indicate that prediction and attention did not interact. Therefore, attention and prediction have opposing, and separate, effects on the N400 during sentence reading. When it comes to integrating words into a sentence context, visual attention does not boost the precision of lexical predictions.

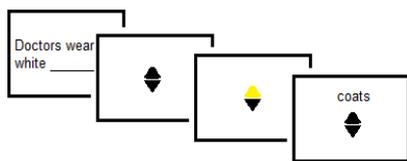


Fig 1. Predictable, validly cued trial

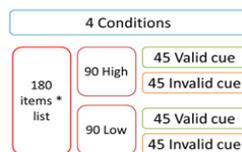


Fig 2. Design

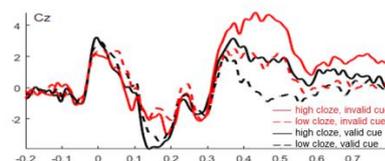


Fig 3. ERP waveform

[1] Critescu & Nobre, 2008, *J Cogn Neurosci* 20:5; [2] Szewczyk & Schriefers, 2013, *J Mem Lang* 68:4; [3] Kok et al., 2011, *Cereb Cortex* 22:9; [4] Huettig et al., 2011, *Acta Psych* 137:2

RESTING STATE EEG POWER PREDICTS COGNITIVE AND LANGUAGE SKILLS

Suzanne Jongman (Max Planck Institute for Psycholinguistics)

suzanne.jongman@mpi.nl

To perform a language task, an individual not only uses language-specific processes but also domain-general skills [1,2]. Cognitive and linguistic task performance has been related to an individual's brain functioning, for instance to power (i.e. the number of neurons discharging synchronously) in certain frequency bands as measured by EEG. High average theta is associated with poor performance on memory tasks, and high alpha with good performance on attention tasks [3]. Similarly, beta frequency modulations have been linked to various aspects of linguistic processing [4]. Even power measured at rest has been shown to predict later task performance: individuals' alpha power is related to intelligence [5] and beta power predicts the level reached in a language training program [6]. This suggests that stable network properties at rest can largely determine performance outcomes. It is unclear however which domain-general skills and which linguistic skills are predicted by oscillatory activity at rest, and by which frequency bands in particular. The present study aimed to provide more clarity and tested whether the relationship between domain-general skills and linguistic skills found behaviorally is driven by shared brain network functioning at rest.

Dutch young adults ($N = 34$, data collection ongoing) performed a battery of tasks aimed to test both language abilities and general cognitive abilities. Language was divided into comprehension (lexical decision RTs), production (picture naming RTs), verbal fluency (number of retrieved items), vocabulary size (Peabody score), novel word learning and artificial grammar learning (RTs for correct test items). Assessed domain-general skills were intelligence (Raven's score), processing speed (factor from simple and choice RT and letter comparison RT), attention (attention network test, factor from alerting, orienting and executive control effects), and working memory (total score forward and backward digit span). In addition, five minutes of eyes-closed resting state EEG (64ch) was recorded and power analyses were performed following [3]. Using Fast Fourier Transform, the power spectrum was calculated for 2-sec epochs, which was log-transformed, and then averaged across all epochs. For each participant and channel, power was calculated across three frequency bands: theta (4-7.5Hz), alpha (8-12.5Hz) and low beta (13-14.5Hz). Power values were selected from only one electrode to reduce the number of correlations, the electrode with the highest average power (theta FCz, alpha PO3, beta PO8).

Theta power correlated with vocabulary size ($r = -.51$, $p < .01$, $BF = 16.5$) and with speed of correct decisions in a grammar learning test ($r = -.46$, $p < .01$, $BF = 7.4$). Alpha power correlated with processing speed: ($r = -.42$, $p < .05$, $BF = 3.6$). These results mostly corroborate the typical theta/alpha dichotomy [3]: individuals with high resting state theta showed limited vocabulary knowledge whereas high alpha was related to quick processing. Less clear is the link between theta and grammar learning as it suggests that higher power results in quicker responding (note that there is no speed-accuracy trade-off). It seems theta power is related to how quickly you process and decide on a newly acquired string. All in all, these results provide evidence that stable brain functioning can predict task performance for some but not all tasks. Moreover, oscillations at different frequencies are related to different cognitive functions, with alpha linked to general processing and theta to linguistic skills.

[1] Huettig & Janse, 2016, *Lang Cogn Neurosci*, 31:1; [2] Shao et al, 2012, *Q J Exp Psychol*, 65:10; [3] Klimesch, 1999, *Brain Res Rev*, 29:2; [4] Weiss & Mueller, 2012, *Front Psychol*, 3:201; [5] Doppelmayr et al., 2002, *Intelligence*, 30:3; [6] Prat et al., 2016, *Brain Lang*, 157.

CROSS-LINGUISTIC INVESTIGATION OF THE REPRESENTATIONS DRIVING THE CHOICE BETWEEN PRONOUNS AND NOUNS

Kumiko Fukumura (University of Stirling), Coralie Hervé (University of Essex), Sandra Villata (University of Connecticut), Francesca Foppolo (University of Milano-Bicocca) & F.-Xavier Alario (Aix-Marseille Université)
kumiko.fukumura@stir.ac.uk

When referring, speakers must decide whether to use a noun or a pronoun; across many languages, pronouns are the most frequently chosen referring expressions (e.g., Leech et al., 2001; New et al., 2001). Recent evidence suggests that speakers are less likely to choose pronouns (over nouns) when referential candidates are similar than when they are dissimilar (e.g., Fukumura & Van Gompel, 2011; Fukumura, Hyönä, & Scholfield, 2013). An important question for theories of language production concerns the level of representation at which such an effect might arise. The *linguistic competition* account predicts that the choice of using a pronoun is determined by the accessibility of the linguistic antecedent that the pronoun is replacing. In this view, speakers use fewer pronouns when the potential antecedents are linguistically more similar, reducing the antecedent's accessibility and pronoun use. Alternatively, the *non-linguistic competition* account predicts that pronoun use is affected by the referent's non-linguistic properties; speakers use fewer pronouns when some non-linguistic properties of the possible referents make them more similar, reducing the referent's non-linguistic accessibility and pronoun use.

We contrasted these two accounts in six referential communication experiments (32 participants, 48 experimental items in each experiment), eliciting full nouns as well as pronominal expressions in three different languages: English, Italian, and French. This was done as the level of representation that affects pronoun use might be language-dependent due to the different pronoun properties. Participants saw a display containing two objects on a computer screen, and read aloud a context sentence (e.g., *The pear above the apple is on Number 2*) to their addressee. The target object then changed location, whilst the competitor remained still. Participants described the change (e.g., *Now [the pear/it] is on Number 3*), such that the addressee could identify the target and its new location. We varied the congruence of the antecedents' semantic categories (e.g. pear and apple vs. pear and cigar) and their phonological similarity (e.g. pear and pen vs. pear and bull). We also varied the context in which the objects appeared; in the non-linguistically similar condition, both target and competitor were placed in a red box, signalling to participants that either could move in the display. In the non-linguistically dissimilar condition, only the target was in a red box, signalling that only the target could move.

In all languages, participants produced fewer pronominal expressions when both target and competitor were in the box than when only the target was in the box, demonstrating an effect of non-linguistic similarity. By contrast, neither semantic similarity (category congruence) nor phonological similarity had an effect; participants produced no fewer pronouns in English and in French and no fewer null pronouns in Italian when the potential antecedents were semantically or phonologically similar than when they were dissimilar. An experiment in French additionally showed an effect of the antecedent's grammatical role: Participants used more pronouns when the referent was mentioned in the subject position rather than when it was mentioned in the non-subject position in the context sentence. This confirmed that our task was sensitive to other variables that affect pronoun use. Semantic and phonological similarity as well as non-linguistic similarity affected the time taken for participants to read aloud the context sentence, demonstrating participants' sensitivity to semantic and phonological similarities. The lack of semantic and phonological effects on the choice of pronouns therefore suggests that speakers do not necessarily access the antecedents' semantic categories or their phonologies when choosing to use a pronominal expression.

To conclude, the choice of using a pronoun is cross-linguistically affected by the accessibility of the referent's non-linguistic representation, not the antecedent's linguistic accessibility, in support of the non-linguistic competition account.

Thursday Poster.33

NORMALIZING VOWELS AT A COCKTAIL PARTY

Hans Rutger Bosker (Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands),
Eva Reinisch (Institute of Phonetics and Speech Processing, Ludwig Maximilian University
Munich, Germany) & Matthias Sjerps (Donders Institute for Brain, Cognition and Behaviour,
Radboud University, Nijmegen, Netherlands)

HansRutger.Bosker@mpi.nl

Different talkers have different average F0 and formant frequencies (for example due to variation in vocal tract length). One way in which listeners cope with such variation is by 'normalizing' vowels for the surrounding context. A clear demonstration of such normalization has been that a vowel ambiguous between /ɪ/ (with low F1) and /ɛ/ (with high F1) is more likely to be perceived as /ɛ/ when preceded by a precursor sentence with a relatively low F1 (and as /ɪ/ after a sentence with high F1).

This contrastive effect of preceding context, known as vowel normalization (or talker, or spectral normalization), has been explained in two ways: 1) as resulting from listeners' sensitivity to the *auditory contrast* between the long-term average spectrum (LTAS) of the precursor and the spectral properties of the ambiguous target vowel; 2) as *speaker-dependent restructuring* of phonemic category boundaries. Importantly, the two accounts make different predictions regarding talker-incongruency between precursor and target, and the role of selective attention.

An auditory contrast account would predict comparable vowel normalization induced by talker-congruent and (LTAS-matched) talker-incongruent precursors, since the effect would be based on the general LTAS. This account would also predict no modulation of the effect by selective attention since auditory contrast effects involve early general-auditory processes that presumably precede attentional modulation. Some support for this account is found in the observation that cognitive load does not modulate vowel normalization. Conversely, a speaker-dependent account would predict talker-incongruent precursors to not elicit any normalization, and attention to modulate normalization since it involves later cognitive adjustments.

In three experiments, native Dutch participants categorized target words ($n=20$; produced by female Talker A) that contained vowels from spectral continua ranging from /ɪ/ (low F1) to /ɛ/ (high F1; e.g., "bid" /bɪt/ *pray* vs. "bed" /bɛt/ *bed*). Target words were preceded by spectrally-manipulated precursor sentences ($n=200$), with F1 shifted up or down by 20% using Burg's LPC methods in Praat.

Experiment 1 (E1) showed that single *talker-congruent* precursors (i.e., from Talker A) with F1 shifted down induced a greater proportion of /ɛ/-responses relative to precursors with F1 shifted up. In Experiment 2 (E2) targets were preceded by *talker-incongruent* precursors (produced by LTAS-matched female Talkers B and C; F1 manipulated within talkers). E2 also demonstrated a vowel normalization effect, albeit significantly reduced, compared to E1. Interestingly, the presence of *some* vowel normalization in E2 challenges speaker-dependent accounts, but at the same time the significant reduction from E1 to E2 challenges a general-auditory account.

Finally, in Experiment 3, the same targets were preceded by *two simultaneous* precursors, one in each ear (location counter-balanced). To resemble typical 'cocktail party' scenarios, one of the precursors was always produced by Talker A, the other by another talker (B or C; F1 manipulated randomly within talkers). Participants were instructed to always attend to Talker A. Transcriptions demonstrated that participants were able to report what the attended talker said. Moreover, vowel categorization consistently followed the F1 of the attended precursor: trials with an attended low F1 induced a greater proportion of /ɛ/ responses (independent of unattended F1), with a similar effect size as in Experiment 2.

Together, these experiments indicate that vowel normalization is sensitive to the overall LTAS, but is also modulated by talker-congruency and selective attention. As such, the outcomes are in support of a combination of *speaker-dependent* adjustment of phonemic category boundaries and *auditory contrast*.

**The effects of context and literality on L1 and L2 idiom processing:
Evidence from self-paced reading**

Sara Beck & Andrea Weber (University of Tübingen)
sara.beck@uni-tuebingen.de

Idioms challenge typical processing models as they have figurative meanings that differ from the compositional meaning of their individual parts. Current research has shown that context can impact access to the figurative meaning of an idiom, but not necessarily the literal meaning as composition is an immediate, obligatory process (e.g., Holsinger & Kaiser, 2013). However, recent research suggests the presence of figurative contexts might hinder literal composition (e.g., Canal et al., 2017). Some idioms, however, have a high potential to be interpreted literally (high-literality), such as *break the ice*, and others do not (low-literality), like *lose one's cool*, and the limits of biasing contexts are still unclear when comparing these idioms. Also, speakers may vary in their ability to integrate context in online processing strategies. In particular, non-native (L2) speakers, are said to have a tendency to be more reliant on literal meaning (e.g., Cieślicka, 2006), and it is not clear that L1 predictions hold across a less-proficient L2 speaker group. Our study therefore asked: How does biasing context impact access to literal and figurative meaning in high- and low-literality idioms for both L1 and L2 readers, and what happens in online processing when contextual expectations are met or turn out to be misleading?

We conducted an English phrase-by-phrase self-paced reading study on L1 (American) English and highly-proficient L2 (German) speakers on sentences containing high- and low-literality idioms. Idioms were embedded in literal and figurative contexts and followed by prepositional phrases that were either congruent or incongruent with these expectations (Table 1). Norming data (plausibility and context) on items was collected and included in the analyses.

		Analysis regions:		IDIOM	RESOLUTION	R+1	
High-Literality	The new schoolboy, who didn't know anyone in his class,	just wanted to	<i>break the ice</i>	with his peers	sooner than later	...	
	The chilly Eskimo, who was eager to catch some fish,			on the lake			
Low-Literality	The emotional writer, who often started political debates,	didn't want to	<i>lose his cool</i>	in his anger	...		
	The overheated runner, who was resting under a tree,			from the shade			

Table 1. Figurative contexts/endings (DARK), Literal contexts/endings (LIGHT).

Using linear mixed-effects models, we analyzed three regions of interest. On the resolution, literal resolutions were read more quickly than figurative ones ($p < .01$) across idiom-types and readers. The following region (R+1) showed faster reading times following figurative resolutions overall ($p < .001$). However, unlike previous studies, interactions showed that congruent contexts and resolutions improve reading times following high-literality idioms for both literal ($p < .05$) and figurative interpretations ($p = .05$), while unambiguous idioms show advantages for figurative resolutions regardless of context ($p < .01$). Across all analyses, no significant differences other than overall reading times were found between L1 and L2 readers. These findings provide further evidence for immediate literal processing. Additionally, as a sentence unfolds in real-time, both reader groups integrate context only when necessary (i.e. ambiguity is likely), and may abandon literal integration processes where unnecessary (i.e. ambiguity is unlikely). Furthermore, it seems that highly proficient L2 readers are just as able to integrate or ignore contextual cues during reading where experience deems it (un)necessary.

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6), 928–943.
 Cieślicka, Anna B.; Heredia, Roberto R. (2011): Hemispheric asymmetries in processing L1 and L2 idioms: effects of salience and context. In *Brain and language* 116 (3), pp. 136–150.
 Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 866–878.

RUN-SPEAKING? SIMULATIONS OF RATE CONTROL IN SPEECH PRODUCTION

Joe Rodd^{*1,2} Hans Rutger Bosker^{1,3} Mirjam Ernestus^{2,1} Antje S. Meyer^{1,3} Louis ten Bosch^{2,1}
¹Max Planck Institute for Psycholinguistics; ²Radboud University, Centre for Language Studies; ³Radboud University, Donders Institute for Brain and Behaviour *joe.rodd@mpi.nl

That speakers can vary their speaking rate is evident, but how they accomplish this has hardly been studied. Consider this analogy: when walking, speed can be continuously increased, within limits. However, to speed up further, humans must run. Are there multiple qualitatively distinct ‘gaits’ in speech planning that resemble walking and running gaits? Or is control achieved by continuous modulation of a single gait? This study investigates these possibilities through simulations of a novel connectionist computational model of the cognitive process of speech production, which mimics the temporal characteristics of observed speech.

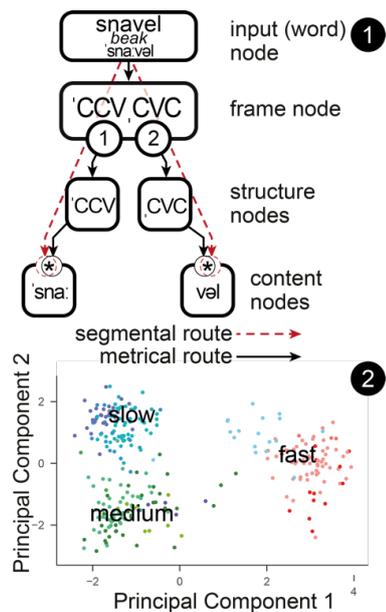
CONNECTIONIST MODEL Our model, illustrated in Figure 1, is derived from Dell, Burger and Svec’s (1997, *Psychol. Rev.* 104(1), 123) model of serial order in language, and sequentially retrieves syllable-level motor plans in response to activation in a word level input node. A frame node mediates, encoding metrical structure and enforcing serial order. This model is the first of its type to predict the precise timing of motor plans and account for the ability to control rate in speech production.

The model has many parameters (connection weightings, thresholds, etc.) that can be adjusted to achieve a specific speaking rate. Different ‘regimes’ (combinations of parameter settings) can be engaged to achieve different speaking rates. We consider each parameter as a dimension of a high-dimensional ‘regime space’, in which the regimes occupy different locations.

MODEL TRAINING Our model approximated the distributions of observed syllable durations and syllable overlap durations in the PiNCeR corpus of Dutch disyllabic words produced at fast, medium and slow speaking rates. Syllable onset and offset were identified from the acoustic signal on the basis of spectral instability as an index of syllable overlap. Together, these duration distributions form a ‘fingerprint’ of the speech production system operating at a given rate. The model was trained separately for each speaking rate, by the natural selection-inspired optimisation algorithm NSGA-III. The training identified parameter values that caused the model to best approximate the ‘fingerprint’ distributions of each speaking rate in the corpus. The fit of the model was assessed by calculating the Kullback-Leibler divergence between the model’s predicted distributions and those taken from the corpus for each speaking rate.

PREDICTIONS In a one gait system, where we ‘speed-walk’ to speak faster, the regimes used to achieve fast and slow speech are qualitatively similar, but quantitatively different. In regime space, they would be arranged along a straight line. Different points along this axis correspond to different speaking rates. In a multiple gait system, where we ‘walk-speak’ for slower speaking rates, but ‘run-speak’ to speak faster, this linearity would be missing. Instead, the arrangement of the regimes would be more disperse, with no obvious relationship between the regions associated with each gait, and an abrupt shift in parameter values to move from speeds associated with ‘walk-speaking’ to ‘run-speaking’.

RESULTS Our model achieved good fits in all three speaking rates. In regime space, the broad arrangement of the parameter settings selected for the different speaking rates is clearly non-axial, suggesting that ‘gaits’ may be present in the speech planning system (see principal component projection in Figure 2). Thus, we provide the first computationally explicit connectionist account of the ability to modulate the speech production system to achieve different speaking styles.



ARE ALL INVALID PARAFOVEAL PREVIEWS CREATED EQUAL?

Leigh B. Fernandez (University of Kaiserslautern), Christoph Scheepers (University of Glasgow) & Shanley E.M. Allen (University of Kaiserslautern)
leigh.fernandez@sowi.uni-kl.de

Reading research has shown that while fixating on a word, we begin preprocessing information from the following word including orthographic, phonological, morphological, and (debatably) semantic information (Schotter et al., 2012). This preprocessing can lead to a reading benefit, in that previewed words have shorter fixation durations when fixated on compared to words that were not previewed (e.g., because they were blocked with a string of uninformative letters i.e. invalid preview). This phenomenon has traditionally been called a *preview benefit* but has recently been renamed the N+1 preview effect (Vasilev & Angele, 2017) since invalid previews may lead to *preview costs* that interfere with reading (Hutzler et al., 2013). However, no research has systematically tested the role of invalid previews on reading times.

In the current study we test the role of invalid previews by manipulating the type of preview using a gaze contingent boundary paradigm. Native English speaking participants (n=24) read sentences (Table 1) with the preview of the critical word (*suit*) manipulated such that an invalid preview (that shared no features with the critical word) was previewed and changed to the critical word upon making a saccade across in an invisible boundary (*). Six preview types were used as shown in Table 1. The identical form (*suit*) served as the baseline and was compared to the invalid previews. English and German non-words were used to test whether participants were sensitive to language specific sub-lexical orthographic information and were based on two language specific orthographic-measures (Schröter et al., 2016).

Table 1: Example stimuli

	Identical	English pseudo-word	German pseudo-word	Illegal	Xs	blank
The tailor trimmed th*e	suit	bame	biet	ryjf	xxxx	

First fixation duration (FFD) and total duration (TD) of the identical preview were compared to each invalid preview and sliding contrasts were used to compare each level to the next. For FFD, only the X preview differed from the Identical, and there was a continuum with FFD increasing from the pseudo-words to the illegal and to the X, and then a decrease from X to blank. This suggests there is a graded effect of uninformative masks with increases in FFD as a mask becomes less “word-like”. For TD, there was a significant decrease from English to German pseudo-words and an increase from German pseudo-words to Illegal (see Figure 1). We plan to investigate this further by comparing a group of L2 English speakers with German L1. Overall these data show that not all previews are equal, and research should be aware that non-word-like invalid previews can inflate reading durations compared to invalid previews that resemble legal words.

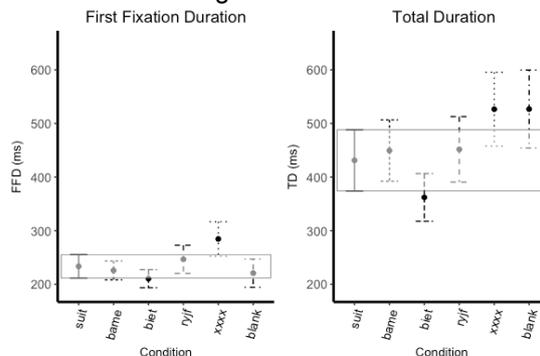


Figure 1. Reading durations by preview type, bars represent standard error, and the box encompasses the baseline area.

Thursday Poster.37

MORPHOLOGICAL AND STEREOTYPICAL GENDER IN PROCESSING AGREEMENT

Anastasia Generalova (St.Petersburg SPBU) & Natalia Slioussar (HSE, Moscow, & SPBU)
slioussar@gmail.com

Russian has three genders (M, F and N). The gender of the noun cannot be unambiguously determined from its inflectional affixes (although there are some clear tendencies), but becomes evident from agreeing adjectives, participles and verbs. Like many other languages, Russian faces the following problem: many nouns denoting professions are grammatically M. How to call a female director or a female author? In Russian, two routes are available.

Firstly, a corresponding grammatically F noun can be formed (e.g. *zhurnalist* 'journalist_M' – *zhurnalistka* 'journalist_F', *uchitel* 'teacher_M' – *uchitel'nica* 'teacher_F'). Unlike German, where the *-in* suffix can be applied to any relevant noun, and like French, Russian uses a variety of suffixes to form such nouns, and many M nouns denoting professions do not have an established F counterpart at all. Secondly, Russian has so-called common gender, and a particular originally M noun can be used with M and F agreeing forms (*nash* / *nasha vrach* 'our_M / our_F doctor'). This route is available for any relevant noun.

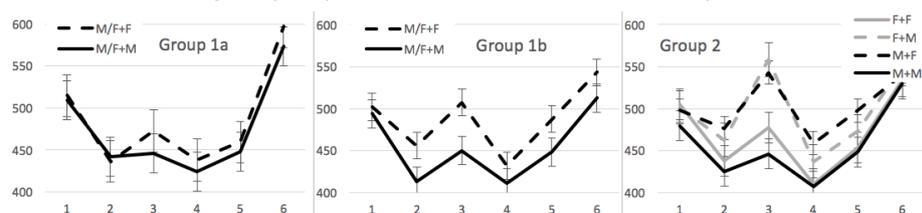
This study focuses on the second option. We conducted a self-paced reading experiment comparing word-by-word reading times for the sentences in two groups: group 1 with subject nouns denoting professions that can be used as common gender nouns, as in (1a-b), and group 2 with subject nouns denoting personal qualities, like in (2a-d). In the latter case, there is always a M and a F noun, common gender is impossible, and (2b) and (2d) contain an agreement error. We wanted to compare the processing of such errors to the processing of (1b) — the sentence is grammatical, but the subject noun is originally M and has zero inflection typical for M nouns, so the readers might react to that. Group 1 had two subgroups, A and B: with professions that are perceived as stereotypically female (as in (1)) or male (e.g. *mjasnik* 'butcher'). Stereotypical norms for Russian were taken from (Garnham, Yakovlev 2015).¹

- (1) a. *Pediatr byl obespokoen iz-za objavlennija karantina.* b. *Pediatr byla...*
 pediatrician_{M/F} was_M worried_M because of the quarantine pediatrician_{M/F} was_F
- (2) a. *Intrigan byl ostorozhen v etom voprose.* b. **Intrigan byla...*
 intriguer_M was_M cautious_M in this question intriguer_M was_F
- c. *Intriganka byla ostorozhna v etom voprose.* d. **Intriganka byl...*
 Intriguer_F was_F cautious_F in this question intriguer_F was_M

Participants were 62 native Russian speakers. We had 32 stimulus sentence sets (16 per group, group 1 had two subgroups A/B and two conditions, group 2 had four conditions, as in (1)-(2)). Two factors were manipulated: (i) whether the gender of the predicate coincides with the (original) gender of the subject and (ii) whether the subject is M or F (in group 2). All sentences had the same structure. We also had 64 grammatically correct filler sentences.

Average RTs in different conditions are presented on Fig. 1. We used RM ANOVAs (by participants and by items) for the statistical analysis. Factor I was found to cause significantly smaller RT differences in the group 1A (stereotypically female professions) than in the group 1B, and in the group 1B than in the group 2. Thus, the idea that every noun denoting profession, even the most stereotypically male one, can be used to refer to a woman, is already present in the mental grammar of Russian speakers. If the profession is perceived as a stereotypically female, there is virtually no reaction to the surface mismatch between the common gender noun that is originally M (and 'looks M' due to its inflection) and the predicate.

Fig. 1. Average RTs (in ms) per region (word) in different conditions.



¹ This study also analyzed reading times for sentences with Russian nouns denoting professions, but only whole-sentence reading times were recorded.

INCREMENTAL GENERATION DRIVES “EFFICIENT” LANGUAGE PRODUCTION

Spencer Caplan (University of Pennsylvania)

spcaplan@sas.upenn.edu

A major testing ground for mechanistic accounts of language production (De Smedt, 1990) is the study of “syntactic optionality”; i.e. given multiple potential syntactic encodings for equivalent semantic sentences, what factors govern the use of one form rather than another (Ferreira and Dell, 2000).

A prominent previous account, the “Uniform Information Density” hypothesis (UID) (Jaeger, 2010), proposes that such syntactic optionality is driven by a speaker’s implicit managing of computable information content to maximize communicative efficiency. On this model, conditional probability serves as a proxy for optimized information content. Previous corpus modeling of optional ‘that’-omission has supported UID (Jaeger, 2010). However, such evidence is potentially problematic. In addition to syntactic confounds (Grimshaw 2009), rates of ‘that’-omission show a great deal of variability by genre, ranging from 1% in formal writing to 85% in conversational speech (Biber 1999). With so much variance attributable to sociolinguistic register, it is not clear what we might learn about the cognitive architecture of the production system.

We propose the English verb-particle (VP) construction (e.g. ‘John picked up the book’ vs. ‘John picked the book up’) as a better case to evaluate theories of optionality. We extracted a large database of VP alternations from COCA, a balanced corpus of modern English (Davies, 2009). From these we can compare a UID account with the general framework of incremental generation (IG) (Bock and Levelt, 2002). Under IG, the architecture of sentence generation requires several components: retrieve lemmas from memory, assign such elements their proper functional roles, as well as assign linear order in adherence with syntactic restrictions. If we assume such modules operate incrementally and in parallel, then variations in the order in which information is delivered from one component to the next can readily affect the linear order elements appear in speech. So long as the system does not intentionally hold retrieved lemmas back in a buffer, any factors which speed up lexical access will also be proxies for spoken linear order (see Rayner, 1998 for a review of lexical access factors). While both UID and IG make convergent predictions regarding conditional probability, only IG predicts that the factors of frequency, definiteness, and constituent length, etc. should all predict linear order in optional constructions. This is because such factors correlate with lexical retrieval times yet are orthogonal to “information density”.

Following Jaeger’s original study, a multilevel logit model was used to evaluate predictions of an IG account of sentence production compared with UID. The dependent variable was the binary outcome of linear order (particle-first rather than object-first) in VP sentences. Evaluating over the entire database, we see a strong correlation between all IG related factors and output order (Blue in Table). This includes the effect of conditional probability posited by both UID and IG. However, when limited to evaluation over even moderately long objects (at least four words), then the predictions of UID are not borne out while the other instantiations of IG remain significant (Green in Table). To whatever degree we can characterize the output of the language production system as “efficient” in information ordering, this is an emergent property of a simple, incremental generation system.

Factor	All VPs (58,619 cases)				Longer object VPs (5,679 cases)			
	Estimate	Std. Err	Z-Value	P	Estimate	Std. Err	Z-Value	P
Freq(obj)	-282.8	38.12	-7.42	~0	-502	77.9	-2.84	0.01
Info(obj verb)	0.039	0.01	8.6	~0	0.02	0.03	0.69	0.49
NP-length	1.0	0.03	40.37	~0	0.63	0.12	5.18	~0
Definite-Obj	-0.59	0.02	-26.9	~0	-1.13	0.15	-7.33	~0

CHINESE WH-IN-SITU AND ISLANDS: A FORMAL JUDGEMENT STUDY

Jiayi Lu (Northwestern University), Cynthia K. Thompson (Northwestern University) & Masaya Yoshida (Northwestern University)

Introduction: Cross-linguistic studies on the processing of wh-question constructions have found that wh-filler-gap dependency (WhFGD) in English and wh-in-situ in Chinese are processed similarly, i.e., the parser forms a long-distance dependency between the wh-element and its licenser [1]. However, syntactic studies on Chinese wh-in-situ have identified a generalization that suggests that WhFGD and wh-in-situ are different. In WhFGD constructions, both argument and adjunct are sensitive to islands, but in wh-in-situ construction, argument is not sensitive to islands, but adjunct is ([2,3,4]) as in (1).

(1) a. Argument	b. Adjunct
Ni yudao le [chi shenme de nvhai] _{RC} ?	*Ni yudao le [weishenme chi jiaozi de nvhai] _{RC} ?
You meet ASP [eat what REL girl] _{RC} ?	You meet ASP [why eat dumpling REL girl] _{RC} ?
“What did you meet the girl who ate?”	“Why did you meet the girl who ate dumplings”
(asking about what the girl ate)	(asking about why the girl ate dumplings)

Through a formal acceptability judgement task, this study examines whether the this generalization is accurate or not. Specifically, this study shows that both argument and adjunct wh-in-situ lead to strong acceptability degradation when embedded within a relative clause (RC) island ([3,6]), suggesting that the generalization is not correct. Thus, WhFGD and wh-in-situ show a very strong parallelism in Chinese.

Background: Previous studies discovered that Chinese wh-in-situ are parsed in similar ways as WhFGD, triggering a decrease in processing accuracy proportional to length of dependencies ([1]). Following this analysis, we should expect Chinese wh-in-situ to be influenced by relative clause (RC)-islands in similar ways as WhFGD ([5]). The observation in (1) would then be unexplained by this "covert" filler-gap dependency analysis of Chinese wh-in-situ. Based on the contrast as in (1), it has been suggested that argument wh-in-situ does not undergo movement and thus is immune from island effects ([4]). However, there has been no study that systematically examined the contrast in (1). Based on informal judgments, it seems that the adjunct wh-in-situ is indeed less acceptable than argument wh-in-situ in an RC island. However, without knowing acceptability of argument and adjunct wh-in-situ in non-island environments, we cannot tease apart island effect and acceptability degradation due to processing additional adjuncts.

Experiment: We employed a formal acceptability rating experiment (n=56) adopting the methodology from [6], which shows that the island effect is characterized as the super-additive interaction of Dependency Length (short vs. long) × Structure (non-island vs. island). We used a 2×2×2 factorial design where Dependency Length (short vs. long) × Structure (non-island vs. island) × Wh-category (DP (who/what) vs. Adverb (why)) are manipulated as independent factors. A linear mixed effects model reveals significant main effects of Length (t=10.497, p<0.001), Structure (t=3.903, p<0.001), and Wh-category (t=3.694, p<0.001). There are significant interactions of Length × Structure in both adjunct (t=4.362, p<0.001) and argument (t=6.890, p<0.001). There is no significant interaction of Structure × Wh-category in the long dependency conditions (t=0.400, p=0.690). There are two important findings. First, arguments show island effect just like adjuncts. Second, the main effect of Wh-category and the lack of interaction of Structure × Wh-category in long dependency conditions suggest that adjuncts are not more sensitive to islands than arguments. Based on this result, we contend that both the argument and adjunct wh-in-situ are processed in the same way (they undergo "movement") as WhFGD constructions. Furthermore, the "argument/adjunct asymmetry" is only apparent, but rather that adjuncts are processed more difficultly than argument in general.

References: [1] Xiang et al., (15) Constructing Covert Dependencies *JML*. [2] Huang (82) Move wh in a language without wh-movement *The Linguistic Review*. [3] Lasnik & Saito, (92) Move α: Conditions on Its Application and Output *MIT Press* [4] Tsai (94) On Nominal Islands and LF Extraction in Chinese *NLLT*. [5] Sprouse Hornstein (13) Experimental Syntax and Island Effects *Cambridge* [6] Sprouse et al., (10) A test of the relation between working memory capacity and syntactic island effects

Thursday Poster.40

AUDITORY STATISTICAL LEARNING RAPIDLY SUPPORTS THE PROCESSING OF LARGER LINGUISTIC CHUNKS ACROSS EARLY CHILDHOOD.

Evan Kidd (MPI-Psycholinguistics, The Australian National University), Joanne Arciuli (University of Sydney), Michael Smithson (The Australian National University), Erin Isbilen (Cornell) & Morten Christiansen (Cornell).

Email: evan.kidd@mpi.nl

Classic psycholinguistic studies showed that memory for sentences is greater than for word lists (Miller & Isard, 1963), suggesting knowledge of language enables humans to process larger chunks of information. Statistical learning (SL) is one cognitive process assumed to support the extraction of regularities from the input. Thus, one potential outcome of rapidly identifying and extracting reoccurring patterns from the input is that newly acquired sequences should allow more information to be processed. In the current study we developed a new child-friendly measure of SL that enabled us to test this proposal. Using the embedded triplet paradigm (Saffran, Aslin, & Newport, 1996), 145 (71 females) children aged 5;6 – 7;7 ($M = 6;7$) were familiarised with a continuous sequence of syllables containing four trisyllabic words [transitional probabilities (TPs) within words = 1.0, TPs between words = 0.33]. After a 10-minute familiarisation period children completed a 32-trial 2AFC task, where they distinguished between target words and foils (TPs = 0). Performance was above chance [$M = .53$, $t(144) = 3.07$, $p = .003$, $d = .26$]. Following the 2AFC task, the children participated in a newly-developed recall task, in which they recalled (i) 3-syllable words versus foils, and (ii) two 3-syllable words ‘chunks’ (i.e., two trained words together) versus 6-syllable foils (a randomization of the syllables from the trained words). Following the proposal that SL supports language acquisition, we predicted that children would more accurately recall trained versus foil items, and that this difference would be greater for 6-syllable items because experience with the trained items would alleviate memory problems associated with repeating long sequences (Gathercole, 2006). The predictions were borne out. Children performed significantly better on trained items than on foils [$\beta = -1.89$, $z = -2.32$, $p = .02$], and while children performed better on 3-syllable versus 6-syllable strings overall [$\beta = 1.77$, $z = 25.8$, $p < .001$], a word length by item type interaction showed that the difference between target and foil repetition was larger for 6-syllable versus 3-syllable sequences [$\beta = -.25$, $z = -2.86$, $p = .004$]. Subsequent analyses revealed that trained word repetition and performance on the 2AFC task were related, and that performance on both SL measures significantly increased with age (*pace* Raviv & Arnon, 2017). The results suggest that the extraction of co-occurring linguistic elements via SL can rapidly lead to memory traces that alleviate the seemingly overwhelming processing load caused by the significant amount of information in the speech stream (Christiansen & Chater, 2016).

References

- Christiansen, M.H. & Chater, N. (2016). *Behavioral and Brain Sciences*, 39, 1-72.
Gathercole, S. (2006). *Applied Psycholinguistics*, 27, 513-543.
Miller, G. & Isard, S. (1963). *Journal of Verbal Learning and Verbal Behavior*, 2, 217-228.
Raviv, L., & Arnon, I. (2017). *Developmental Science*, 21, e12593.
Saffran, J., Aslin, R., & Newport, E. (1996). *Science*, 274, 1926-1928.

ASPECTS OF SEMANTICS AND THEIR EFFECTS ON APHASIC PICTURE NAMING

Leonie Lampe (Macquarie University, IDEALAB), Solène Hameau (Macquarie University), Nora Fieder (Macquarie University, Humboldt-Universität), Lyndsey Nickels (Macquarie University)
leonie.lampe@students.mq.edu.au

Introduction: In word production, both the target word and other semantically related words receive activation. Depending on the theory, these co-activated representations may impact the production of the target word. However, the nature of semantic representation and processing remains unclear and therefore, which semantic factors influence lexical activation. Semantic feature norms (e.g., McRae et al., 2005) permit the calculation of several new semantic variables, which relate to aspects of the representation of meaning of a target word. However, to date, few studies have investigated their effects on word production, particularly in a standard picture naming paradigm rather than a semantic interference task (e.g. picture-word interference). This study presents the first simultaneous examination of effects of six semantic variables on picture naming in people with aphasia.

Methods: Picture naming accuracy and response types of 175 people with aphasia were retrieved from the Moss Aphasia Psycholinguistic Project Database (MAPPD; Mirman et al., 2010). 89 MAPPD items were selected that also appeared in the McRae et al. (2005) feature norm database and in the University of Florida Free Association Norms (Nelson et al., 2004). McRae et al.'s database was used to derive the feature-based semantic variables: *number of near semantic neighbours*, *number of semantic features*, *typicality*, *distinctiveness*, and *intercorrelational density*. *Strength of the first associate* was derived from Nelson et al..

Results: Generalised Linear Mixed Effect Models with length, frequency, age of acquisition, familiarity, visual complexity, imageability, and name agreement as control variables were used to determine the effect of the semantic predictor variables on naming accuracy. Moreover, we tested for interactions between the semantic variables and participant's semantic abilities (scores on the Pyramids and Palm Trees test, PPT score; Howard & Patterson, 1992).

Naming accuracy increased for words that were shorter ($p < .001$), higher in frequency ($p = .012$), acquired earlier in life ($p = .005$), and more familiar ($p = .039$) as well as for participants with better semantics (higher PPT score; $p < .001$). Surprisingly, none of the semantic variables predicted naming accuracy, with only a marginal effect of number of semantic features ($p = .094$; increased accuracy for items with many features). Bayesian correlations between the semantic variables and naming accuracy corroborated the null effects: $BF_{01} > 3$ for all variables, except for strength of the first associate ($BF_{01} = 1.370$) and number of semantic features ($BF_{01} = 0.357$), where we could not adjudicate between H_1 that there was an effect on naming accuracy and H_0 that there was no such effect. Moreover, there were no significant interactions with PPT score.

Analyses will also be reported for the effects of the semantic variables on different error types for the whole sample, as well as for a subgroup of participants with semantic impairments.

Discussion: In contrast to previous research, we found no reliable effects of the semantic variables on picture naming in aphasia. Importantly, previous publications used smaller samples of participants, and few studies took into account both individual patient variation and a large number of control variables, which might have distorted their results.

Our results suggest that conceptualisation and lexical selection processes during word production are largely unaffected by these semantic variables. Hence, the semantic factors investigated here cannot be processed to a degree that affects word production. However, there was some evidence that a larger number of semantic features may increase activation of the target's lexical representation, resulting in higher naming accuracy.

References: Howard & Patterson (1992). Thames Valley Test Co.; McRae et al. (2005). *Behav Res Methods*, 37, 547–559; Mirman et al. (2010). *Cogn Neuropsychol*, 27, 495–504; Nelson et al. (2004). *Behav Res Methods*, 36(3), 402–407.

Thursday Poster.42

**DELIBERATIVE PROCESS IN SHARING INFORMATION WITH DIFFERENT
AUDIENCES: EYE-TRACKING CORRELATES**

Beatriz Martín-Luengo (National Research University - Higher School of Economics), Yury Shtyrov (Aarhus University, National Research University - Higher School of Economics) & Andriy Myachykov (Northumbria University, National Research University - Higher School of Economics)

bmartinluengo@hse.ru

People tailor the information they share depending on the target audience. When questioned about difficult general knowledge in informal contexts, interlocutors prefer to provide the answers that include more information (*plural answers* with several alternatives), and their willingness to report these answers is high. In formal contexts, specific answers (*single answers* including one alternative) are preferred. Also, the amount of withheld answers typically increases in formal contexts. In other words, the answers in different social contexts vary as a function of an informativeness-accuracy trade-off in relation to the specificity of the answers and the willingness to report them.

Here, we used eye-tracking methodology to further investigate the differential decision making process underlying the answer choice in formal and informal social contexts. It has been proposed that the eye movements preceding decision making are indicative of the underlying cognitive processes. Participants were asked to answer difficult general knowledge questions. The social context for each question was presented (formal vs. informal). Following this, participants were asked to decide on the specificity of their answers (single or plural) and whether they wanted to report or withdraw their selected answer in the suggested context. We analyzed their gazes directed to the visually presented answer options and compared the eye-tracking data between contexts.

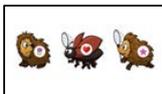
Growth curve analysis of the participants' gaze fixation patterns related to the answer choice revealed the following. In informal contexts, there were more fixations to both, single and plural reported options with the differences in the linear and cubic terms indicating an early discard of the withheld options. In formal contexts, there were more fixations in the single reported and the plural withheld conditions with participants also fixating their preferred choices earlier. Overall, our results offer a new perspective on the informativeness-accuracy trade-off in socially grounded communication: While in formal contexts we might have a pre-established type of answer, in the informal contexts we consider which information would be more valuable for the interlocutors.

Thursday Poster.43

(Specifically) Language-impaired Processing of Relative Clauses in German

This study examines how German-speaking children process subject (SR) and object (OR) relative clauses with an embedded full noun phrase (NP) or a 3rd person pronoun (pro; cf. Table). To date, there is only sparse evidence on the impact of embedded 3rd person pronouns on relative clause comprehension. While a pronoun is expected to facilitate OR processing, as compared to ORs with a full NP, the impact of embedded pronouns on SRs is less clear. Structural intervention accounts predict no effect of the NP vs. pronoun in SRs (Belletti et al., 2012). By contrast, discourse-based accounts predict a processing disadvantage in SRs with an embedded pronoun (Mak et al., 2008). Importantly, severe difficulties documented cross-linguistically make relative clauses a likely clinical marker of Specific Language Impairment (SLI; Frizelle & Fletcher, 2014). With respect to German, it is known that children with SLI produce fewer fully-fledged relatives than their typically developing controls. However, employing an implicit receptive measure such as eye-gazes has the potential to reveal whether children with SLI are nonetheless able to process relative clauses in a qualitatively similar way to typically developing children or not.

Towards this end, 74 monolingual German-speaking children, divided into three groups based on their developmental profile (SLI group, typically developing Language-Match (LM) controls and chronological Age-Matched (AM) controls), participated in an eye-tracking-while-listening experiment. Participants watched an animated visual display while listening to one of four experimental conditions (cf. Table). Their eye-gazes towards the three animals were recorded, and the Proportion of Looks to the Target animal (PLT, reported in the Table along with 95% Confidence Intervals) in the 1100 ms-long temporal region (starting from sentence offset) was taken as a measure of accurate processing.

Experimental conditions	Visual display	PLT (95% CI)		
		LM N=30; 6yo	SLI N=15; 7yo	AM N=29; 7yo
SR_NP e.g., Wo ist der Igel, der <u>den Käfer</u> fängt? (Where is the hedgehog that tickles the beetle?)		.38 (.07)	.37 (.08)	.38 (.04)
SR_pro e.g., Wo ist der Igel, der <u>ihn</u> fängt? (Where is the hedgehog that tickles him?)		.27 (.07)	.29 (.1)	.23 (.07)
OR_NP e.g., Wo ist der Igel, den <u>der Käfer</u> fängt? (Where is the hedgehog that the beetle tickles?)		.21 (.05)	.32 (.08)	.41 (.07)
OR_pro e.g., Wo ist der Igel, den <u>er</u> fängt? (Where is the hedgehog that he tickles?)		.27 (.07)	.30 (.07)	.39 (.07)

The analysis of the eye-gaze data, performed using Bayesian linear mixed models, indicates a processing disadvantage in SRs when the embedded object is a pronoun, rather than a full NP. This pattern is present in all groups, but it is less pronounced in the SLI group. As predicted by the discourse-based approach, the difficulty in interpreting a pronoun as direct object/patient in SRs may stem from the pronouns' discourse-pragmatic properties. Pronouns are typically used to refer to given entities, hence privileging a subject interpretation. Children with SLI reveal a weaker sensitivity to the pragmatic requirement on the use of pronouns, a difficulty that goes beyond their hitherto documented morpho-syntactic deficit. As for ORs, the LM group shows the expected pronoun facilitation, whereas the AM group identifies the target in both OR types alike. By contrast, children with SLI show a similar proportion of target identification in both OR types, but overall lower than the AM group. These results reveal an atypical trajectory of SR and OR processing in children with SLI.

Belletti et al. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*.

Frizelle & Fletcher (2014). Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders*.

Mak et al. (2008). Discourse structure and relative clause processing. *Memory & Cognition*.

What Use for Wrong Guesses? Disconfirmed Predictions Boost Novel Word Learning

Chiara Gambi (Cardiff University), Martin Pickering (University of Edinburgh) & Hugh Rabagliati (University of Edinburgh)
gambic@cardiff.ac.uk

The current debate on disconfirmed predictions focuses on whether they lead to immediate processing costs. What is typically not discussed (though see [3]) is the effect of disconfirmed predictions on the state of lexical representations in memory. Inspired by prediction-based accounts of memory [2], we ask whether we can improve memory for a novel word-object association by leading the comprehender to generate more precise linguistic predictions, which are then disconfirmed; the more precise the prediction, the larger the prediction error when disconfirmed, leading to a stronger memory representation.

To investigate the role played by linguistic prediction in shaping lexical representations, we created a new web-based paradigm. In Experiments 1-3 (combined N = 138), participants first listened to 8 unique non-words (e.g., *cheem*) within sentences that were more or less predictive of a familiar word (e.g., *Peppa will eat the...* is more predictive of *apple* than *Peppa will get the...*), while they looked at a depiction of the familiar word (apple) and at a novel object. On filler items (4) the sentence ended with a familiar word instead. Participants mostly mapped the novel word onto the novel object (trials on which they failed to do so were excluded); i.e., they followed mutual exclusivity [1]. Predictive sentences should generate higher prediction error (as participants expect *apple* more strongly). After a short interval (~ 5min), we asked participants to *Pick the cheem* out of an array of three novel objects (chance: 33%). We then compared the likelihood of selecting the correct object (retention accuracy) as a function of the predictive power of the verb used during the learning phase, using mixed-effect logistic regression. Provided they had selected the novel object during learning, adults were more likely to remember words learned under conditions of higher prediction error (combined analysis: $B = .61$, $SE = .22$, $z = 2.79$, $p < .01$; see Fig. 1A). The three experiments differed minimally in testing mode (lab or online) and the task used to fill the post-learning interval, and yielded the same pattern of results.

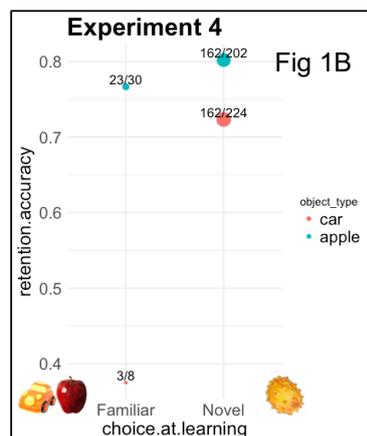
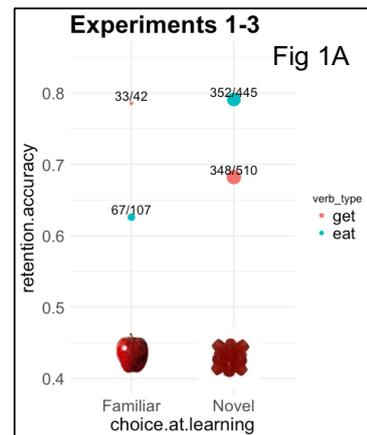
However, predictive verbs tend to have richer semantics than non-predictive verbs, and may thus provide more distinctive memory cues regardless of prediction error. To rule out this alternative explanation, in Experiment 4 (N = 58) participants always heard predictive verbs (e.g., *eat*); instead, we varied predictability by presenting either a depiction of the predictable familiar word (apple, as in Exp. 1-3) or of a different, unpredictable familiar word (e.g., car) alongside the novel object. Again, participants were more likely to remember words learned under conditions of higher prediction error ($B = 1.30$, $SE = .62$, $z = 2.10$, $p < .05$; see Fig. 1B).

Our findings suggest that the process of disconfirming a linguistic prediction leads to the formation of stronger lexical representations for a new word. An interesting question for future research is whether this effect extends to known words. In any case, we have shown that disconfirmed linguistic predictions play an important role in shaping lexical representations.

[1] Golinkoff, R., Hirsh-Pasek, K., Bailey, L., & Wenger, N. (1992). *Dev. Psych.*, 28, 99-108.

[2] Henson, R. N., & Gagnepain, P. (2010). *Hippocampus*, 20, 1315-1326.

[3] Rommers, J., & Federmeier, K.D. (2018). *Cortex*, 101, 16-30.



**PREDICTION ERRORS DUE TO TENSE BIASES DO NOT AFFECT
STRUCTURAL PRIMING**

Roger P.G. van Gompel (University of Dundee)
r.p.g.vangompel@dundee.ac.uk

Traditionally, structural priming models have assumed that structural priming is due to residual activation of the prime structure or of cognitive procedures involved in producing this structure (e.g., Bock, 1986; Pickering & Branigan, 1998). In contrast, more recent error-based learning models assume that structural priming is due to (implicit) learning from prediction errors (Chang et al., 2006; Jaeger & Snider, 2013). Comprehenders predict the upcoming structure of a prime sentence, and when this prediction is incorrect, the activation of the incorrectly predicted structure is reduced. Thus, structural priming should be stronger when a structure is less predictable.

Until now, the strongest evidence for error-based learning has come from studies showing that priming is stronger when a structure occurs infrequently with its head verb than when it occurs frequently with it (Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2013). However, head-centred residual activation models may be able to explain these results by assuming that the amount of activation that a single prime contributes is proportional to the total number of times a language user has come across that structure with that head verb.

We investigated priming of transitive/intransitive structures (Van Gompel et al., 2012) and manipulated their predictability by using either a past progressive or pluperfect tense. Participants read the primes aloud and provided a spoken completion to the targets:

- Intransitive prime: When the doctor {was supervising, had supervised}, he was very kind.
Transitive prime: When the doctor {was supervising, had supervised} the trainees, he was very kind.
Target: Although the teacher {was supervising, had supervised}

Frazier et al. (2006) showed that the processing of transitive/intransitive ambiguities was affected by the verb's tense. They argued that the direct object in a transitive sentence provides an end point to the sentence. Progressives denote an ongoing activity and therefore do not need an endpoint, whereas non-progressive past tense can be interpreted as an event and a direct object provides an endpoint to it.

Consistent with Frazier et al., a pretest in which the primes were cut off after the verb (*supervising/ed*) showed more transitive completions after a pluperfect tense and more intransitives after a progressive tense. Thus, if comprehenders use the tense of the verb to predict whether the following structure is transitive or intransitive, transitive priming should be stronger with progressive tense and intransitive priming with pluperfect tense. Head-centred residual activation models that assume that structural activation is not stored separately for different tenses (Pickering & Branigan, 1998) predict no effect of tense.

We observed clear structural priming, with more transitive target completions following transitive (65%) than intransitive primes (40%). There were also more transitive completions when the target verb was in the pluperfect (63%) than progressive tense (42%). However, priming was no stronger when the prime structure was unpredictable (24% priming, transitive in progressive tense, intransitive in pluperfect tense) than when it was predictable (25%, transitive with pluperfect, intransitive with progressive). Thus, there was no evidence for learning from prediction errors caused by verb tense biases. Instead, the results support residual activation models, in which predictability of structures plays no role.

**DOES NOUN CAPITALIZATION HELP?
AN EYE-TRACKING STUDY ON GERMAN**

Margreet Vogelzang, Esther Ruigendijk, Tobias Mundhenk & Nanna Fuhrhop
(University of Oldenburg)
margreet.vogelzang@uni-oldenburg.de

German is unique in its use of noun capitalization. It has been suggested that capitalization as in German may have processing benefits; by specifically marking a certain part of speech (here N as the head of the Noun Phrase), processing of syntactic structure is facilitated (cf. Günther, 1998; Günther & Nünke, 2005). There are however additional cues that indicate a noun phrase. The most common one is a determiner. The influence of capitalization on processing may thus be context-dependent, i.e. dependent on this other cue. Precisely this context-dependency is investigated in the current study: is there an effect of capitalization on reading and is this affected by the presence of other cues such as an article, meaning that ability to predict an upcoming noun differs with respect to structure?

We ran an eye-tracking study with 30 participants, measuring fixations during sentence reading. Sentences either contained a correctly capitalized noun or not, and were presented either with or without a determiner (see Table 1).

First pass duration is used as measure of reading time. Results show that on the critical noun, incorrect decapitalization lengthens reading time (263ms vs. 244ms, $p < 0.05$), but the presence of a determiner does not influence reading time. When examining the critical noun and the spillover region (verb) together as one region of interest, an interaction of capitalization and determiner emerges, with the no-determiner + incorrect decapitalization condition requiring longer to read (573ms vs. 508, 518, and 519ms, $p < 0.05$, see Fig. 1).

Concluding, evidence was found that the influence of capitalization is context-dependent. This indicates that capitalization indeed aids processing, but only then, when no other cue is present: both the presence of capitalization on N and the presence of a determiner led to lower reading times than when neither was present. The finding that reading times did not increase when decapitalization was combined with a determiner indicates that decapitalization does not cause an overall surprise effect, and so the longer reading times for incorrect decapitalization without a determiner reflect effects related to the processing of a category (N). Thus, signaling a noun phrase seems to be important for processing. Based on these first results, further experiments with more complex constructions (complex noun phrases, other uses of articles) will be done.

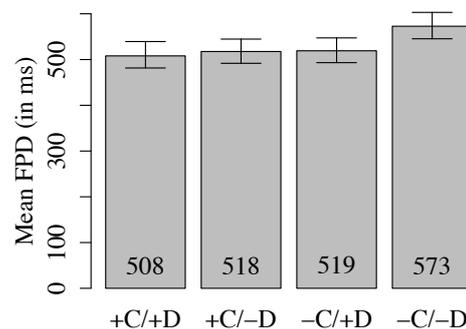


Figure 1. Mean first pass duration (FPD) on the critical noun and the spillover region

Cap	Det	Sentence
1	1	Morgen wollten sie <i>die</i> Kerzen herstellen, obwohl...
1	0	Morgen wollten sie Kerzen herstellen, obwohl...
0	1	Morgen wollten sie <i>die</i> kerzen herstellen, obwohl...
0	0	Morgen wollten sie kerzen herstellen, obwohl...

Tomorrow they wanted to make (the) candles, although...

Table 1. Example test item in German in each condition, with Cap = Capitalization, Det = Determiner. Test sentences consisted of two clauses each; a critical main clause and a dependent clause. The main clause varied only on the basis of the experimental manipulations. The dependent clause varied for every sentence manipulation.

Thursday Poster.47

SPOKEN IDIOMS ARE DECOMPOSED: EVIDENCE FROM ERP AND EYE TRACKING

Ruth Kessler (University of Tuebingen) & Claudia K. Friedrich (University of Tuebingen)

ruth.kessler@uni-tuebingen.de

It is a matter of debate, whether idioms, such as “to kick the bucket”, are represented as chunks in the mental lexicon. If so, lexical processing would not necessarily involve decomposition of idioms into single words. Event-related potentials (ERPs), on the one hand, confirmed this assumption: ERPs for words with semantic relation to an idiom constituent did not differ from ERPs for unrelated words (Rommers, Dijkstra & Bastiaansen, 2013). Eye fixation data recorded in a visual world study, on the other hand, revealed that participants consider distractor words with semantic relation to individual idiom constituents (Holsinger, 2013). Both studies differed in the modality in which participants received the idioms (written in the ERP study, spoken in the visual world study). Here, we investigate the processing of spoken idioms in an ERP study (Experiment 1) and in a visual world paradigm (Experiment 2). In both experiments, participants listened to highly predictive idioms (e.g. *to let the cat out of the bag*). We measured whether the expectation of the idiom final word (*bag*) would also lead to co-activation of semantically related words (*basket*) compared to unrelated words (*stomach*). In Experiment 1, participants (N = 40) listened to idiomatic phrases while EEGs were recorded. The phrase final word was either the correct idiom final word, the semantically related word or the unrelated word. In a N400 time window (300-500ms), correct and related words showed reduced amplitudes compared to unrelated words. In a later time window (700-1000ms), only ERPs for correct words, but not for related words, differed from unrelated words. Thus, semantic relatedness seemed to affect the N400 amplitudes, but not late positivity. In Experiment 2, participants (N = 31) listened to idioms up to the final word while written versions of the correct completion, the related distractor and unrelated distractors appeared simultaneously on a computer screen. Participants fixated related distractors more often than unrelated distractors already early during the anticipation of the correct completion, but not during later processing stages. Taken together, both experiments give evidence for early, automatic activation of single semantic features of idiom constituents, but also chunk-like access during later processing. It appears that at least in spoken idiom processing, chunks are at least not solely determining lexical processing.

Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech*, 56(3), 373-394.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762-776.

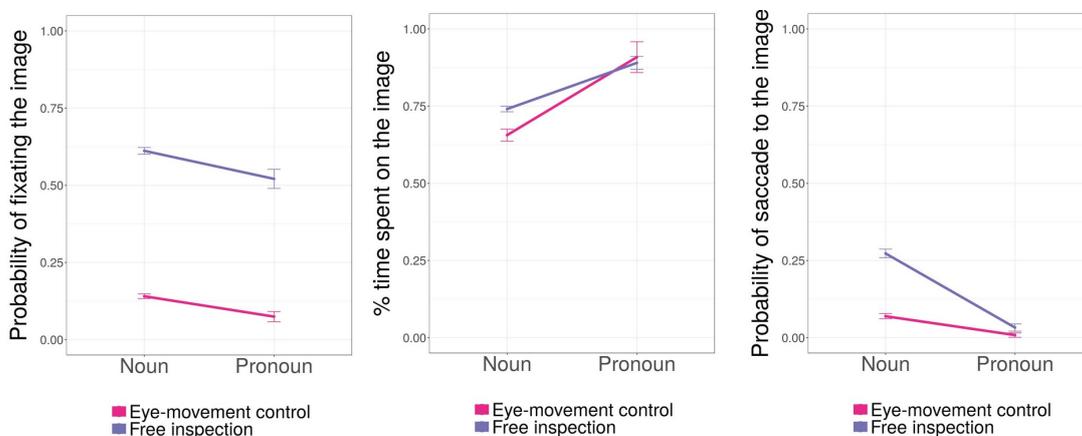
Thursday Poster.48

EYE-MOVEMENT CONTROL IN THE VISUAL WORLD PARADIGM

Anna Laurinavichyute (Higher School of Economics, Moscow; University of Potsdam),
 Anastasiya Lopukhina (Higher School of Economics, Moscow)
 anna.laurinavichyute@uni-potsdam.de

The presumably widely shared, although not often explicitly articulated assumption behind the use of the Visual World paradigm is that participants tend to *automatically* look at the object that undergoes linguistic processing at the current moment. The assumption is based on close correspondence in time between the onset of the reference to the object/action and fixations to its depiction. However, close correspondence in time does not necessarily imply automaticity, and lack of automaticity would affect the interpretation of experimental results: if language-mediated eye-movements are not automatic, then the absence of fixations on the object cannot be seen as absence of linguistic processing of the corresponding word. In two experiments, we tested the automaticity of language-mediated eye movements by probing whether they could be canceled by volitional control. In the *eye-movement control* experiment, participants (N=39) were asked to not look at the object that is currently being referred to. In the *free inspection* experiment (N=41), participants had a classical 'look and listen' task. Experiments included the same 32 stories without comprehension questions and differed only in the task. In addition, we aimed to test whether different types of references – noun-object and pronoun-object – invoke different patterns of eye-movement responses, and whether they are differentially affected in the eye-movement control experiment.

In the eye-movement control experiment, the probability of at least once fixating the referred image and the number of incoming saccades to that image were lower ($\hat{\beta}=-1.45$, $p<.0001$; $\hat{\beta}=-.85$, $p<.0001$). When fixated, there was a difference between the proportion of time spent on the images referred to by nouns and pronouns: for nouns, less time was spent on the image in the eye-movement control experiment ($\hat{\beta}=-.096$, $t=-5.37$), no such difference was found for pronouns. Finally, probability of fixating an image referred to by a pronoun and the number of incoming saccades to those images were lower in comparison to nouns ($\hat{\beta}=-0.32$, $p<.001$; $\hat{\beta}=-1.06$, $p<.0001$). We found that the type of reference plays a role in the visual world paradigm: images referred to by pronouns tend to be fixated less often across the board. Most importantly, although participants were generally able to control their eye-movements, in 7% of cases they were still executing saccades to the image currently being named. Whether every participant was as effective in suppressing fixations, or some were more so than others, is a question for future research.



LEXICAL BOOST FROM THE MATRIX VERB

Laura Wakeford & Roger Van Gompel (University of Dundee) & Leila Kantola (Umeå University)
ljzwakeford@dundee.ac.uk

Speakers tend to repeat a recently encountered syntactic structure rather than constructing one anew. Studies have shown that this *structural priming*, from a prime to the target sentence, is enhanced if the prime and target share lexical content; a phenomenon termed the *lexical boost* (LB). Models differ in how they account for these effects. In the *Residual Activation Model* (Pickering & Branigan, 1998) the LB is caused by the lingering activation of the link between the licensing syntactic head and its argument structure in the prime. Other models suggest that the LB is not confined to the repetition of a syntactic head (e.g., Chang et al., 2006; Reitter, et al., 2011). For example, in the *Implicit Learning Model* (Chang et al., 2006), structural priming in the absence of word repetition is due to implicit learning, whereas the LB originates from an explicit memory trace of the repeated word and the structure with which it occurred in the prime.

Contrary to the Residual Activation Model, a LB has been reported for repeated words other than the verb (Scheepers et al., 2017; see also McClean et al., 2004), leading Scheepers et al. to suggest that the LB is not bound to repeating the verb licensing the argument structure. They found a LB from words that were arguments of the verb in the same clause as the primed structure, which raises the possibility that a LB occurred because these words were activated simultaneously with the primed structure. We investigated whether a LB occurs when a word outside the clause with the primed structure is repeated by manipulating the repetition of a matrix verb that dominated the clause with the primed structure.

In **Experiment 1**, 112 native speakers provided written completions to 40 experimental prime-target sentence pairs. We manipulated (1) whether the prime encouraged a prepositional object (PO) or a double object (DO) completion and (2) matrix verb (MV) repetition. For example:

	Matrix verb	PO fragment	DO fragment
Prime:	The painter {hesitated,decided} to show	{the long ladder... ,	the eagerly apprentice...}
Target:	The colleague {hesitated,decided} to show ...		

A higher proportion of PO completions followed a PO prime than DO prime (60% vs. 46%; $p < .001$). Critically, the priming effect was stronger following MV repetition (18%) than without (9%), qualified by an interaction ($p < .05$).

In **Experiment 2**, forty participants read aloud 40 prime-target sentences in a comprehension to production task. Primes were full sentences that primed either a transitive or intransitive structure (i.e., the head verb of the primed subordinate clause either took an explicit object, or did not); as above, MV was also manipulated. For example:

	Matrix verb	Intransitive	Transitive
Prime:	The ambulance driver {longed,attempted} {to approach. ,	to approach the accident.}	
Target:	Although the patient {longed,attempted} to chew...		

Critically, there was a LB for intransitive primes (24% vs. 29% intransitive completions; $p < .05$), but not transitive primes (69% vs. 69% transitive completions; $p = .74$). This suggests that the transitive structure, being the most frequent structure in English, is not associated with specific words in the sentence. This is in line with Van Gompel et al. (2012) who reported a LB with head verbs following an intransitive prime structure, but not a transitive one.

We suggest that the LB does not just occur when the repeated word is an argument of the verb in the same clause as the prime structure, but also with words outside this clause. The LB reported here with the matrix verb may occur because it acts as a retrieval cue and activates a memory trace (e.g., Chang et al., 2006), or because the matrix verb activates its argument structure which includes the primed structure.

**AGAINST REACTIVATION OF SYNTACTIC TRACES IN FILLER-GAP DEPENDENCIES
IN THE VISUAL WORLD PARADIGM**

Anna Laurinavichyute (Higher School of Economics, Moscow; University of Potsdam), Olga Dragoy, Maria Khudyakova (Higher School of Economics, Moscow), Irina Sekerina (City University of New York)
anna.laurinavichyute@uni-potsdam.de

According to the *active filler strategy* [1], in filler-gap dependencies, the human parser postulates a gap immediately at encountering a filler. Previous studies of trace reactivation in English *wh*-questions in the Visual World paradigm demonstrated increased fixations on the gapped object at and after the gap site, presumably reflecting completion of the dependency [2, 3]. However, the studies conducted in English contain a confound: the gapped object is always the answer to the *wh*-question; therefore, it is impossible to disentangle the reactivation of the trace from the pragmatically oriented strategy of looking at what seems to be the answer [4].

In a Russian Visual World experiment ($N=40$), we employed embedded relative clauses instead of questions to decrease the pragmatic pressure of answering the question. We contrasted *wh*-movement (1-2) with scrambling (3-4), which allowed us to disentangle the moved/scrambled object and the referent of *who*. We tested 32 items (mixed with 32 filler stories) in 4 conditions. The three animate protagonists (*boy, girl, teacher*) and a location (*school*) were depicted for every item in the 4 corners of the screen. Participants received a 'look and listen' task, no comprehension questions were asked.

On Monday a boy and a girl walked past the teacher. Suddenly the boy₁ pushed the girl₂, which surprised the teacher₃. He expelled both from school₄. Nobody realized,

- | | |
|----------------------------------------------------------------------------------|-------------------------------------|
| (1) kogo ₂ malchik v ponedel'nik tolknul _{_2} – | who-ACC boy-NOM (on Monday) pushed_ |
| (2) kogo ₃ malchik v ponedel'nik udivil _{_3} – | who-ACC boy-NOM surprised_ |
| (3) kto ₁ devochku ₂ v ponedel'nik tolknul _{_2} – | who-NOM girl-ACC pushed_ |
| (4) kto ₃ devochku ₂ v ponedel'nik vygnal _{_2} – | who-NOM girl-ACC expelled_ |
- ... at/from school.

We analyzed proportion of fixations to the object, contrasting conditions (1-2) where the gapped object coincides with the referent of *who*, and (3-4) where the scrambled object is not the referent of *who*: **wh-operator** (*kogo/kto*) – no difference; **agent** (*boy*) in (1-2) / **scrambled NP** (*girl*) in (3-4) – proportions of fixations to the gapped object increased across all conditions; **temporal adverbial** – increase in fixations to the scrambled NP (*girl*) in (3-4), probably because it was named in the previous region; **verb** (+gap) – proportion of fixations to the gapped object increased in (1-2), while the proportion of fixations to the scrambled object decreased in (3-4); **location**: the trends from the verb region progressed.

While these results are compatible with the active filler strategy for the *wh*-questions (1-2), they do not support reactivation of the scrambled NP at the verb in (3-4) as we observed a significant decrease in its activation. Two explanations are possible. First, the scrambled NP, in contrast to the gapped object, does not leave a trace, hence no reactivation. We argue for the second, more parsimonious, explanation: Eye movements in the Visual World paradigm reflect pragmatic rather than syntactic processing, i.e., participants seek to establish referential relationships and thus fixate the referent of *who* regardless of the dependency type. The latter explanation is in line with the goal-based linking hypothesis account of the eye movements in the Visual World paradigm [4].

[1] Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Nat Lang Linguist Theory*, 5(4), 519. [2] Dickey, M.W., Choy, J.J., & Thompson, C.K. (2007). Real-time comprehension of *wh*-movement in aphasia: Evidence from eyetracking while listening. *Brain Lang*, 100(1), 1. [3] Thompson, C.K., & Choy, J.J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *J Psycholinguist Res*, 38(3), 255-283. Salverda, A.P., Brown, M., & Tanenhaus, M.K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychol*, 137(2), 172.

**HIGH VERB FREQUENCY AS AN ACCESSIBILITY PARAMETER
PROMOTING EARLY VERB PLACEMENT
IN MAIN CLAUSES IN THREE SEMI-FREE WORD ORDER LANGUAGES**

**Gerard Kempen (Max Planck Institute for Psycholinguistics, Nijmegen)
Karin Harbusch (Computer Science Faculty, University of Koblenz-Landau, Koblenz)**

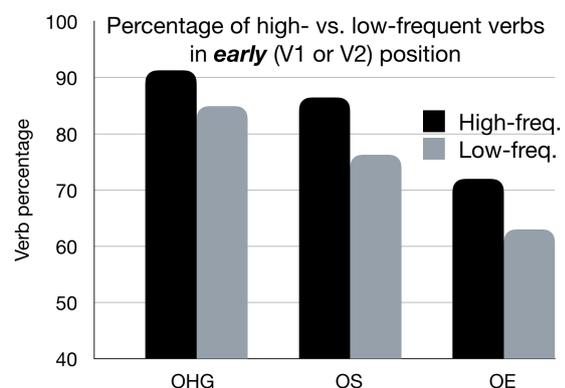
An important determinant of constituent order in clauses is CONCEPTUAL AND LEXICAL ACCESSIBILITY (or AVAILABILITY) of the constituents. Empirical evidence in support of this effect is virtually restricted to NOMINAL constituents, i.e. to nouns as heads of NPs. Many studies since Osgood & Bock (1977) have shown that speakers tend to assign early positions to NPs headed by highly accessible nouns. For instance, NPs referring to animate entities or to the current discourse topic are more likely to receive an early position as subject than inanimate NPs or NPs belonging to the comment. Conspicuously absent are data on effects of accessibility of VERBAL clause constituents: finite full verbs, auxiliary verbs, and modal verbs. The reason must be the immovable position of verbs in many modern languages: Demonstrating accessibility effects on linearization presupposes a minimum of linearization flexibility. In order to close this gap, we turned to Old High German (OHG), Old Saxon (OS; one of the ancestors of modern Dutch), and Old English (OE). In these languages, which were spoken and written before circa 1100 CE, the position of finite verbs is considerably more variable than in their present-day counterparts. A practical reason for selecting these languages is the availability of substantial syntactically analyzed text corpora (“treebanks”): YCOE for OE (Taylor 2007); B4-TATIAN for OHG (Petrova et al. 2009), and HELIPAD for OS (Walkden 2016). The number of different verbs (lemmas) occurring in these treebanks varies considerably, due to widely differing corpus sizes: ~350 different verb lemmas in the OHG corpus, ~650 OS, and ~5000 OE lemmas. We calculated the total frequency of a verb lemma as the sum of the frequencies of all its (un)inflected forms in all spelling variants.

Finite verbs in MAIN clauses of the target languages can be linearized as first (“V1”), second (“V2”) or later (“V>2”) constituent. (We leave subordinate clauses out of consideration here because the three languages impose differing requirements w.r.t. the presence of introductory subordinating conjunctions.) In order to arrive at a dichotomous variable for linear position, we combined—somewhat arbitrarily—the V1 and V2 cases into a group called “early”, and renamed the V>2 group “late”. Likewise, in each treebank, we dichotomized the frequency spectrum into a High-frequent group comprising the 50 lemmas with highest total frequencies, and a Low-frequent group containing the remaining verb lemmas.

As shown in the figure, high lemma frequency indeed promotes early verb placement (black bars higher than grey bars). The mean placement difference between high- and low-frequent verbs is significant in each treebank (χ^2 tests with $df=1$, all yielding $p < .005$, 1-sided). (Discussion of the between-language differences is beyond the scope of this presentation.)

This result confirms the hypothesis that verb frequency functions as an accessibility parameter capable of exerting effects similar to noun accessibility. High-frequency verbs represent frequently used concepts and lexical entries, which we assume have a higher residual (resting) level of activation. Hence, the speaker can retrieve these verbs from the mental lexicon more easily and rapidly than infrequent verbs. This temporal advantage enables frequent verbs to land in ANTERIOR clause positions allowed by the grammar more likely than infrequent verbs.

The finding reported here is relevant not only for psycholinguistics but also for historical linguistics: The observed accessibility–anteriority link may have been a factor in the diachronic development of (S)OV–to–(S)VO main-clause word order in Germanic languages.



ABOUT THE IMPORTANCE OF IMPLICIT CAUSALITY TO UNDERSTAND RELATIVE CLAUSE PROCESSING

Céline Pozniak & Barbara Hemforth (LLF, CNRS, univ. Paris-Diderot, Labex EFL)
celine.pozniak@gmail.com

Implicit causality (IC) of verbs like “trouble” or “hate” foregrounds the subject or the object of a sentence, influencing discourse and syntax preferences (e.g. Kehler et al., 2008; Rohde et al., 2011). In offline and online experiments in French and English, we show that syntactic factors like structural/linear distance cannot explain processing of subject (SR) and object relatives (OR) alone, and that IC is an essential factor that needs to be taken into account.

As a reminder, the head of a restrictive relative clause (RC) is what the RC is about and what needs to be foregrounded. From studies on IC (e.g. Garvey and Caramazza, 1974; Kehler et al., 2008; McKoon et al., 1993), verbs can foreground the subject (NP1-biased: *Mary troubles Peter because she is pretty*) or the object (NP2-biased: *Mary hates Peter because he is mean to her.*) of a sentence by biasing their perception as probable causes of the eventuality. We predict that RC processing should become harder when the verb bias conflicts with the necessity of having the RC head foregrounded in particular for ORs, i.e. subject-biased verbs should make the comprehension of ORs difficult since the head is the object of the sentence. We ran acceptability judgments in French and English as well as an eye-tracking reading experiment in English, varying the verb’s IC bias (see table [1]). As for the choice of verbs, we used a recently established corpus on IC- biases of verbs for French (Mertz et al., in prep) and the corpus from Ferstl et al. (2011) for English where highly biased verbs include different thematic roles, e.g. stimulus/experiencer (“troubled”), agent/evocator (“applauded”). We manipulated the type of verbs (NP1-biased, NP2-biased) and the type of RCs (SRs and ORs).

Subject relative	NP1-biased	<i>The teacher that troubles the lawyer will not give classes next semester.</i>
	NP2-biased	<i>The teacher that hates the lawyer will not give classes next semester.</i>
Object relative	NP1-biased	<i>The teacher that the lawyer troubles will not give classes next semester.</i>
	NP2-biased	<i>The teacher that the lawyer hates will not give classes next semester.</i>

Table [1] Example of sentences used for the acceptability judgment task (only for English here)

Study 1 & Study 2

Study 1 presents acceptability judgments in French and English and confirms that IC bias influences RC processing across languages (74 French native speakers, 63 English native speakers; 20 items). ORs with NP1-biased verbs were judged significantly less acceptable than all other RCs both in French and in English ($p < .05$). An interaction of verb type and RC type was found ($p < .01$). *Study 2* shows a similar pattern in eye tracking while reading with the same materials, mainly at the main clause verb (21 English native speakers). Reading times for ORs with subject biased verbs are significantly higher in this condition than in all other conditions ($p < .05$). No significant difference was found between any of the other conditions.

Study 3

From the fact that IC bias only affected ORs, it might be concluded that syntactic factors could be at stake, meaning that IC bias is just a secondary factor enhancing effects of an intervening subject (DLT; Gibson, 2000 / Relativized Minimality; Rizzi, 1990). In French ORs, the subject can also be postverbal (*The teacher that_{obj} hates the lawyer_{subj}*), implying that the subject does not intervene in the filler-gap dependency. Acceptability judgments for the 2 types of French ORs, materials adapted from *Studies 1* and *2* (34 French native speakers), show the same effect of verb bias as the earlier studies ($p < .01$) and no interaction between verb type and subject position, meaning that the effect of verb bias applies equally to both ORs.

Conclusion

Our studies provide on- and off-line evidence that the foregrounding of the subject or the object coming from verb biases plays an important role in the comprehension of RCs independent of syntactic factors. This effect is robust across languages. ORs can be as acceptable and easy to understand as SRs with the right verb bias. This means that beyond possible syntactic factors or semantic factors such as head animacy (Gibson, 2000; Traxler et al., 2005, Vasishth et al., 2013), processing effects based on the function of the construction (attributing information about its head) are essential to the understanding of RC processing.

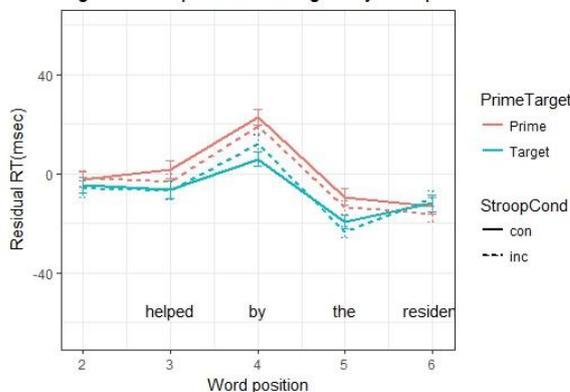
EXECUTIVE FUNCTION ADAPTATION AND SYNTACTIC PRIMING

Edith Kaan, Emma Leone & Yucheng Liu (University of Florida)
 kaan@ufl.edu

Executive Function (EF) adaptation is the observation that a difficult cognitive control trial, for instance, an incongruent Stroop task, can make it easier to process a following difficult task, such as processing a syntactically ambiguous sentence (Hsu & Novick, 2016). However, EF encompasses many functions, and it is not clear what function is adapted to. The present study further explores the effect of EF adaptation on sentence processing by interleaving Stroop trials with syntactic priming of reduced relative clauses (RR, see 1a). In RR sentences the first verb (“helped”) is ambiguous between a main clause verb and a non-preferred passive participle in a relative clause. The word “by” (in 1a) strongly biases towards the latter interpretation. Prior studies have shown that a reduced relative clause is easier to read when it is preceded by a sentence with a similar structure (Tooley & Traxler, 2008). If a preceding hard EF trial facilitates the processing of an RR prime by better inhibiting the unintended main clause reading, we would expect the processing of a following target RR (1b) to be even more facilitated after primes that are preceded by a hard versus an easy EF trial. In addition, a main clause (MC) target (1b’) will be expected to be harder to process in this case, since the inhibited structure needs to be reactivated. If, on the other hand, a hard EF trial leads to shallower processing because of resource depletion (Engle et al. 1995), RR priming effects should be smaller after a hard compared to an easy EF trial; the processing of the MC targets, being the preferred structure, should not be affected by the preceding EF task.

- 1a. RR Prime: The students helped by the counselor were grateful for the aid.
- 1b. RR Target: The surgeons helped by the resident were exhausted by the operation.
- 1b’. MC Target: The surgeons helped the resident revive the dying man on the cot.

Figure 1: RR primes and targets by stroop



We conducted a word-by-word moving window self-paced reading study in which 56 native English speakers read prime-target pairs of the type illustrated in (1): 36 with an RR target; 36 with an MC target; half of the pairs per prime-target type were preceded by a congruent Stroop trial, half by an incongruent. Materials were Latin Squared; the prime and target shared the verb to encourage priming effects. Additionally, 72 distractor items were included to equate the number of main clauses and reduced relative clauses in

the study and to vary the number of sentences intervening between the Stroop trials. As expected, RR target sentences were read faster than the RR primes at the disambiguating “by” position ($b=-12.01$; $SE=3.7$; $t=-3.3$). This priming effect was larger when the RR prime was preceded by a congruent than an incongruent Stroop trial ($b=10.5$, $SE=6.2$, $T= 1.7$, fig.1). In addition, the object noun in the MC targets (“resident” in 1b’) was read faster when the RR prime was preceded by a congruent than an incongruent Stroop trial ($b = 7.9$, $SE = 4.6$, $T= 1.7$). These results are not expected under the view that EF adaptation leads to a better inhibition of the initially preferred structure or to more superficial processing of complex structures. An alternative explanation is that an incongruent Stroop trial leads to sustained monitoring, resulting in longer reading times at the point of disambiguation in both types of target. The weakness of the effects may be due our version of the Stroop task being rather difficult. We are currently running a follow-up study with a more traditional Stroop task.

Thursday Poster.54

Active Antecedent Search in Cataphora Processing: Insights from Neural Oscillations

Ashley Glen Lewis (Haskins), Dave Kush (NTNU), Andrew Jahn (Haskins), Luca Campanelli (Haskins), Clinton Johns (Haskins) & Julie Van Dyke (Haskins)
lewis@haskins.yale.edu

Reading studies [1,2] suggest that participants can construct a backwards anaphoric (a.k.a., cataphoric) dependency between an unbound pronoun and a subsequent antecedent. Unlike forwards anaphora, processing cataphoric dependencies involves an *active search* for an antecedent *later* in the sentence. In the N (no constraint) example below, reading the pronoun *he/she* triggers a search for a potential antecedent, which it finds in the main subject *John*. Evidence for active search comes from gender-mismatch effects: a candidate antecedent NP is read more slowly when it mismatches the preceding pronoun in gender (there should be no mismatch effect if the parser were not actively searching for a potential coreferential relationship for the pronoun) [2]. Some researchers have argued that active search is sensitive to grammatical constraints [1]: the parser does not consider NPs as potential antecedents if they are in grammatically illicit positions. For example, the parser should never consider coreference between *He/She* and *John* in the C (constraint) example below, because coreference is ruled out by Principle C [3].

N: After **he/she** met the girl that was wearing orange pants by the store *John* jogged home.

C: **He/She** met the girl that was wearing orange pants by the store after *John* jogged home.

Initial evidence for a grammatically sensitive active search mechanism came from self-paced reading studies, which showed mismatch effects at NPs in grammatically licit positions, but not at positions ruled out by Principle C [1]. However, recent eye tracking work suggests that structural constraints are only used relatively late during processing to rule out illicit coreference [4]. For sentences like N and C above the critical interaction (mismatch effect at the potential antecedent *John* for N but not C) appears only in late eye movement measures [4]. This suggests that active search may initially be insensitive to grammatical constraints, indiscriminately positing coreference relations between unbound pronouns and subsequent matching NPs, some of which must later be filtered out based on grammatical constraints. Given that the critical interaction in [4] was only marginal the debate is far from settled. This study uses the time-course of neural signatures, which are sensitive to qualitative aspects of processing, to address the interplay between structural (top-down) and lexical (bottom-up) information during cataphora processing.

Neural oscillations in the beta frequency range (13-30 Hz) provide a signature of the maintenance (increase) or change (decrease) of the representation of the current sentence-level meaning [5] and are sensitive to grammatical gender violations [6]. Beta oscillations are therefore an excellent candidate to provide a window into the neurophysiological time-course of the kind of active search engaged during cataphora processing. We used electroencephalography (EEG) to investigate beta oscillations while participants (N = 24) read sentences like N (no constraint) and C (constraint) requiring active search for an antecedent. For each of these conditions the preceding pronoun either matched (MA) or mismatched (MI) the grammatical gender of the potential antecedent (target word) *John*.

Time-frequency analysis of power with source reconstructed data (beamformer approach) demonstrated an interaction between structural constraint (N/C) and gender (MA/MI) for low beta power (13-19 Hz) between 240-880 ms after the target word, [$F(1,23) = 7.65, p = 0.011$]. This was driven by lower beta power in the MI than the MA condition for the N sentence types [$t(23) = 3.21, p = 0.004$]. These effects were restricted to left frontal and temporal regions, most notably left inferior frontal gyrus, which is implicated in syntactic prediction [7]. Only for the MI condition in the C sentence types beta power increased early (~150-200 ms) after the onset of the potential antecedent, suggesting active maintenance of the sentence representation, possibly to rule out the formation of a coreferential relationship with the preceding pronoun. For the MA condition in the C sentence types the grammatical gender matches that of the preceding pronoun and the coreferential relationship cannot be ruled out this early, likely due to competition between bottom-up and top-down signals. This offers support for an early influence of structural constraints on active antecedent search.

References: [1] Kazanina et al., 2010; [2] van Gompel et al., 2003; [3] Chomsky, 1981; [4] Drummer et al., 2018; [5] Lewis et al., 2015; [6] Lewis et al., 2016; [7] Matchin et al., 2017.

Thursday Poster.55

THE STATISTICAL SIGNIFICANCE FILTER LEADS TO OVEROPTIMISTIC EXPECTATIONS OF REPLICABILITY

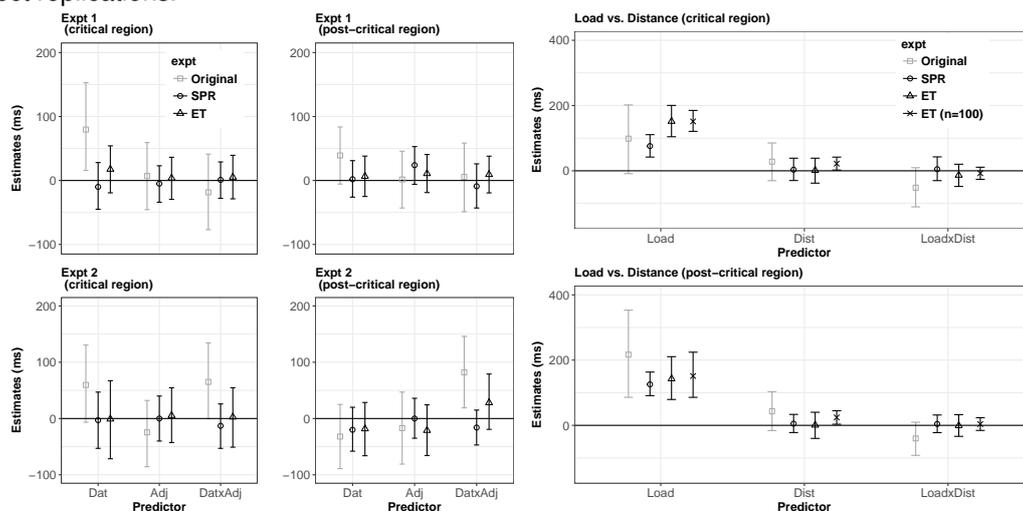
Shravan Vasishth (Potsdam), Daniela Mertzen (Potsdam), Lena Jäger (Potsdam), Andrew Gelman (Columbia)

Treating a result as newsworthy, i.e., publishable, because the p-value is less than 0.05 leads to overoptimistic expectations of replicability. The underlying cause of these overoptimistic expectations is Type M (magnitude) error [1]: when underpowered studies yield significant results, the effect size estimates are *guaranteed* to be exaggerated and noisy (have large standard errors). Type M error arises because the standard error in a low-powered study is large, which implies that the mean will fluctuate greatly under repeated sampling. Thus, under low power, significant effects will always be overestimates and are unlikely to replicate. We refer to the use of $p < 0.05$ as a criterion for publishability as the *statistical significance filter*.

These issues have been discussed repeatedly in psychological science since at least the 1970s [2], and many researchers today believe that awareness of these issues is widespread. Despite this rising awareness, even today top journals continue to publish underpowered studies. We therefore felt it would be useful to empirically demonstrate the adverse consequences of the statistical significance filter by carrying out direct replication attempts of published statistically significant results from a recent eyetracking reading study on locality and expectation effects in German [3]. A detailed discussion of the design and interpretation of [3] cannot be provided in this short abstract; for details, see the published paper [4] (<https://osf.io/eyphj/>).

We ran seven experiments (268 subjects in total) and could not replicate a single effect reported as significant in [3]. We also demonstrate the contrast between these small-sample studies and a larger-sample study (100 participants); the latter generally yields less noisy estimates but also a smaller magnitude of the effect of interest. The small magnitude looks less compelling but is more realistic.

Our repeated null results show that using the p-value as a publication criterion is of little value when estimates are noisy. We suggest that researchers and journals avoid focusing exclusively on statistical significance to evaluate the reliability of studies. Instead, the focus should be on obtaining accurate and replicable estimates. We suggest that researchers (i) carry out direct (not just conceptual) replications in order to demonstrate the existence of an effect; (ii) move their focus away from statistical significance and attend instead to reducing the standard error of their estimates (e.g., by increasing sample size, or improving quality of measurements), and (iii) pre-register their designs and planned analyses in venues like osf.io. Journals can encourage these practices by introducing a new short-article type, pre-registered direct replications.



[1] Gelman & Carlin, 2014. [2] Lane & Dunlop, 1979. [3] Levy & Keller, 2013. [4] Vasishth, Mertzen, Jäger, & Gelman, 2018 (osf.io/eyphj/).

Thursday Poster.56

INTEGRATION AND ANTICIPATION PROCESSES OF THE SPEAKER AND MEANING IN ADULTS WITH AND WITHOUT AUTISM SPECTRUM DISORDER: EVIDENCE FROM EYE-TRACKING AND ERPS

Mahsa Barzy (University of Kent), Jo Black (University of Kent), David Williams (University of Kent) & Heather J Ferguson (University of Kent)
mm951@kent.ac.uk

Evidence suggests that when comprehending language, people build mental models that include knowledge about the speaker. For instance, using event related potentials (ERPs), a semantic anomaly-type response (i.e. N400 effect) was observed when typically developing (TD) adults listened to sentences in which the content of the message mismatched with the voice of the speaker (van Berkum, et al., 2008). It is claimed that individuals with autism spectrum disorders (ASD) experience specific difficulties integrating information from the context to build pragmatic mental models while comprehending language (Happe, 1996). Therefore, we present two pre-registered experiments that examined whether adults with ASD exhibit comparable anticipation and integration processes for speaker and meaning as TD adults.

Experiment 1 employed the visual world paradigm, and tracked the timecourse of anticipatory biases about a speaker's meaning, based on their voice. Forty-eight participants (N=24 in each group, matched on gender, age and IQ) listened to sentences, in which the voice of speaker was either consistent or inconsistent with the intended message (e.g. "On my last birthday, I got an expensive electric shaver/car" in a child or an adult's voice), and concurrently viewed visual scenes that depicted these consistent and inconsistent objects alongside distractor objects. Participants were instructed to select the picture that best matched the audio description, and eye movements were recorded throughout. Behavioural results showed that for age and gender speaker types all participants were slower to select the correct object when it was inconsistent with the speaker's voice than when it was consistent. Eye-tracking results revealed a visual bias towards the object that was consistent with voice of the speaker group well before the disambiguating word onset. Hence, participants rapidly integrated the speaker's voice and used this to anticipate the content of forthcoming language. Growth curve analyses revealed a Group x cubic fit interaction, showing that the TD group fixated the target earlier than the ASD group (2240ms vs. 1800ms before disambiguation). The anticipatory bias in the TD group subsequently declined prior to a rapid increase (960ms before disambiguation), whereas the ASD group showed a consistent increase in target bias from 1800ms.

In Experiment 2 we recorded ERPs to explore how consistency between the speaker's voice and message influences integration processes. Forty-eight participants (N=24 in each group) listened to sentences of the same type as in Experiment 1 (e.g. "I tried to refresh my lipstick in front of the mirror" in a man or a woman's voice). EEG activity was recorded from 30 active electrodes, time-locked to the onset of the disambiguating target word, which was either consistent or inconsistent with the speaker's voice. A third sentence condition included a semantic anomaly (e.g. "I tried to refresh my seashell in front of the mirror"), and thus provided a baseline of anomaly detection effects on the N400 for comparison with speaker consistency effects. Results revealed an enhanced N400 for inconsistent sentences relative to consistent sentences, which was comparable to the N400 elicited by anomalous sentences. Further analyses revealed group differences in the topography of N400 effects, suggesting that different neural generators may be involved. Overall, these results show that contrary to previous suggestions of pragmatic dysfunction, people with ASD are sensitive to integration between speaker and meaning.

Thursday Poster.57

COMPREHENDING INFORMATION STRUCTURE BY KOREAN ENGLISH L2ERS: AN ERP STUDY

Wonil Chung (Dongguk University) & Myung-Kwan Park (Dongguk University)
wonilchung@naver.com

This paper investigates ERP responses to violations of information structure (IS) in answers to *wh*-questions in Korean English speakers, where focus structure is incorrectly aligned in 'it'-clefts. Cowles, Kluender, Kutas, and Polinsky (2007) found two types of ERP response in answers to *wh*-questions. One is that all words in focused position showed a large positivity understood as characteristic of sentence-final elements, and in fact the sentence-final words of the sentences containing them did. They suggested that focused elements might trigger integration effects, like those in sentence-final position. The other is that inappropriately-focused referents showed a right negativity. They suggested that this N400-like effect was elicited by comprehending structurally-encoded focus cues and discourse-level restrictions.

To this aim the experimental materials for our ERP study consisted of 60 sets with two types of stimuli (congruent and incongruent), adopted from Cowles, et al. (2007). Each trial contains a set-up context with the introduction of three discourse participants, and then a *wh*-question consisting of one participant as an agent and two participants as an undergoer of an event, and a target sentence that was constructed as an *it*-cleft, with its pivot marked for focus with a congruent or incongruent participant, schematically represented below.

- (1) set-up: Who did the queen silence with a word, the banker or the advisor?
- (2) congruent target: It / was/ the banker/ that/ the queen/ silenced/.
- (3) incongruent target: It /was/ the queen/ that/ silenced/ the banker/.

Twenty Korean English L2ers (13 males, mean age: 23.7, SD: 1.7) with a high level of English proficiency (mean score on TOEIC: 920.2, SD: 33.3, range: 850-980) participated in this experiment. ERPs were measured at the critical phrase (a cleft pivot: 'the banker' or 'the queen') and all the following expressions (i.e. words/phrases) in a sentence. We found that, first, all the expressions in cleft-pivot focus position registered a large positivity. Likewise, the final expressions in the congruent condition recorded a positivity, but those in the incongruent condition didn't. Second, the expressions in cleft-pivot focus position in the incongruent relative to the congruent condition elicited N400 ($F(1,19)=7.28, p<0.05$) at right anterior regions at the 300-400 ms interval and widespread P600 ($F(1,19)=31.35, p<0.001$) at the 600-700 ms interval, namely, a bi-phasic RAN-P600. The word immediately after the pivot (e.g., 'that') in the incongruent relative to the congruent condition elicited an ELAN ($F(1,19)=9.34, p<0.01$) at 100-200 ms interval, and the sentence-final expressions in the incongruent relative to the congruent condition evoked a sustained negativity with marginal effects at the 300-500ms interval.

We take the results in this experiment to indicate that the N400 evoked at the cleft pivot in the incongruent condition reflects a violation of IS called for by the congruence between the preceding *wh*-question and its answer in a given context, and the P600 at the same position is a signature of syntactic integration difficulty due to the misfit of a non-focused constituent in a syntactic position reserved for focused expressions. At the same time, we suggest that the sustained negativity at the sentence-final elements in the incongruent condition is a neural correlate of increased syntactic complexity owing to the IS-wise mis-alignment of syntactic constituents.

Thursday Poster.58

**THE PROCESSING OF FOCUS ALTERNATIVES:
EVIDENCE FROM NEUROIMAGING**

Katharina Spalek (Humboldt-Universität zu Berlin) & Yulia Oganian (University of California,
San Francisco)

katharina.spalek@hu-berlin.de

Linguistic focus signals the existence of alternatives that are relevant for the interpretation of an utterance. Focus can be realized through prosodic pitch, morphology or syntax. For example, a speaker claiming that “[NEUROIMAGING]_F is a very useful tool to investigate focus alternatives” and putting a pitch accent on “neuroimaging” might imply that other methods, measuring reaction times, for example, are less well suited. The speaker thereby causes the listener to contrast the focused element (neuroimaging) with possible alternatives (reaction time, EEG, eyetracking).

A focused element and its alternatives often come from the same semantic (taxonomic) categories. Even if they do not come from the same category, they can replace each other in a given context, which is how semantic relatedness is described by latent semantic analysis. Thus, one might hypothesize that the processing of focus alternatives is a process akin to semantic priming, possibly enhanced by the increased prominence of a focused element. Alternatively, one might hypothesize that the processing of focus alternatives is part of discourse processing: What can (or cannot) be a focus alternative, is determined by the utterance context. For example, “his dog” is a valid focus alternative for focused “Mary” in the sentence “John went for a walk with [MARY]_F”. However, in the sentence “John proposed to [MARY]_F”, “his dog” is not a valid alternative.

In a cross-modal priming paradigm with functional magnetic resonance imaging (fMRI), we directly pitted these hypotheses against each other. 26 native German speakers listened to German sentences that were spoken either with subject focus or object focus (see Examples 1a and 1b for an English rendition).

(1a) Carsten has picked [CHERRIES]_F from the tree.

(1b) [CARSTEN]_F has picked cherries from the tree.

These sentences differ in their alternative sets. While for sentence 1a, the alternative set is the set of elements that can be picked from trees (e.g., {peaches, apples, cherries, quinces, plums,...}), for sentence 1b, the alternative set consists of individuals who could have done the picking, for example {Marie, Carsten, Damon, Leah, ...}. Following a variable time delay after the prime sentence, participants had to respond to a written probe word, e.g. PEACHES. In case 1a, the probe is related to the sentence content and a member of the alternative set. In case 1b, the probe is still related to the sentence content, but it is not a member of the alternative set. In an unrelated case, the probe PEACHES was preceded by a different sentence, for example “Sarah has tuned the violins in the music room”. Twenty-four sentences were tested in each condition, and twenty-four additional filler sentences were included to minimize strategic processing. Participants had to press a button if the probe word had appeared in the sentence (this only happened for filler trials). We compared neural responses to the probe word on the three conditions. Both related conditions induced less activation than the unrelated condition in the bilateral superior temporal gyri, replicating previous findings on semantic priming. This effect did not differ between related alternatives and related non-alternatives. So, this finding suggests a general semantic priming for probes related to the sentence content which is not affected by the alternative status of the probe word. By contrast, for the contrast between related alternatives (e.g., PEACHES primed by sentence 1a) and related non-alternatives (e.g., PEACHES primed by sentence 1b), we observed stronger activations for alternatives than for non-alternatives in the precuneus and in the frontomedian wall. Both areas have been observed in brain imaging studies of coherence processing. This finding places the processing of focus alternatives at the level of discourse processing. Moreover it suggests the interesting possibility that focus alternatives, activated through focus, contribute to the coherence of connected text/ discourse.

Thursday Poster.59

L2 LEXICAL ENGAGEMENT OF GRAMMATICAL FUNCTIONS IN RECENTLY LEARNED WORDS: INSIGHTS FROM AN EYE-TRACKING STUDY

Verónica García-Castro (University of York-University of Costa Rica) & Prof. Leah Roberts (University of York)
vgc505@york.ac.uk

Lexical engagement refers to the interaction of lexical items with other lexical entries and sublexical representations in the mental lexicon (Leach and Samuel, 2007). While a number of studies have shown that online lexical engagement of recently learned words is possible in adult monolingual speakers (Mehravari et al, 2015; Tamminen and Gaskell, 2012), studies of L2 lexical engagement of newly learned words are very scarce (Elgort et al, 2015). We explored this issue in the current experiment, investigating the extent of L2 online lexical engagement in recognition and resolution of subject-object ambiguities in temporarily ambiguous sentences. We wanted to investigate whether L2 learners were able to detect and resolve subject-object ambiguities, and thus if the recently learned words showed lexical engagement with other lexical items and lexical levels (i.e. semantic and syntactic). 25 adult Spanish L2 learners of English and 25 English controls took part. In order to keep the lexical knowledge constant, participants were trained on a set of 14 pseudowords acting as verbs like “gwap” “hirp,” which were then inserted into English sentences. The treatment consisted of a learning/training phase where participants read 24 repetitions of these novel items set in meaningful (English) sentences. After the training task, participants immediately undertook a recognition and recall test (lexical configuration). L2 lexical engagement was tested through a garden path eye-tracking study (day 2) (modelled on Roberts and Felser, 2011) in which participants read semantic plausible and implausible sentences like 1 below containing the target word (novel item, i.e. “gwap”).

1) While the child gwapped [ate] the ice-cream/puppy dropped to the floor.

Results of linear mixed-effect modelling showed significant L1 effects on the disambiguating region of the sentence in first fixations ($\beta = 31.85$, $SE = 14.17$, $t = 2.247$, $p < 0.05$), first pass times ($\beta = 45.89$, $SE = 20.04$, $t = 2.289$, $p < 0.05$) and regressions into the region ($\beta = -0.18875$, $SE = 0.06559$, $t = -2.878$, $p < 0.05$) because L2 learners read the segments more slowly than English native speakers. Effects of first fixations in the disambiguating region have been related to syntactic processing difficulty in L1 (Frazier and Rayner, 1982; Pickering and Traxler, 1998; Clifton and Staub, 2011) and this was also evidenced in the L2. Both types of learners used semantic and lexical cues to resolve the ambiguity (Papadopoulou, 2005), but it was more effortful and slower in the L2. In terms of plausibility, there was an effect in total reading times of the ambiguous determiner phrase where plausible sentences elicited longer reading times than implausible sentences. For instance, L2 learners took longer to process the direct object (i.e. “the puppy” in 1 above) of the implausible sentences. This effect shows that learners had semantically integrated and engaged the recently learned pseudowords to an extent that they were able to identify the semantic and syntactic implausibility in the direct object. In addition, they seem to have strongly committed to the semantic processing of the initial analysis (the ambiguous DP as the direct object, and not the subject of the following clause); hence, they found it hard to abandon that first analysis in the plausible sentences. Overall, the results of this study indicate that L1 and L2 learners activated and engaged semantic and syntactic knowledge of the recently learned pseudowords to parse and resolve subject-object ambiguities.

THE ROLE OF EXPECTATIONS IN THE PRODUCTION BIAS TOWARDS SHORT DEPENDENCIES

Idoia Ros, Adam Zawiszewski and Itziar Laka (UPV/EHU)
idoia.ros@ehu.eus

Languages favour short dependencies [1]. We investigated whether the magnitude of this preference varies as a function of case morphology or of word order-based expectations. Case marking provides cues that reduce uncertainty during processing [2]. Alternatively, certain word orders increase the expectation of upcoming linguistic elements [3]. Both reduced uncertainty and high expectations translate into a lessened processing cost of long dependencies, which could reduce the need to avoid them in production [1,4]. A cued-recall experiment was conducted in Basque, Polish and Spanish (n=67), for which 24 transitive and 30 ditransitive sentences were used. Crucially, these three languages differ regarding case morphology and expectation values in the relevant syntactic category in the type of sentences used in our experiment. Regarding case marking, Basque and Polish present rich morphological systems, whereas Spanish case marking is scant. Regarding word order based-expectations, the *cloze value* of the syntactic category after a long dependency is higher in Basque than in Polish and Spanish ditransitive canonical sentences (.71 vs. .58), but lower in Basque than in Polish and Spanish transitive canonical sentences (.60 vs. .89 and .87). Three conditions were created by manipulating the length of sentence constituents: All-Short, Long-O, Long-IO/S. Participants had to read and recall the constituents in order to produce the target sentence later on. We measured the number of shifted non-canonical orders produced by our participants in the Long-O versus the All-Short condition (note that shifted orders in the Long-O but not in the All-Short condition shorten dependency length). Our results reveal consistent cross-linguistic variation in the bias towards short dependencies (Fig. 1). This variation is not derived from differences in case morphology, but is in line with word order-derived cloze values: Polish and Spanish pattern identically and radically differ from Basque. In ditransitives, Polish and Spanish showed a stronger preference for shortening dependency length than Basque (shifted orders were produced approx. 50% less often). In transitive sentences, Polish and Spanish present no preference for short dependencies, whereas Basque does (shifted orders are produced approx. 20% more often). Our results suggest that the strength of word order-based expectations can alleviate the pressure towards avoiding the production of long dependencies.

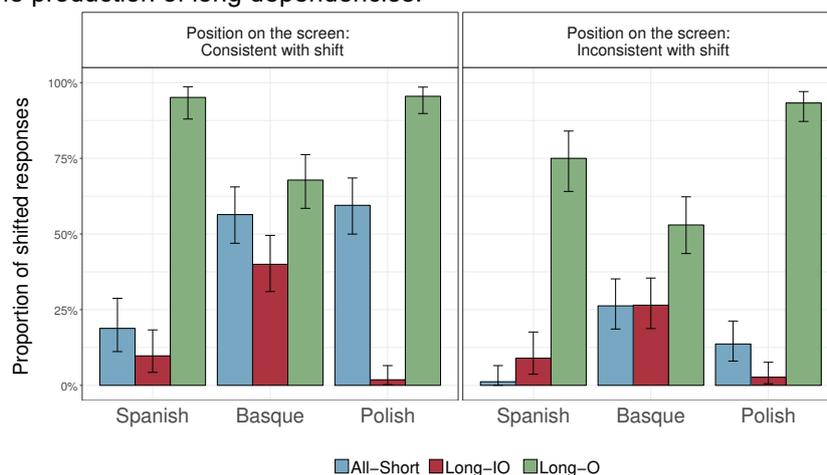


Fig. 1. **Proportion of shifted orders in ditransitive sentences.** The position of the constituents on the screen was counterbalanced (consistent vs. inconsistent with the expected direction of shift).

References: [1] Futrel, R.I et al. (2015). *PNAS*, 112(33), 10336-10341 | [2] Tily, H. J. (2010). PhD diss. | [3] Levy, R. (2008). *Cognition*, 106(3), 1126-1177 | [4] Hawkins, J. A. (1994). NY: CUP.

Thursday Poster.61

EFFECTS OF CATHODAL tDCS ON ONLINE ACQUISITION OF NOVEL WORD FORMS

Daria Gnedykh (Saint-Petersburg University), Nadezhda Mkrtychian (Saint-Petersburg University), Diana Kurmakaeva (Saint-Petersburg University), Evgenii Blagoveschenskii (Saint-Petersburg University), Svetlana Kostromina (Saint-Petersburg University) & Yury Shtyrov (Aarhus University, Saint-Petersburg University)
d.gnedykh@spbu.ru

Whereas behavioral and neurophysiological observations suggest differences in the acquisition of new abstract and concrete words (Schwanenflugel & Shoben, 1983; Hoffman, Rogers, & Lambon Ralph, 2011; Wang et al., 2010; Hoffman et al., 2015), the available evidence with respect to the mechanisms underpinning acquisition of novel surface forms and single-word semantics remains contradictory. To clarify these mechanisms, a causal approach targeting specific brain areas appears necessary. Here, we have applied cathodal tDCS over Wernicke's area to address contextual acquisition of new concrete and abstract words. Since previous studies, which used different tests to assess tDCS effects, produced mixed results (e.g., Brückner & Kammer, 2016; Jacobson et al. 2012; Floel et al 2008; Bastani & Jaberzadeh, 2011), we have employed a set of different behavioural measures to quantify the outcomes of contextual word learning both immediately after the tDCS/learning session and after a 24-hour delay.

All stimuli were tri-syllabic written word forms, grouped into sets of 10, matched statistically on their lemma and final syllable frequency. Novel items were created by modifying existing words in participants' native language (Russian) to produce a lexical competition effect (e.g. mandarin - *mandanal), and were rotated across conditions (used as new abstract words, new concrete words, unrelated fillers or competitors). Acquisition of new word forms and, simultaneously, their novel meanings was achieved by written word-by-word presentation of 5 sentences describing situations through which the participant could deduce the meaning of novel words (e.g., "in medieval times, they treated lice with mandanal"). The meanings of new words were also novel, to avoid learning synonyms of familiar lexicon items. The sentences were constructed such that their length and number of words were balanced and novel words were always presented in their base form.

The sample included two groups, 24 healthy monolingual Russian speakers each (age 17-35). Before the learning session, one group received cathodal tDCS of Wernicke's area for 15 minutes, while the other (control) group received placebo (so called sham) tDCS. Both immediately after learning and on the next day, learning outcomes were assessed using free recall, recognition and lexical decision tasks. The results were analysed in terms of reaction times and accuracy (number of correct answers) and compared between and within groups using Wilcoxon Test.

Free recall task indicated limited ability of subjects to recall the forms acquired, with no differences between sham and cathodal groups. The recognition task did not show significant differences between the two groups either. However, on Day 1, the lexical decision task showed better accuracy in the cathodal tDCS than sham group for responding to pseudoword competitors of novel abstract words ($p \leq 0.015$). On Day 2, for both sham and cathodal tDCS in lexical decision and recognition tasks, the scores were lower for novel abstract and concrete words than the respective existing competitor words (all p -values ≤ 0.006); conversely, the latency of response times for novel abstract and concrete words in comparison with these competitors was increased.

These findings suggest that cathodal tDCS over Wernicke's area, in comparison to placebo stimulation, reduces lexical competition for novel words, which can be interpreted as a sign of reduced integration of these novel items into the mental lexicon. This effect is most pronounced for abstract words, highlighting Wernicke's area role in acquiring abstract semantics. Future research should use different stimulation regimes, sites and polarities to scrutinise tDCS effects on language comprehension and learning, potentially leading to development of neurostimulation protocols for e.g. clinical and educational applications.

Supported by RF Government grant contract No.14.W03.31.0010.

NOT ALL ISLANDS ARE CREATED EQUAL: SPEEDED JUDGMENTS IN SPANISH

Claudia Pañeda (Universidad de Oviedo), Sol Lago (Universität Potsdam) & Claudia Felser (Universität Potsdam)
panedaclaudia@uniovi.es

Extracting elements from sentence regions which are thought to disallow extraction (so-called 'islands') usually leads to unacceptability (e.g. **What* do you wonder [*whether* we cooked _]?). It has often been claimed that processing factors may influence the perception of these sentences [1,2], but previous empirical work supporting their unacceptability mainly used offline judgment tasks [3,4]. We extended this work to a different language, Spanish, using a speeded acceptability judgment task. This task increases processing demands by using a more challenging presentation mode and restricting the amount of time that participants have to reflect on their acceptability intuitions. We evaluated the acceptability of four different types of Spanish islands. Our findings replicate island effects previously observed in other languages but also revealed clear differences between island types.

Method. Participants included 80 adult native speakers of Spanish (age: 23.8, 49 female). Sentences were presented word by word (SOA=400 ms) with a 2000 ms response deadline (Table 1). We tested four island types ('if', 'complex NP', 'subject' and 'adjunct') with a 2 x 2 design that manipulated the presence vs. absence of islandhood-inducing cues such as *si* 'if' (compare 1a,c vs. 1b,d) and the distance between the gap and the extracted element (short vs. long: 1a,b vs. 1c,d). In previous studies, the presence of a statistical interaction between these two factors has been used as evidence of an island effect, as it shows that there is a portion of the unacceptability of island sentences that cannot be explained by the mere addition of the two factors [3,4].

Results and discussion. Overall, island sentences (e.g. 1d) were less acceptable than predicted by the mere addition of the effects of dependency length and the presence of islandhood cues, as witnessed by a significant 2 x 2 interaction. Importantly, the size of the interaction, as well as the mean acceptability of the island conditions, varied between island types. Acceptability was lowest for subject islands (12%, SD: 33%), followed by complex NP (18%, SD: 39%), adjunct (31%, SD: 46%) and *if* islands (42%, SD: 50%). These results confirm the existence of island effects in Spanish, but they also indicate that not all islands sentences are equally unacceptable. The low ratings of subject and complex NP islands replicate previous rejection patterns found in English and Norwegian, whereas the relatively high acceptability of *if* islands resembles previous results for Norwegian *om* but not English *whether* [4]. These data challenge grammatical theories that attribute island effects to the violation of a single underlying principle, which would have predicted similar acceptability patterns across islands [5]. Rather, they support accounts that allow for gradedness [2] or differentiate between weak and strong islands [6].

Table 1. Sample item set for an *if* island.

- 1a. *-island/-long* ¿Quién __ piensa [que hemos recibido una carta]?
'Who __ thinks [that we have received a letter]?'
- 1b. *+island/-long* ¿Quién __ pregunta [si hemos recibido una carta]?
'Who __ asks [if we have received a letter]?'
- 1c. *-island/+long* ¿Qué piensas [que hemos recibido __ por correo]?
'What do you think [that we have received __ by post]?'
- 1d. *+island/+long* ¿Qué preguntas [si hemos recibido __ por correo]?
'*What do you ask [if we have received __ by post]?'

[1] Kluender & Kutas (1993) *Lang Cognitive Proc*; [2] Hofmeister & Sag (2010) *Language*; [3] Sprouse, Caponigro, Greco, & Cecchetto (2016) *Nat Lang Linguist Th*; [4] Kush, Lohndal, Sprouse (2017) *Nat Lang Linguist Th*; [5] Müller (2011) *Constraints on displacement*; [6] Szabolcsi (2006) *The Blackwell companion to syntax*.

EFFECTS OF SYNTACTIC AND SEMANTIC STRUCTURE ON PRODUCTION PLANNING

Monica L. Do (University of Southern California) and Elsi Kaiser (University of Southern California)

monicado@usc.edu

Overview: We conduct two eye-tracking studies investigating the relative contributions of **syntactic versus semantic structure** during production planning. Speech latencies and eye-movements from a ‘see-and-describe’ task (Exp1) show that initial formulation of a message is guided by the semantic structure of the utterance. But the process of encoding that message linguistically is neither purely syntactically- nor semantically-guided. Instead, it is the mapping between syntactic and semantic prominence that guides linguistic encoding. Additional data from post-experiment questionnaires, speech errors, and from Exp2, indicate that our results *cannot* be attributed to non-linguistic factors. **Exp1** (n=34): Unlike Agent-Patient/AP (1a) and Experiencer-Stimulus/ES verbs, *Stimulus-Experiencer/SE (1c) verbs exhibit a syntax-semantics mismatch*: The semantically most prominent role, the Experiencer (Grimshaw, 1990; Jackendoff, 1987), is the syntactic *object*. Participants first saw an AP, ES, or SE verb, and then saw an image with two characters depicting the action/psychological state for that verb. The task was to describe the image.

- (1a) Agent-Patient: Leslie confronted Ann.
- (1b) Experiencer-Stimulus: Leslie feared Ann.
- (1c) Stimulus-Experiencer: Leslie frightened Ann.

Because there is no correspondence between syntactic versus semantic prominence in SE verbs, we use them to tease apart their respective influences on linguistic encoding. Prior work (Griffin & Bock, 2000) using active-passive pairs showed a privileged role for syntactic prominence, but other factors (e.g. surface form and humanness) may have contributed to those results. **Predictions:** If encoding starts with the most syntactically prominent argument, speakers should fixate subjects before objects, even when the subject is less prominent semantically. But, if encoding starts with the most semantically prominent argument, speakers should fixate the semantically prominent Experiencer first, even when it is the object. **Exp1 Results** (Figs.1,2a): 200-400ms after image onset, when conceptually *forming their message*, speakers preferentially fixate the semantically prominent argument (A or E), regardless of whether it is the subject or object ($p < .01$). But, *during linguistic encoding* (400+ms after image onset), SE verbs show slower speech onsets ($p < .01$) and reveal subject/object competition in eye-movements ($p < .01$). In contrast, speech onset times and eye-movements in this time window show that speakers have no difficulty encoding ES or AP verbs. Thus, it is *not* the case that sentences with Experiencer and Stimulus arguments are categorically harder to encode than Agent-Patient arguments. Rather, difficulty in linguistic encoding is driven by the *mismatch* between syntactic and semantic prominence, as exhibited by SE verbs. We suggest that encoding is not primarily driven by syntactic or semantic prominence (cf. Griffin & Bock, 2000), but by the *mapping* of syntactic to semantic prominence.

Exp2 (n=18): A possible concern is Exp1’s results stem from non-linguistic factors (e.g. visual salience, image interpretability), not sentence planning. If so, these should be reflected in a non-linguistic Picture Inspection task. **Exp2’s results** (Fig.2b): Absent a language-production task, images for our verb types do *not* elicit different eye-movements. Also, eye-movements in Exp1 differ from Exp2 in precisely the predicted conditions and time windows of interest ($p < .05$). This – along with post-experiment questionnaires and speech error data – confirms that Exp1’s results are due to language planning processes, not extra-linguistic artefacts.

Fig 1 Mean speech latencies by verb condition. Errors bars show +/-1SE.

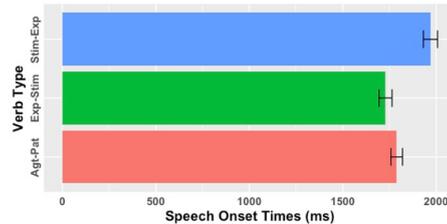
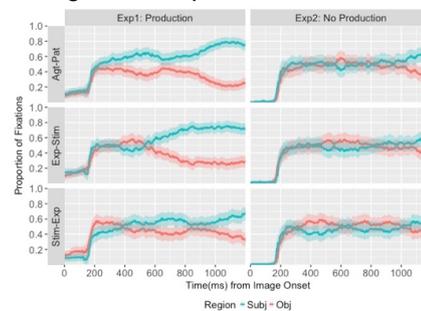


Fig 2 Eye-movements with 95% CIs in Exp1 (left, 2a) & Exp2 (right, 2b) during planning, before speech.



PARALLELISM EFFECTS IN PRONOUN DEPENDENCY RESOLUTION

Kathleen G Hall (Northwestern University), & Masaya Yoshida (Northwestern University)
KathleenHall2018@u.northwestern.edu

Introduction: The parallelism effect is pervasive in the processing of coordinate structures. Past studies have shown that the parser prefers conjuncts to be structurally parallel [1,2]. However, it is unclear what sort of structures are subject to this parallelism preference. Is it limited to parallelism in terms of phrase structure? This work examines parallelism effects in pronoun-antecedent dependencies, and demonstrates that parallelism in terms of dependency length is also inspected during online processing of coordinate structures.

Pronoun resolution: During the pronoun resolution process, once the pronoun is identified, the parser has to search through previously processed material, and evaluate and select an appropriate antecedent for the pronoun to resolve the pronoun-antecedent dependency. The selection of an appropriate antecedent has been shown to be affected by various factors, such as grammatical function [3], subjecthood [4], and coherence relations [5]. To this catalogue, we add parallelism in terms of dependency length. For example, in (1), both of the conjuncts include a pronoun and two possible antecedents for the pronoun.

(1) John said that Mary hates his mother and Jane said that Susan hates her father.

In the first conjunct, the matrix subject *John* is serving as the antecedent for the pronoun *his*. The pronoun in the second conjunct shows ambiguity: Both matrix subject *Jane* and the embedded subject *Susan* can be the antecedent. We show that, facing such an ambiguity, the parser prefers the matrix subject *Jane* as the antecedent for *her*, due to how the pronoun is resolved in the first conjunct, i.e., to maintain parallelism of the dependency length.

Experiment 1: An offline binary forced-choice fill-in-the-blank sentence completion task (n=48) was conducted. Participants read and complete sentence fragments like (2c) by choosing a pronoun (*his* or *her*). The fragment (2c) was either preceded by (2a) or (2b), or presented as is without a preceding conjunct. In (2a) the first conjunct has long dependency (*John-his*), and in (2b), it has short dependency (*Mary-her*).

(2)a./b. John said that Mary hates **his/her** mother and (2c)

(2)c. Jane said that Max loves ___ (his/her) father.

The result shows that parallel dependencies are strongly preferred: In (2a) a long-dependency is strongly preferred in the second conjunct (72%, $p < .00001$), but, in (2b), a short dependency is preferred (73%, $p < .00001$). No such bias was seen in (2c) ($p > .1$).

Experiment 2: An eye-tracking while reading experiment was conducted to further probe the effects of parallelism in online processing, in which the Length of the dependency (Long vs. Short) and Parallelism among the conjuncts (Parallel vs. Unparallel) were manipulated as independent factors in 2x2 factorial design, as shown in (3).

(3) a. Michael thought Emma missed **her** dog and Sarah thought Max despised **his** cat ...

(3) b. Emma thought Michael missed **her** dog and Max thought Sarah despised **his** cat ...

(3) c. Michael thought Emma missed **her** dog and Max thought Sarah despised **his** cat ...

(3) d. Emma thought Michael missed **her** dog and Sarah thought Max despised **his** cat ...

The results of this experiment demonstrate a clear parallelism effect, as well as an effect of locality. Model comparison of converging maximally inclusive models to reduced models revealed that *his cat* in the Parallel conditions (3a/b) was read faster than in the Unparallel conditions (3c/d) in total fixation time ($X^2=3.4$, $p=.06$). The Short conditions (3a/d) also showed facilitated processing in total fixation time ($X^2=6.6$, $p<.05$) and first fixation duration ($X^2=5.1$, $p<.05$). In the first spill-over region, the regression path durations were significantly shorter in the Parallel conditions (3a/b) than the Unparallel conditions (3c/d).

Conclusion: Taken together, these results demonstrate that both locality and parallelism can affect pronoun resolution: they appear to constrain the search space during antecedent retrieval. While locality may be more easily accommodated within a cue-based retrieval system, the observed parallelism effect may pose challenges for current models of dependency resolution.

References: [1] Frazier and Clifton (00) JPR, [2] Sturt et al. (10) JML, [3] Smyth (94) JPR, [4] Crawley et al. (90) JPR, [5] Rohde and Kehler (13) LCP.

EVALUATING PREDICTION-BY-PRODUCTION: PRODUCTION-LIKE ACCESS TO ORTHOGRAPHIC AND PHONOLOGICAL FORMS OF PREDICTABLE WORDS

Aine Ito (Humboldt University of Berlin) aine.ito@hu-berlin.de

Prediction models [1,2] propose that people use their language production system for prediction, and when they do so, they pre-activate representations of a predictable word like they do so in production. Prediction-by-production is likely to be involved to predict precise information like word form. Thus, I examined whether pre-activation of orthographic and phonological forms pattern with production models in a visual world eye-tracking study. Orthography is generally accessed as people hear a word [3], but not when people utter a word [4], unless there is a visual context that facilitates orthographic access. So, under prediction-by-production, orthography should not be pre-activated during listening without a relevant visual context. But if pre-activation occurs like activation in comprehension, orthography and phonology should interact [3] and orthography should be accessed without a relevant visual context. I manipulated the visual context using Japanese logogram (kanji) and phonogram (hiragana) to create word pairs that were orthographically related in kanji (e.g., 魚 - 角) but not in hiragana (e.g., さかな - つの).

80 participants heard sentences that contained a predictable word (cloze probability $M=87%$), viewed a scene depicting 4 words, and clicked on a mentioned word or the background if none of the words was mentioned. The scene contained a critical word (the predictable target word, a word that was orthographically/ phonologically related to the target word or an unrelated word, Figure 1) and 3 unrelated distractor words. The number of strokes, length and frequency of the words were matched across the conditions. Participants saw the identical set of words in kanji ($N=40$, Exp 1) or hiragana ($N=40$, Exp 2).

Figure 1 plots the model fit of a growth curve analysis [5]. It tested condition by visual context interaction with linear and quadratic time to capture overall and time-course differences from the scene onset + 200 ms to the target word onset + 200 ms (200ms lag: estimate for saccade initiation). Target, orthographic and phonological conditions all attracted more overall looks than the unrelated condition ($p<.05$). The orthographic effect interacted with the visual context ($p<.01$), driven by a significant effect in kanji ($p<.05$) but not in hiragana ($p=.5$).

The lack of the orthographic effect in hiragana but its presence in kanji demonstrates a critical role of visually available orthographic information on pre-activation of orthographic form. This finding fits with production models [3] but not with comprehension (auditory word recognition) models [4], providing support for prediction-by-production.

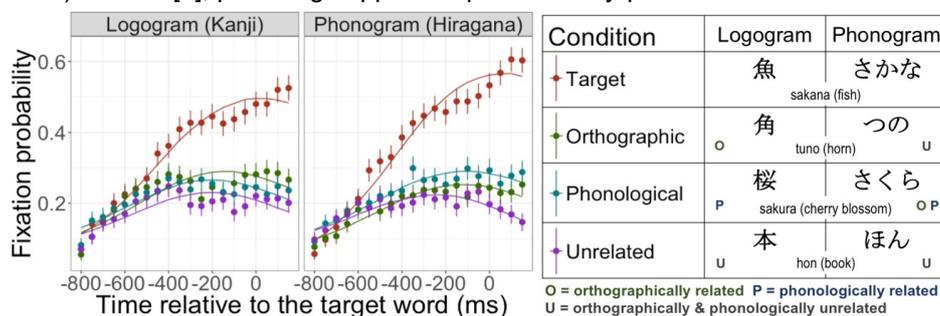


Figure 1. Model fit (left) and example critical words in 4 conditions (right).

References. [1] Pickering & Gambi (2018). *Psychol. Bull.* [2] Huettig (2015). *Brain. Res.* [3] Frost & Ziegler (2007). Oxford University Press. [4] Alario, et al. (2007). *Cognition.* [5] Mirman, et al. (2008). *J. Mem. Lang.*

CUE RELIABILITY AND ADAPTIVE RE-WEIGHTING IN SPOKEN WORD RECOGNITION

Wednesday Bushong (wbushong@ur.rochester.edu) & T. Florian Jaeger (fjaeger@ur.rochester.edu)

During spoken word recognition, listeners integrate bottom-up auditory cues from the signal with top-down cues (e.g., lexico-semantic context). Work on cue integration [e.g., 1, 2] has generally assumed that cue weights are static. In everyday language use, however, the relative reliability of cues can vary, e.g., between talkers. We explore whether listeners are sensitive to changes in the relative reliability of cues, and learn to re-weight cues accordingly. Specifically, we investigate what happens when two cues exhibit unexpectedly high degrees of conflict in recent input. We hypothesize that in such situation, the cue that is deemed less reliable (or more likely to vary across situations) will be down-weighted over time [3, 4]. We investigate this for the integration of bottom-up acoustic and top-down semantic cues in spoken word recognition.

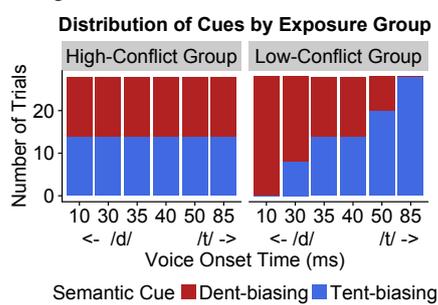
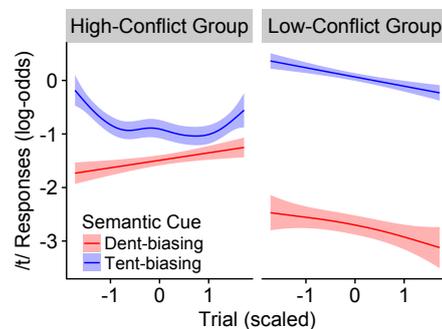


Figure 1: Between-Group design.

Methods. We present listeners (N=106) with sentences like “When the ?ent in the [fender/campground] was noticed...”, and they judge whether they heard “tent” or “dent” (following [5,6]). Two cues are varied: the acoustics of the sound range from /t/-/d/ (using VOT); and a binary semantic cue biases toward either “tent” or “dent”. We divide subjects into two exposure groups differing in the amount of cue conflict they encounter. In the *High-Conflict Group*, the acoustic and semantic cues are uncorrelated, creating frequent conflict. In the *Low-Conflict Group*, we increase the correlation between semantic and acoustic cues, thus decreasing such conflicts (Fig 1).

Results. We found strong main effects of both VOT ($\hat{\beta} = 0.1, p < 0.001$) and semantic cue ($\hat{\beta} = 0.9, p < 0.001$), confirming that both cues affect categorization (replicating [5,6]). Critically, the semantic cue effect was much larger in the Low-Conflict Group than in the High-Conflict Group (significant at all VOTs tested; $\hat{\beta}_s \geq 0.43, ps < 0.001$). This difference was driven by a three-way interaction between Group, trial, and semantic cue: semantic cues were down-weighted over time for the High-Conflict group, but not the Low-Conflict group ($\chi^2 = 12.34, p < 0.01$). Fig 2 illustrates this through fits from a generalized additive mixed model. A second study not reported here conceptually replicated this effect using a different design and stimuli.



Conclusions. Listeners re-weight cues depending on their reliability over time: specifically, listeners who encounter high levels of conflict between acoustic and semantic cues over time down-weight the semantic cues, relying instead primarily on acoustic cues. This effect emerges over time, suggesting that listeners cumulatively track the correlations of cues in their exposure to guide cue re-weighting. These findings highlight the adaptivity of spoken word recognition, and point to implicit learning processes that continuously update through life to support processing.

[1] Oden & Massaro (1978) *Psych Review*. [2] Toscano & McMurray (2010) *Cognitive Science*. [3] Jacobs (2002) *TICS*. [4] Atkins et al. (2001) *Vision Research*. [5] Connine et al. (1991) *JML*. [6] OMITTED FOR REVIEW

Thursday Poster.67

KOREAN NEGATIVE POLARITY ITEMS MEET AN ERP STUDY

Myung-Kwan Park (Dongguk University), Wonil Chung (Dongguk University) & Sanghoun Song (Incheon National University)
parkmk@dongguk.edu

Previous neuro-imaging studies have examined the neural processes of licensing negative polarity items (NPIs) in English such as 'any' and 'ever' in the anomalous sentences without NPI-licensing elements like negation, which elicited an N400 followed by a P600 compared to their grammatical counterparts (cf. Drenhaus et al., 2004, 2005, 2006). They suggested that the failure in licensing of NPIs engenders semantic integration costs (N400), but the additional P600 component from unlicensed NPIs reflects different aspects of processing them. Xiang, Grove and Giannakidou (2016) also noted that explicit and implicit negative meanings were integrated into the grammatical representation in distinct ways, leading to a difference in the P600, and calling for a separation of semantic and pragmatic integration during NPI licensing.

Turning to Korean, it is controversial whether the two NPIs in Korean such as the nominal 'amwu N-to' (any N) and the adverbial 'te isang' (any more) are licensed by nonveridical contexts like interrogatives and '-ki cen-ey' (before)-clause in Korean, although it is well established that they are licensed by an overtly negated predicate (Lee, 1999; Hwang, 2013). Thus, in order to examine how Korean NPIs enter into licensing relation during online processing, this study conducted two ERP experiments in addition to offline and online acceptability tasks with 'amwu-N-to' (Experiment 1) and with 'te isang' (Experiment 2) within four different contexts: (a) NEGative (EXPLICIT); (b) INTerrogative (IMPLICIT); (c) Korean 'BEFORE'-clause (IMP); (d) POSitive (unlicensed).

Twenty-one right-handed normal functioning Korean native speakers (14 males, mean age 23), participated in the two experiments. In offline acceptability, there was a significant effect of Licensor, $F(3,60)=142.75$, $p<0.001$, in EXP 1, and a significant effect of Licensor, $F(3,60)=79.36$, $p<0.001$, in EXP 2. ERPs were measured at the NEG/INT/'BEF'-clause verbal complex that the two NPIs, 'amwu N-to' and 'te isang,' associate with.

In Experiment 1 with the nominal 'amwu N-to', there was a main effect of Licensor ($F(3,60)=7.28$, $p<0.001$) at the 300-450 ms interval. In comparison between the unlicensed condition and each of the three licensed conditions, the unlicensed condition versus the EXP NEG condition elicited N400 ($F(1,20)=18.7$, $p<0.001$), but the unlicensed condition versus the two IMP INT/'BEF' conditions evoked no ERP effect. The two IMP conditions, INT and 'BEF', relative to the EXP NEG condition elicited N400 ($F(1,20)=9.34$, $p=0.006$ for INT; $F(1,20)=13.90$, $p=0.001$ for 'BEF').

Likewise, in Experiment 2 with the adverbial 'te isang', there was a main effect of Licensor ($F(3,60)=2.81$, $p=0.056$) at the 500-650 ms interval. In comparison between the unlicensed condition and each of the three licensed conditions, the unlicensed condition vs. the EXP NEG condition elicited anterior P600 ($F(1,20)=5.10$, $p=0.035$ at midline and right regions) besides N400 ($F(1,20)=4.20$, $p=0.054$). The unlicensed condition relative to the Korean 'BEF' condition showed N400 ($F(1,20)=7.51$, $p=0.013$) along with a sustained negativity at the 150-550 ms interval, but no ERP effect arose in the unlicensed vs. the INT condition. In comparison of the two IMP conditions relative to the EXP NEG condition, the INT condition elicited a marginal anterior P600 at left regions ($F(1,20)=6.94$, $p=0.016$), and the 'BEF' condition showed anterior P200 ($F(1,20)=4.70$, $p=0.042$) at right regions and anterior P600 ($F(1,20)=12.79$, $p=0.002$) at left regions.

The results indicate that, first, N400 was evoked by both NPIs at issue in illegal environments like a positive clause. Since N400 is regarded as a neural index of incomplete semantic integration, it follows that the Korean NPIs' licensee-licensor relation is resolved via semantic processes. Second, 'amwu+N+to'-containing non-negative conditions elicited N400, whereas 'te isang'-containing question and '-ki cen-ey' clauses elicited anterior P600. We take the latter anterior P600 component to reflect not a violation of NPI licensing but a cognitive load of discourse/pragmatic processing due to the lexical meaning of 'te isang'. Third, directly comparing the two NPIs in terms of neural profiles, we find that during processing 'amwu+N+to' is cognitively more demanding than 'te isang'.

(DIS-)CONFIRMATION OF LINGUISTIC PREDICTION BY NON-LINGUISTIC CUES

Torsten Kai Jachmann, Heiner Drenhaus, Maria Staudte, Matthew W. Crocker (Saarland University)

jachmann@coli.uni-saarland.de

We present findings from an ERP study (24 German right-handed participants, 19–28 yrs) investigating the influence of speaker gaze on listeners' retrieval and integration of expected (True) and unexpected (False) nouns in a shared visual scene. In face-to-face interactions, speakers gaze at objects about 800ms prior to their mentioning [1, 2, 3]. Previous eye-tracking studies provided evidence that participants use such a gaze cue to anticipate upcoming object references, expressed by a higher inspection rate of the gazed at object compared to a competitor [4, 5]. Here, we investigated whether the gaze cue toward a later mentioned object does indeed prepare, or even advance, referential processing. We utilized a stylized face performing gaze cues time-aligned to an auditory sentence.

Each experimental item consisted of a visual scene containing three objects that either differed in size (small, medium, large) or brightness (bright, medium, dark). After three seconds, a stylized face was displayed in the centre of these objects. Gaze cues were time-aligned to an auditory sentence describing a comparison of two of the objects of the form "Das Haus ist größer als das dargestellte Auto, denke ich" ("The house is bigger than the displayed car, I think"). The gaze cue preceding the mentioning of the second noun was manipulated in a way that it either was directed toward the consequently named object (Gaze) or redirected toward the listener (No-Gaze). As a second manipulation, the noun rendered the sentence either as True or False given the visual context (the resulting four conditions were fully counterbalanced). Each participant was presented with 38 items per condition.

We analysed ERPs time-locked to the onset of the gaze cue for the four conditions (Gaze–True, Gaze–False, No-Gaze–True and No-Gaze–False). Overall, a gaze cue toward an object led to a long lasting positive deflection starting at 150ms. Additionally, a gaze cue toward the unexpected object (Gaze–False) showed a negative deflection in the time-window between 200 and 450ms compared to its expected counterpart (Gaze–True).

The ERPs time-locked to the onset of the noun following the gaze cue region showed no effects if a gaze cue toward an object was present (Gaze–True, Gaze–False). In the absence of a preceding gaze cue, however, we observed a larger negativity in the N2 time-window (150–300ms) as well as in the N4 time-window (300–450ms) for the False condition compared to the True condition. Importantly, a late positivity is only observed when no gaze cue precedes the mentioning of the referenced object.

The effect in the N2 time-window on the noun can be interpreted as a PMN. This early effect is explained as a mismatch between the expected word form given a context and the actual activated word candidates given the speech signal listeners perceived [6, 7].

We argue that the N4 effects in both regions (gaze cue and noun) reflect expectations about the upcoming referent (the one that makes the utterance true), that can be formed once the comparator (*größer/bigger*) is heard. That is, an N4 effect is only elicited when the comparator is followed either by an unexpected gaze cue towards the false referent (in the Gaze–False), or by the unexpected noun, when no gaze cue is present (No-Gaze–False).

The positive deflection observed in both the gaze cue and noun regions can be interpreted as an integration process of the provided information about the discourse-relevant object. In the absence of a preceding gaze cue, the integration of the noun into the mental representation of the situation can only be done on the noun itself. A preceding gaze cue however allows the integration at this (earlier) point in time.

In sum, we provide evidence that speaker gaze towards a referent is processed much like verbal mentioning of the referent itself: Unexpected gaze cues and nouns result in similar ERP effects, while the presence of a gaze cue eliminates effects on the subsequent noun.

References: [1] Meyer et al., 1998; [2] Griffin and Bock, 2000; [3] Kreysa, 2009; [4] Staudte and Crocker, 2011; [5] Staudte et al., 2014; [6] Connolly and Phillips, 1994; [7] Hagoort and Brown, 2000

ARE 'GREAT' AND 'NOT GREAT' THE SAME? ERP EVIDENCE ON THE PROCESSING OF IRONY AND SENTENCE NEGATION

Stefanie Regel & Thomas C. Gunter (MPI for Human Cognitive and Brain Sciences)
regel@cbs.mpg.de

Neuropragmatic models of understanding communicative intents have proposed different mechanisms underlying the processing of complex language phenomena, such as irony (i.e., a form of indirect negation (Giora, 1995), cf. (Gibbs, 2002; Grice, 1975)), or sentence negation (Kaup & Zwaan, 2003). Though both these phenomena communicate unfulfilled expectancies, the way these intents are conveyed differ: While irony implicitly conveys the speaker's intent – mostly an opposite meaning than stated literally – (e.g., *That's great* in reply to a rather unpleasant event), sentence negations explicitly communicate the negated state of affairs by means of a negation marker (e.g., *That's not great*). Whether the processing of both phenomena involve similar, or rather different neurocognitive mechanisms contributes to those neuropragmatic models. The present study addresses this issue by directly comparing the processing of ironic and negated sentences relative to literal ones by using event-related brain potentials (ERPs). Participants read sentences that achieved either an ironic, negated, or literal meaning depending on the preceding context, and differed merely in the negation marker (for an example see above). As experimental task, a recognition test of eight items was applied. ERPs analyzed for the critical word (e.g., *great*) revealed different ERP patterns in response to irony and sentence negation (see Figure 1). For irony an enhanced P200-P600 pattern in comparison to literal sentences is replicated (Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013). In contrast, for sentence negation an enhanced N400-P600 pattern was observed. These findings of different ERP patterns seen for both language phenomena imply an involvement of different neurocognitive mechanisms initially, but partially similar ones during later stages of processing (i.e., pragmatic reanalysis) as indicated by P600. Both irony and sentence negation may engage inferential processes of the sentence meaning whereby the appropriate communicative intents are derived.

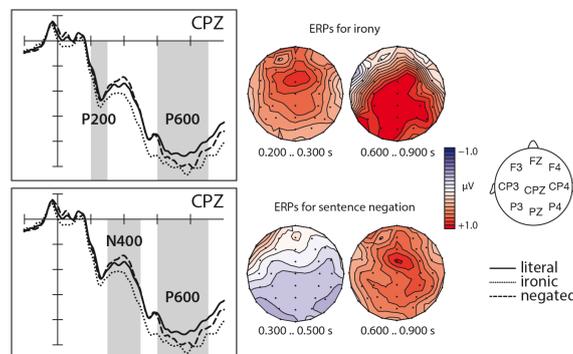


Figure 1. Grand average ERPs analyzed at the critical word that achieved either a literal (solid line), ironic (dotted line), or negated (dashed line) sentence meaning. The zoom of the CPZ electrode illustrates the ERPs in response to irony (upper panel) and sentence negation (lower panel).

References

Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, 34(4), 457-486.

Giora, R. (1995). On irony and negation. *Discourse Processes*, 19(2), 239-264.

Grice, P. H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts: Syntax and semantics* (pp. 41-58). New York: Academic Press.

Kaup, B., & Zwaan, R. A. (2003). Effects of Negation and Situational Presence on the Accessibility of Text Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 439-446.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, 23(2), 277-293.

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration Operations during Irony Processing. *Plos One*, 8, 1-10.

Thursday Poster.70

SYNTACTIC AND SEMANTIC PROCESSING IN POOR COMPREHENDERS: EVIDENCE FROM EYE-TRACKING AND COMPUTATIONAL MODELING

Luca Campanelli (Haskins Laboratories & The Graduate Center, CUNY),
Nicole Landi (Haskins Laboratories), & Julie A. Van Dyke (Haskins Laboratories)
campanelli.l@gmail.com

Individuals identified as Poor comprehenders (PCs) have significant difficulty in the domain of reading comprehension despite typical intellectual ability and word reading skill. Research studies in the US and UK suggest that PCs comprise about 10% of school age children and adults [1]. Although progress has been made, the majority of previous work has utilized offline (standardized) assessments of comprehension relevant skills, leaving the source of comprehension deficits in PC poorly understood [1]. The present eye-tracking study addressed those limitations by employing two methods for directly querying the syntactic representation created during incremental comprehension. First, similarly encoded syntactic elements have been shown to create retrieval interference in long-distance dependency formation [2]. Thus, the accuracy of the encoding of syntactically similar elements is negatively correlated with performance – contrary to the typical case in which comprehension ability is positively correlated with processing performance. Secondly, we employed syntactic surprisal [3] as a linking function to assess the nature of readers' linguistic representations as they are computed. We derived word-by-word predictions from linear (n-gram) and context-free (CFG) grammar models and conducted competitive model fits to eye-tracking data ([4]).

Method. We recruited a sample of 51 native English speakers (ages 13-19; 28 female) affording a broad range of reading comprehension skill, indexed via standardized assessment (KTEA: 70-146, M=101, SD=19). In order to isolate poor comprehenders, only participants with no history of learning or cognitive impairments and word decoding scores above 90 were included. The experimental material consisted of 40 item-sets with four conditions that parametrically decreased cue-diagnostics as illustrated in (1). In the control condition 1a, subject and verb (e.g., *father* and *smiled*) were adjacent, therefore requiring no memory retrieval. In 1b-d there were five words intervening between the dependent elements: 1b included no distractors, the embedded clause in 1c contained a syntactically matching distractor (*shirt* is grammatical subject), and 1d included a distractor that was both syntactically and semantically plausible (*clown*).

- | | |
|-------------------|-------------------------------------------------------------------------------------------------------|
| (1) a. NoLoad | The shirt was colorful and the father smiled proudly during the entire game. |
| b. Load-NoInt | The father with the very colorful <u>shirt</u> smiled proudly during the entire game. |
| c. Load-SynInt | The father who the colorful <u>shirt</u> pleased smiled proudly during the entire game. |
| d. Load-SynSemInt | The father who the colorful <u>clown</u> pleased smiled proudly during the entire game. |

Results and Discussion. For the effects of reading and condition, similar patterns emerged for ACC and RT to comprehension questions and for eye movements at the critical verb (e.g., *smiled*) and the spillover region [First pass (FP), Probability of Regression (PR), Regression Path (RP), Total time (TT)]. Overall, performance decreased as cue-diagnostics was reduced (e.g., longer fixations in 1d than 1c, 1c than 1b, and 1b than 1a), and skilled comprehenders performed better than PC. Most interesting, for ACC, RT, RP and TT at both critical and spillover regions, we found significant interactions of reading with 1c, indicating that the negative effect of syntactic interference was greater in good than poor readers. For PR and RP at the spillover region, we also found an interaction between reading and 1d, such that the semantic interference effect was greater in good than poor comprehenders.

For the surprisal analyses [4], 2-gram, 3-gram, and CFG language models were used to model eye movement data for whole sentences. For all dependent measures, effects of 2- and 3-gram were comparable in all readers. Most interestingly, for FP, RP, and TT, but not PR, we found significant interactions between CFG and reading, indicating that the more sophisticated grammar models eye-movements of good readers, but not poor readers.

Overall, it appears that PCs use the same memory retrieval mechanisms as skilled comprehenders, as the effect of semantic interference was comparable in all readers both offline (ACC and RT) and online at the critical region. The interactions between comprehension ability and 1c and the computational modeling results point to impaired syntactic processing as a key distinguishing component for comprehension skill.

References. [1] Landi & Ryherd, 2017; [2] Van Dyke, 2007; [3] Hale, 2014; [4] Brennan et al., 2016.

Thursday Poster.71

ON DURATION AND COMPLEXITY: THE HORSE RACED FASTER WHEN EMBEDDED

Nino Grillo (University of York), Miriam Aguilar (CLUNL), Leah Roberts (University of York),
 Andrea Santi (University College London) & Giuseppina Turco (Paris 7/Paris Diderot)
 nino.grillo@york.ac.uk

In current psychological models, and our everyday intuition, a simple correlation exists between relative task complexity and completion duration (when successful). Since Donders' experiments in 1867, (reaction/response) time measures have been correlated with complexity and have consistently provided key insights into processes and mechanisms of the mind. We argue that, while generally sound, in the domain of language, and in particular when prosodic effects on duration are taken into account, this simple correlation can lead to dangerous oversimplifications. (Explicit and Implicit) prosodic properties play a central role in sentence processing [1]. Prosody modulates durational properties of words and phrases to reflect their structural and interpretive properties. We show that these effects can lead to apparently paradoxical cases of shorter durations for more complex structures.

Prosody, it is assumed, does not always disambiguate syntax. The contrast between Main Verb (MV) and reduced-Relative Clause (r-RC), is one classic case of such mapping failure:

A [_{DP}The [_{NP}horse [_{CP}raced past the barn]]] fell. **B** [_{DP}The horse] [_{VP}raced past the barn and fell].

Despite their centrality in shaping theories of sentence processing, no experimental work to date has investigated the prosody of these sentences. We present evidence from production and comprehension that, contrary to previous assumptions [2,3], this contrast *is* prosodically disambiguated but that this disambiguation is best observed when the relevant clauses are embedded within a matrix clause which provides a baseline pace. Prosodic disambiguation obtains through pace modulation, with *faster* pace associated with the embedded/reduced relative reading and *regular* pace (no change) with main verb analysis. The essential contribution of the matrix sentence is to provide a baseline pace without which it is impossible to establish whether a change took place. *Importantly, duration is solely determined by prosody and independent from complexity: faster pace is associated with the more complex structure.* **Experiment 1: Planned Production.** (Higher) Attachment site has been previously shown to correlate with (separate) phrasing [4,5,6,7]. This is often observable in terms of durational differences between the two readings, with shorter durations for more deeply embedded strings and longer durations when the same string attaches to a higher position. We compared the prosodic properties of r-RCs (**A**), where the VP is embedded within the DP it modifies, and MVs (**B**), where the VP is in a sisterhood relation with the same DP. **Methods:** Five native English speakers produced 16 experimental utterances per condition (interspersed with 48 unrelated fillers) adapted from previous experiments in the relevant literature [8,9]. Each sentence was embedded within short introductory sentences containing declarative verbs (2,3). Intro strings were neutral with respect to the disambiguation and present solely to provide a baseline tempo. **Predictions:** Prosody predicts *shorter* duration for the r-RC than the MV parse, while the well-known higher complexity of r-RCs leads to the opposite prediction.

2. **Reduced-RC:** Jason claims that *the student pushed into the row of traffic* got badly hurt.

3. **Main Verb:** Jason claims that *the student pushed into the row of traffic* and got badly hurt.

Results. English speakers make use of temporal cues to disambiguate between MV and r-RC readings: the ROI (*the student pushed into the row of traffic*) was significantly *shorter* in the r-RC than in the MV condition ($t=-2.729$, $p=0.0155^*$). This disambiguation is observable already at the subject DP (*the student*), similarly shorter in r-RCs than MV ($t=-2.425$, $p=0.0167^*$). **Experiment 2:** 120 English speakers participated in a **forced-choice cloze task** with auditory stimuli produced by one of the participants in Experiment 1, and thus unaware of the goals of the study. Sentences were cut so as to remove the disambiguation regions (i.e. *(and) got badly hurt*), which were presented in text format as forced choice. The crucial manipulation involved presence or absence of intro providing reference tempo (i.e. *Jason claims that* in 2,3). While garden path effects were still present, comprehension was significantly better when r-RCs were preceded by the matrix sentence ($z= 5.271$ $p<0.0001$), while the opposite effect obtained in the MV condition ($z=-2.045$, $p=0.049$).

Ref: [1] Frazier & Gibson 2015 [2] Fodor 2002 [3] Wagner & Watson 2010 [4] Hirschberg & Avesani 1997 [5] Poschmann & Wagner 2015 [6] Wagner 2010 [7] Grillo & Turco 2016 [8] Crain & Steedman 1985 [9] Ni et al. 1996

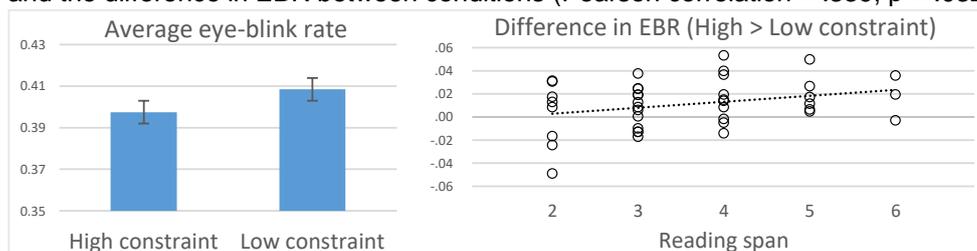
Event-based Eye Blink Rate as an Index of Working Memory Gating and Updating: Predictive Pre-updating and Individual Differences in Working Memory Capacity

Tal Ness (Tel Aviv University) and Aya Meltzer-Asscher (Tel Aviv University)
talness@mail.tau.ac.il

Extensive research indicates that spontaneous eye blink rate (EBR) is an effective indirect measure of dopamine (DA) activity in the striatum (Jongkees & Colzato, 2016). Additional work suggests that phasic DA signals drive gating and updating of working memory (WM), in line with the prefrontal cortex basal ganglia WM model (Hazy, Frank, & O'Reilly, 2006; D'Ardenne et al., 2012). In a recent event-based eye-blink rate (ebEBR) study, blink rate was shown to increase when WM was updated in a non-linguistic, reference-back task (Rac-Lubashevsky & Kessler, 2017). To our knowledge, this technic was never utilized in a language study. Recent research (Ness & Meltzer-Asscher, 2017) found evidence from event-related potentials for predictive pre-updating, namely updating of strongly predicted content into WM representation prior to confirmation of the prediction by the input. In the current study, we tested whether predictive pre-updating is also reflected in ebEBR. We used word pairs in which the first word was either highly predictive of the second (e.g. *global warming*) or not (e.g. *vegetable soup*). In highly predictive pairs, upon updating the first word into WM, if pre-updating occurs then the predicted second word is also updated, and no gating or updating is expected upon presentation of the second word. Therefore, reduced EBR is expected in the high constraint condition. Since the tendency to engage in pre-updating is affected by individual differences in WM capacity, this effect should also depend on WM capacity.

Methods: Participants were 40 native Hebrew speakers (12 male, age: M = 25, range = 20-33). 160 word pairs in Hebrew, 80 from each condition, were used. Based on a cloze probability questionnaire in which 30 (different) participants were given the first word in each pair and asked to provide a completion, the average constraint was 88.6 in the high constraint condition, and 24.3 in the low. 40 anomalous word pairs were used as fillers. Pairs were presented word by word (SOA=1000ms), and participants were instructed to press a button as quickly as possible if a word pair is anomalous. Participants' WM capacity was assessed via a reading span (RS) task. Following the methodology of Rac-Lubashevsky and Kessler (2017), blinks were measured using EOG electrodes above and below the eye and ebEBR was calculated as the average amount of blinks per second in the 4 seconds following the presentation of the second word in the pair.

Results: Results showed a significant decrease in EBR in the high (relative to low) constraint condition ($t(39) = -3.276, p = .002$). Additionally, a significant correlation was found between RS and the difference in EBR between conditions (Pearson correlation = .339, $p = .032$).



Discussion: Decreased EBR was observed in the high (relative to low) constraint condition, reflecting pre-updating. This effect was correlated with RS, strengthening the link to pre-updating. By providing corroborating evidence for pre-updating using ebEBR, we hope to establish this technique as a valuable tool to study other linguistic phenomena.

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, 109(49); Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. *Neuroscience*, 139(1); Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews*; Ness, T., & Meltzer-Asscher, A. (2017, March). Predictive pre-updating: ERP correlates and the influence of working memory capacity. *Poster presented at the 31th Conference on Human Sentence Processing (CUNY), Davis, California*; Rac-Lubashevsky, R., Slagter, H. A., & Kessler, Y. (2017). Tracking real-time changes in working memory updating and gating with the event-based eye-blink rate. *Scientific reports*, 7(1).

Thursday Poster.73

EARLY SENSITIVITY TO NUMBER AGREEMENT: WHAT PUPILLOMETRY REVEALS ABOUT L1 ACQUISITION OF GERMAN

Background. When acquiring a first language (L1), children must be able to detect the relevant grammatical distinctions and relations in their ambient language to acquire full syntactic competence. We refer to this emerging ability of responding to syntactic relations in different sentence domains as sensitivity. The aim of this pupillometry study is to investigate German-speaking toddlers' sensitivity to subject-verb agreement and its violations in sentences such as:

“The child opens the present.” versus “The children opens the present.”

Sensitivity to subject-verb agreement has been shown to be present before the age of three years in previous studies in French (Nazzi et al., 2011), Dutch (Polisenska, 2010) and English (Sundara et al., 2011). In accordance with pupillometry studies on other aspects of language (e.g., Tamási et al., 2017) we expect higher pupil dilation when listening to sentences containing a violation of agreement compared to sentences with correct agreement.

Experiment. Pupil size measures were collected with an eye tracker in a single-picture looking paradigm from German-speaking 30- and 36-month-old children (N=54) and an adult control group (N=17). Participants were presented with pictures displaying a transitive action on a computer screen and pre-recorded sentences. Sentences were either grammatical (congruent subject-verb agreement) or ungrammatical (incongruent because of violation of the number feature). Pupil size change was calculated from both eyes for 2500 ms from the verb onset in the sentence and evaluated with a linear mixed effects analysis.

Results. The adult group showed significantly larger pupil dilations in sentences with agreement violation compared to sentences without such a violation, establishing pupillometry as a suitable tool to measure sensitivity to agreement. In children, we found a three-way interaction between age of the children, subject number and grammaticality. This means that sensitivity to agreement is dependent on subject number in the children, with a developmental change between the two age groups. At 36 months children showed a pattern similar to the adult group, with larger pupil dilations for incongruent agreement, but only in the singular condition. At 30 months, no differences could be found.

Conclusion. The pupillary response in adults can be interpreted as an increased processing effort for these sentences. Children at 30 months do not seem to be sensitive to number agreement violations. At 36 months a sensitivity can be observed if the sentence subject is singular. These results suggest that German-speaking children have acquired knowledge of subject-verb agreement by the age of 3 years, but that their detection of violations is still restricted to the singular. This is in line with findings for other languages (Polisenska, 2010, Nazzi et al., 2011, Sundara et al., 2011).

References:

- Brandt-Kobe, O. C., & Höhle, B. (2014). The detection of subject-verb agreement violations by German-speaking children: an eye-tracking study. *Lingua*, 144, 7–20.
- Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, 120(1), 119–135.
- Polišenská, D. (2010). *Dutch children's acquisition of verbal and adjectival inflection*. Utrecht: LOT.
- Sundara, M., Demuth, K., & Kuhl, P. K. (2011). Sentence-position effects on children's perception and production of English third person singular-s. *Journal of Speech, Language, and Hearing Research*, 54(1), 55–71.
- Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology*, 153, 140–148.

Thursday Poster.74

WHAT MAKES A HOUSE A HOME? MECHANISMS OF LEXICAL ALIGNMENT IN PRESCHOOLERS' REFERENTIAL COMMUNICATION

Laura Lindsay (University of Edinburgh), Zoe Hopkins (University of Edinburgh) & Holly Branigan (University of Edinburgh)
L.M.Lindsay@sms.ed.ac.uk

When conversing, people must choose between alternative names to refer to an object (e.g. *cup* vs *mug*). An important determinant of referential choice is lexical alignment: the tendency of speakers to converge on the same name for a particular object (Brennan & Clark, 1996). Lexical alignment (or entrainment) has been found in school-aged children as well as adults (Branigan, Tosi, & Gillespie-Smith, 2016;). More recent research has shown that in a highly structured picture-naming 'snap' task, even 3-4 year old children display lexical alignment (Lindsay, Hopkins, & Branigan, in prep). However, the nature of the 'snap' task promoted the possible influence of lexical priming mechanisms on children's referential choices, and made it less likely that perspective-taking and social-affiliative mechanisms – which mediate alignment in adults (Branigan, Pickering, Pearson, & McLean, 2010) – would play any role. In the present study, we examined whether 3-4 year olds show lexical alignment in a less structured context involving a more demanding task. We also ask whether social-affective mechanisms influence alignment when affiliation goals are salient.

Participants (N=70) played a novel referential communication game – the "Moving House" game – with an experimenter. In the game, the participant and experimenter each moved items from a moving truck into different rooms of a house. Experimental items had two alternative names. Pre-tests established that children knew and understood both names, but had a strong preference for one alternative. The experimenter told the participant where to place items from the truck into their house (prime rounds); the participant and experimenter then switched roles, and the participant directed the experimenter where to place her items (different exemplars of the same category, e.g., a different cup; target rounds). During prime rounds, we manipulated the name that the experimenter used to name target items (preferred vs. dispreferred), and examined children's likelihood of producing the dispreferred name for a different exemplar of the same category in target rounds. Before the game, we also manipulated children's affiliative motivation, by showing half the participants a video depicting third-party ostracism, and the other half a control video (Over & Carpenter, 2009). Children were more likely to use a dispreferred name (*mug*) if the experimenter had previously used a dispreferred name than a preferred alternative (*cup*; .33 vs. .12, $p < .001$). Their tendency to lexically align was not affected by having watched the ostracism vs. control video (.35 vs .31; $p = .46$).

Our findings show that 3-4 year old children spontaneously lexically align with a conversational partner, even when this means using a normally dispreferred name. Importantly, such alignment is not restricted to simple and highly structured tasks where referential communication was not necessary for task success: in the Moving House task, children had to determine the correct location for each object, and communicate both the identity of the relevant object and its location to the experimenter. Moreover, their alignment occurred over a substantial number of intervening turns and associated time delay. However, such alignment was not influenced by exposure to third party ostracism.

Together these results suggest that preschoolers' strong tendency to spontaneously lexically align is not restricted to highly structured and cognitively undemanding tasks, and that it is not primarily driven by social-affective mechanisms whereby children imitate in order to promote affiliation, though we cannot rule out that such mechanisms may be operative under some conditions. The fact that alignment occurred for different exemplars of a category also suggests that it does not arise from a simple memory association between a referent and an episode of language use. Our results are compatible with lexical priming mechanisms, but the durability of alignment effects – in the presence of greater task demands – also suggests the possibility of perspective-taking. Children may have used shared linguistic context (i.e., the experimenter's previous name use to infer what she was likely to understand), and aligned to enhance communicative success.

Thursday Poster.75

ONLINE PROCESSING OF CASE IN AUDITORY AND WRITTEN SENTENCES AS REVEALED BY EYE MOVEMENTS IN NATIVE SPEAKERS AND L2 LEARNERS

Cheryl Frenck-Mestre (Aix-Marseille Univ, CNRS, LPL), Seung Kyung Kim (Aix-Marseille Univ, LPL), Hyeree Choo (Seoul National Univ), Alain Ghio (Aix-Marseille Univ, CNRS, LPL) & Sungryong Koh (Seoul National Univ)
Cheryl.frenck-mestre@univ-amu.fr

Eye movement data from studies conducted in Korean, Japanese and German show that native speakers exploit case marking to compute sentence structure (Koh, 1997; Mitsugi & MacWhinney, 2010) and to anticipate upcoming arguments (Frenck-Mestre et al., submitted; Henry et al., 2017; Hopp, 2015; Kamide et al., 2003a, 2003b; Mitsugi & MacWhinney, 2016). The capacity for adult learners to exploit nominal case marking to predict upcoming elements is less clear (Frenck-Mestre et al., submitted; Henry et al., 2017; Hopp, 2015; Mitsugi & MacWhinney, 2016). We examined the online use of Case in 2 eye movement experiments, using auditory (Exp 1) and written materials (Exp 2). Declarative single clause sentences were presented, with either accusative (NOM-ACC) or dative (NOM-DAT) nominal case marking. Both canonical (SOV) and scrambled (OSV) word order were used. Whether or not scrambling incurs a processing cost for native speakers is debated (Mitsugi & MacWhinney, 2016; Miyamoto & Takahashi, 2004). Native Korean speakers and native French speakers learning Korean who had never resided in Korea participated. In Exp. 1, participants viewed 2 scenes for 1 second prior to hearing an auditory sequence and choosing which of the two scenes correctly depicted the sentence. Exp. 2 presented the same sentences but in written format followed by a short yes/no question. General linear mixed effects models were performed on the eye movement and accuracy data for Exp. 1. Results showed that native Koreans (N=16) used case marking immediately, and looked significantly more at the correct image starting from the second noun independent of Case or Order. Koreans showed ceiling level accuracy for all conditions. For the L2 learners (N=27), no anticipatory looks to the correct image prior to the final auditory verb were observed. At the verb, a significant interaction between Case and Order obtained; learners looked more at the correct image for dative utterances independent of word order (63% and 59% for SOV and OSV order), but showed a decrease in looks to the correct image for scrambled compared to canonical accusative utterances (44% vs. 63%). The same interaction obtained for accuracy, with a dip in performance for scrambled compared to canonical accusative utterances (29% vs. 86% correct) but not for dative (64% vs. 77%). Results for written presentation (Exp. 2) revealed ceiling level accuracy for native Koreans (N = 16). For Learners (N = 16), a significant interaction obtained between Case and Order for accuracy, with chance level accuracy for scrambled but high accuracy for canonical accusative sentences (50% vs. 78%) and both canonical and scrambled datives (85% and 77%). Reading times showed different patterns for dative and accusative sentences, for both native Koreans and learners. For accusatives, no effects of either Case or Order were found during the first pass for either group, however; total reading times showed increased reading times at both N1 and N2 for scrambled sentences for learners. For datives, a significant interaction obtained during first pass reading times with longer reading times for N1 than N2 for scrambled but not for canonical sentences, for both Koreans and learners. Overall, our results show the anticipated use of case only for native speakers but a processing cost for scrambled datives for native Koreans as well as L2 learners in written sentences.

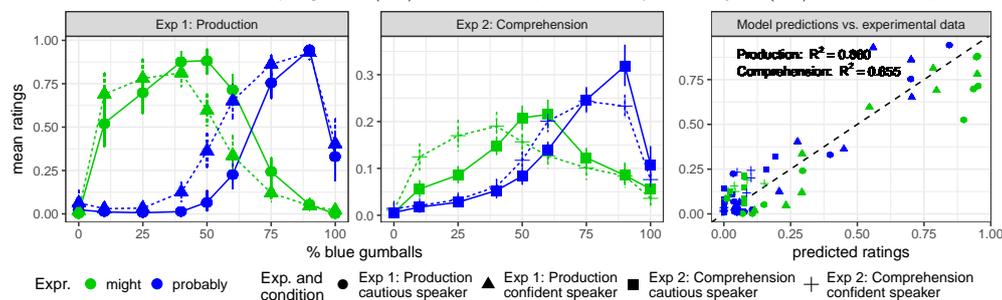
ADAPTATION TO VARIABLE USE OF EXPRESSIONS OF UNCERTAINTY

Sebastian Schuster (Stanford University) and Judith Degen (Stanford University)
 sebschu@stanford.edu, jdegen@stanford.edu

Speakers exhibit considerable production variability at all levels of linguistic representation. Listeners deal with such variability by adapting to it and updating expectations [1-4]. We extend the research on semantic/pragmatic adaptation and investigate whether/how listeners adapt to varying uses of the uncertainty expressions *might* and *probably* (“**You might / You’ll probably** get a blue gumball”). In a norming study we found that speakers vary in their mapping between these expressions and the probability of getting a blue gumball. Confident speakers use *probably* if the objective probability of getting a blue gumball is 60%, whereas cautious speakers use *might*. How listeners should adapt to such variability is still poorly understood. We propose a novel **computational model** within the Rational Speech Act (RSA) framework [5] that treats adaptation as Bayesian update of listeners’ beliefs about the speaker’s underlying semantic thresholds and show that such a model accounts for adaptation observed across both production (Exp. 1) and comprehension (Exp. 2).

Exp. 1 (production): 61 MTurk participants saw 20 exposure trials (10 critical, 10 filler) followed by 36 test trials. Exposure trials showed a video of a speaker describing a gumball machine filled with blue and orange gumballs (critical trials: 60% blue gumballs). Participants in the *confident* and *cautious* condition were exposed to the speaker producing the *probably* and *might* form, respectively. Fillers were intended to boost trust in the speaker: on 5 trials, the speaker described a typical (25% and 90%, as estimated in the norming experiment) probability with the respective other uncertainty expression. On the other fillers, the speaker said of a 100% blue machine “You’ll get a blue one”. On test trials, participants distributed 100 points across the *might* and *probably* utterances and a blanket *something else* option, for 9 blue gumball proportions. As the left panel below shows, *probably* was rated higher than *might* for a larger range of probabilities in the *confident speaker* cond. than in the *cautious speaker* cond. Like [4], we quantified this difference by fitting a spline for each expression and participant and then computing the area under the curve (AUC). We found that the average difference between the AUC of *might* and *probably* is smaller in the *cautious speaker* cond. ($t(59) = -4.98, p < 0.001$), suggesting adaptation of expectations about *might* and *probably* use.

Exp. 2 (comprehension) tested whether the change in production expectations also affects interpretation. 77 MTurk participants saw the same exposure phase as in Exp. 1, followed by 6 test trials. On test trials they heard the exposure speaker produce either the *might* or the *probably* utterance and rated for 9 different gumball distributions, how likely it was that the speaker saw that gumball machine when producing the utterance. We normalized ratings to sum to 1 and computed the expected value of the resulting probability distribution over percentage of blue gumballs for each utterance. As the center panel below shows, participants’ updated expectations transferred to interpretation: the expected value was lower in the *confident speaker* than in the *cautious speaker* condition for both utterances (*might*: $t(75) = -3.05, p < 0.01$, *probably*: $t(75) = -3.08, p < 0.01$).



Conclusion: We showed that semantic/pragmatic adaptation simultaneously affected production and comprehension, captured by a model that suggests a communicatively efficient system.

References: [1] Kleinschmidt and Jaeger. 2015. *Psych. Review*. [2] Kamide. 2012. *Cognition*. [3] Fine et al. 2013. *PLOS ONE*. [4] Yildirim et al. 2016. *JML*. [5] Goodman and Frank. 2016. *TICS*.

Thursday Poster.77

EYE MOVEMENT DATA AND THE CAUSES OF RELATIVE CLAUSE DIFFICULTIES

Douglas Roland (Waseda University), Yuki Hirose (The University of Tokyo) & Gail Mauner (University at Buffalo)
 droland@aoni.waseda.jp

While object relative clauses (ORCs) such as (2) typically take longer to read than analogous subject relative clauses (SRCs) such as (1), the opposite has been found when the embedded NP is pronominal. Several studies using self-paced reading have found that pronominal ORCs such as (5) are read more quickly than pronominal SRCs such as (4) (e.g., Reali & Christiansen, 2007; Roland et al. 2012; Heider et al., 2014). This has been taken as strong evidence for expectation-based accounts of language comprehension (e.g., Levy 2008), since memory-based accounts (e.g., Warren & Gibson, 2002; Gordon et al. 2001) do not predict that ORCs would ever be easier than analogous SRCs.

However, Roland et al. (2015) report that this reversal in costs is not observed in eye-movement data, suggesting that the reversal observed in self-paced reading is due to the nature of the self-paced reading task, rather than an inherent reversal in processing costs. However, it is difficult to individuate the effects of structure, NP type, and linear word order, because SRCs and ORCs differ in both word order and structure, and pronouns are likely to be read more quickly than full NPs.

We addressed these difficulties in an eye-tracking study using a comparison first seen in Staub (2010). We compared sentences containing full NP and pronominal ORCs, such as (2) and (5), with sentences containing the same word sequence in sentential complements, such as (3) and (6). While memory and expectation-based accounts both predict that full NP ORCs would be more difficult than pronominal ORCs, they differ in where that difficulty should emerge. Expectation-based accounts predict difficulties at the noun (*the chef*), due to the surprise of seeing a full NP rather than a pronoun or verb. Memory-based accounts predict difficulties at the verb (*telephoned*), due to retrieval and integration costs. While none of the theories explicitly address how the difficulties would be reflected in eye movement patterns, it is reasonable to assume that difficulty would be reflected in measures of first pass reading and regression (and not as rereading after regressing back from a later region).

Results at the NP (*you, the chef*), partially supported expectation-based accounts. Pronouns were read faster than full NPs in both first fixation ($p = .066$) and first pass duration ($p < .001$). Crucially, NP type interacted with sentence type in measures of probability of regression ($p < .001$), go-past ($p < .001$) and total reading ($p < .001$) times. This was due to an increase in regressions for full NP ORCs ($p < .001$), but not pronominal ORCs ($p = .249$). While there was no facilitation for pronominal ORCs relative to baseline, the additional difficulty for full NP ORCs is broadly consistent with expectation-based accounts. Memory-based accounts do not predict interactions this region.

Results at the verb (*telephoned*) were consistent with the predictions of memory-based accounts. We observed interactions between structure and NP type, indicating difficulty for full NP ORCs in first fixation ($p = .006$), first pass ($p < .001$), second pass ($p = .005$), and total reading ($p < .001$) times. The only effect of significance in go-past reading time was slower reading in ORCs ($p = .010$).

Taken as a whole, our results are consistent with expectation-based difficulties at the ORC NP and memory-based difficulties at the ORC verb. Because both pronominal and full NP ORCs are more difficult than their SC baselines, our results support Roland et al.'s conclusion that the reversed patterns observed in self-paced reading reflect task-based factors, and not an underlying reversal in processing costs.

NP type - Structure

- | | |
|------------|------------------------------------------------------------------------------------|
| 1 Full SRC | The event planner that telephoned the chef planned on having a party. |
| 2 Full ORC | The event planner that the chef telephoned planned on having a party. |
| 3 Full SC | The event planner worried that the chef telephoned her to cancel the party. |
| 4 Pro SRC | The event planner that telephoned you planned on having a party. |
| 5 Pro ORC | The event planner that you telephoned planned on having a party. |
| 6 Pro SC | The event planner worried that you telephoned her to cancel the party. |

APPROACHING SCALAR DIVERSITY THROUGH (RSA with) LEXICAL UNCERTAINTY

Chao Sun^{1,2} & Richard Breheny² (¹Leibniz Center for General Linguistics, ²University College London)

Previous research [1-2] suggests that different scalar expressions give rise to scalar inferences (SIs) at different rates. This phenomenon has become known as scalar diversity (SD). [2] showed that while quantifiers and modal expressions consistently gave rise to SIs, there was much greater variability within adjectives and verbs. As for the source of SD, [2] found that a modest amount of the variation in SI rates might be explained by boundedness of scales and 'semantic distance' - differences in perceived strength between scale mates. However a large amount of variation remains unexplained. Here we approach SD using insights from Rational Speech Act approaches that allow for local lexical enrichments, RSA-LU [3-4]. For scalar term 'some', the literal meaning could be characterised in terms of a set of possibilities where *some and not all* is the case or where *all* is the case: $\{\exists \& \neg \forall, \forall\}$. Possible enriched interpretations of 'some' are $\{\exists \& \neg \forall\}$ and $\{\forall\}$. In general, for scalar term W with literal meaning $\{\omega \& \neg \sigma, \sigma\}$, possible enriched meanings are $\{\omega \& \neg \sigma\}$ and $\{\sigma\}$. RSA-LU approaches lexical enrichment by assigning a prior probability to each of the possible lexical interpretations and capture the effect of lexical uncertainty by taking a weighted average over possible lexical interpretations at the level of the pragmatic listener, L1. According to RSA-LU, the likelihood of a scalar enrichment even for sentences containing an unembedded scalar term is impacted by the prior likelihood of each scalar enrichment for that term. It therefore predicts scalar diversity is possible if the priors for lexical enrichments vary across different scalar terms. We test this prediction by measuring the liability for each of the $\{\omega \& \neg \sigma\}$ and $\{\sigma\}$ enrichments for each scalar term, W , found in [2]. This is achieved by our 'so' and 'i.e.' tasks respectively. We find the predicted correlations with results on a replication of [2]'s inference task.

Exp. 1 We used 43 scalar expressions found in [2] to construct sentences of the form *S so not W*, where S is stronger than W , e.g. 'The student is brilliant so not intelligent'. 40 participants were asked to indicate how natural these constructions are on a Likert scale. *S so not W* sentences should be more coherent to the extent that W can be locally enriched to have an upper bound meaning $\{\omega \& \neg \sigma\}$. **Exp. 2** We used these scalar expressions to construct sentences of the form *W, i.e. S*, where S is stronger than W , e.g. 'The student is intelligent, i.e. brilliant'. Another 40 participants rated the naturalness of these constructions. *W, i.e. S* sentences should be more coherent to the extent that W can be locally enriched to the meaning of S , $\{\sigma\}$. In both experiments, each participant judged 43 experimental sentences and 7 fillers. **Exp. 3** In order to obtain a continuous measure of participants' judgments on the availability of SIs for each scalar pair $\langle W, S \rangle$, we replicated the inference task from [2] using the 0-100scale.

Results: As predicted by RSA-LU, we found the rating of '*S so not W*' was positively correlated with the rate of SIs ($r=0.35, p=.02$), while the rating of '*W, i.e. S*' was negatively correlated with the SIs rate ($r=-0.79, p<.001$). The ratings from two naturalness tasks do not correlate ($r=-0.22, p=.15$). A multiple regression analysis was conducted to predict the rate of SIs determined in Exp. 3 from ratings obtained from Exp.1-2 and other factors explored in [2]. The model accounted for 68% of the variance ($R^2=.75, F(9,32)= 10.51, p<.001$), with the so-task rating explaining 11.53%, the i.e.-task rating, 4.6% and boundedness of scales explaining 3.5%.

Discussion: We provide evidence that liability for lexical enrichments differ across different scalar terms. Also, the more likely a scalar term is to be enriched to its upper-bounded meaning, the higher the rate of SIs; the more likely a scalar term is to be enriched to the meaning of its stronger scale mate, the lower the rate of SIs. As predicted by RSA-LU, we relate SD to liability of local enrichments for each scalar term.

Reference: [1] Doran, R. (2009). *International Review of Pragmatics*, 1, 211–248. [2] van Tiel, B. van Miltenburg, E. Zevakhina N. & Geurts, B. (2016), *J Sem*, 33: 137-175. [3] Bergen, L., Levy, R., & Goodman, N. D. (2016). *Semantics and Pragmatics*, 9. [4] Potts et al. (2016) *J Sem*

Thursday Poster.79

ERPS DIFFERENTIATE BETWEEN TYPE OF LINGUISTIC INFORMATION DURING WORKING MEMORY MAINTENANCE OF SENTENCES

Matteo Mascelloni (University of Osnabrück/Trento), Roberto Zamparelli (University of Trento), Francesco Vespignani (University of Trento), Thomas Gruber (University of Osnabrück), Jutta L. Mueller (University of Osnabrück)
jutta.mueller@uos.de

Working memory for sentences may comprise both processes of language comprehension during encoding and processes of language production, especially during active maintenance via the phonological loop (Baddeley & Hitch, 1974; Meltzer et al., 2016). While the former processes are accessible via controlling the input with fine grained temporal resolution, the latter are more difficult to assess directly as language production is typically initiated and controlled internally. Thus, not much is known about the temporal characteristics of neurophysiological processes supporting different types of linguistic information while it is maintained in the phonological loop. In the present ERP study we aimed to tackle that question by devising a cued-production task. Native German participants (N=24) read different types of sentences word-by-word before undergoing a rehearsal phase during which a visual cue on the screen triggered the silent production of each individual word. In order to ensure task compliance ~10% of the sentences contained an unexpected overt production cue. Crucially, the sentences could either be entirely correct, or contain a semantic or a syntactic violation as in examples i) – iii).

- i) Der Maler verkauft das Bild für 500 Euro.
the painter sells the picture for 500 Euro.
- ii) *Der Maler verschüttet das Bild für 500 Euro.
*the painter spills the picture for 500 Euro.
- iii) *Der Maler verkauft den Bild für 500 Euro.
the painter spills the [+incorrect gender marking] picture for 500 Euro.

ERPs were time-locked to three potentially critical points in the sentences during reading as well as during silent production, namely to the verb, the subsequent article and the following noun. We used cluster-based permutation tests to detect significant ERP effects. While semantic and syntactic violations during reading elicited an N400 effect (362 – 515 ms after noun onset; semantic violation) and a P600 effect (509 – 950 ms after noun onset; syntactic violation), different ERP patterns occurred during the silent production phase. The silent production of semantically violated sentences elicited an *early fronto-central* negativity (114 – 213 ms) *at the verb* while the silent production of syntactically violated sentences elicited a *late right-frontal* positivity (580 – 650 ms) *at the article* following the verb. Thus, even though the linguistic violations do not come by surprise during silent production, they are associated with specific ERP effects indicating processes which are temporally and qualitatively different from the ones elicited during comprehension. Most likely, these effects are specific to the planning of sentence containing semantic and syntactic violations. In conclusion, active working memory maintenance processes for sentences are likely to comprise specific stages of language production that are potentially linked to ERP correlates of syntactic and semantic planning.

References:

- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
- Meltzer, J. A., Rose, N. S., Deschamps, T., Leigh, R. C., Panamsky, L., Silberberg, A., Links, K. A. (2016). Semantic and phonological contributions to short-term repetition and long-term cued sentence recall. *Memory & Cognition*, 44(2), 307–329.

Thursday Poster.80

VERBAL AND NON-VERBAL PREDICTORS OF WORD COMPREHENSION AND WORD PRODUCTION

Florian Hintz, Suzanne R. Jongman (both MPI for Psycholinguistics), James M. McQueen
(Radboud University), Antje S. Meyer (MPI for Psycholinguistics)
Florian.Hintz@mpi.nl

Individuals differ greatly in how quickly and accurately they recognize and produce words. However, despite obvious individual differences, psycholinguistic theories have thus far largely ignored variability among language users. One reason is that previous studies assessed listening and speaking abilities almost exclusively in highly educated, homogeneous student populations, aiming for low participant variability. However, more recent research suggests that including individual differences may be a powerful way to address a wide range of topics, among others potential interactions between the language system and non-verbal cognitive systems.

Using an individual differences approach, the goal of the present study was to investigate which verbal and non-verbal systems are engaged during word processing. Participants were recruited from universities (N = 78) and vocational colleges (N = 60) so as to sample from diverse educational backgrounds. We administered an auditory lexical decision (LD) task and a picture naming (PN) task; both have been instrumental in formulating theories of comprehension and production. Target words in both tasks varied in word frequency, while phonological neighborhood density and word prevalence (degree to which a word is known by a representative sample of population) were controlled. We also administered tests tapping verbal and non-verbal participant variables. The Peabody Picture Vocabulary Test (PPVT) was used to assess participants' receptive vocabulary size. Participants carried out various non-verbal processing-speed tasks. Using principal component analysis, we derived a common factor representing participants' processing speed. Lastly, participants' non-verbal intelligence was measured using Raven's Advanced Progressive Matrices (APM). Participant variables were weakly to moderately correlated (Processing speed x Raven's APM: $r = -.32$, Processing speed x PPVT: $r = -.43$, Raven's APM x PPVT: $r = .50$). We fitted mixed-effects models for four dependent variables (accuracy and reaction time [RT] in LD and PN); all models included 'Word frequency', 'PPVT', 'Processing speed' and 'Raven's APM' as continuous predictors (all scaled and centered).

Lexical decision task: Accuracy (d') depended on vocabulary size and non-verbal intelligence such that participants with larger vocabularies and higher non-verbal intelligence identified words more accurately than participants who scored lower on these tests. RT for existing words in the LD task depended on target word frequency, processing speed and non-verbal intelligence: High frequency words were responded to faster than low frequency words. Individuals who responded faster on the processing speed tasks had faster RTs to existing words than individuals with low processing speed abilities. Interestingly, higher non-verbal intelligence was associated with longer RTs to existing words. Additional analyses showed that this unexpected result could be explained by a speed-accuracy tradeoff.

Picture naming: Accuracy (correctly named items) was predicted by word frequency, vocabulary, processing speed and non-verbal intelligence. Pictures with frequent names were more often named correctly than pictures with infrequent names. Larger vocabularies, higher non-verbal intelligence and faster processing speed also led to more accurate picture naming. RT (speech onset latencies) was predicted by word frequency and processing speed: High word frequency and faster non-verbal processing speed abilities were associated with faster picture naming.

Taken together, our data suggest that word frequency and non-verbal processing speed are important predictors of the speed with which individuals recognize and produce words. Accuracy in these tasks was additionally predicted by participants' non-verbal intelligence and receptive vocabulary size. Adding to a growing body of research, our study thus suggests that linguistic processing at the word level engages a multitude of verbal and non-verbal abilities. These abilities must be taken into account by theories of word comprehension and production.

THE EFFECTS OF VISUAL INFORMATION IN LATE BILINGUAL LANGUAGE PROCESSING. EYE-TRACKING STUDY

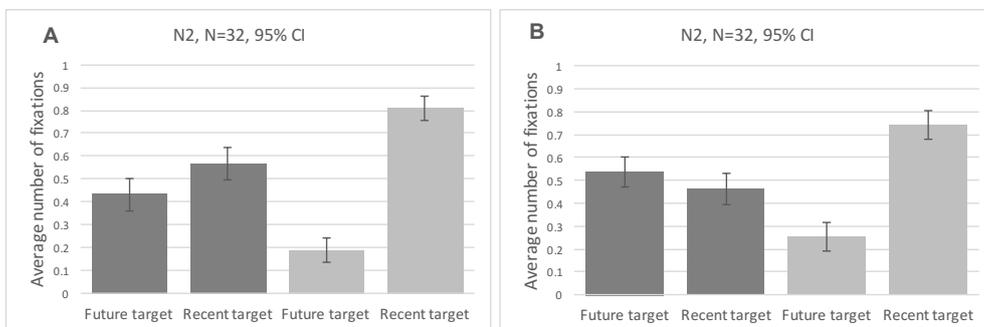
Dato Abashidze, Matthew Martin, Pavel Trofimovich & Kim McDonough
 Concordia University, Montreal, Canada
 dato.abashidze@concordia.ca

Recently seen events influence gaze patterns during language comprehension. When monolingual participants see an action event and they hear a related sentence, they preferentially inspect targets of recent events over plausible future-event targets, independent of tense information [1]. Bilinguals, who experience increased processing demands due to competition between their two languages [4], may have weaker lexical access [2] and weaker predictive processing abilities [3] relative to monolingual speakers. It is currently unclear how bilinguals can use recent event information as predictive cues and how much they rely on visual cues during spoken sentence processing.

This eye-tracking study ($N = 32$) examined French-English bilinguals' reliance on recently seen events, focusing on their ability to predict a plausible future action during spoken sentence comprehension. Participants, exposed to the second language (English) after the age of seven, saw a videotaped actor performing an action (e.g., sweetening strawberries, see [1] for sequence of events in an experimental trial) and then listened to an English (NP1-Aux-Verb-NP2) sentence (e.g., *The experimenter has sweetened the strawberries*) referring to that recently performed action or heard an alternative sentence (e.g., *The experimenter will sweeten the pancakes*) referring to an equally plausible action that the actor would perform next (sweetening pancakes). Eye movements to the recent and future objects were analyzed during the sentences.

Preliminary results indicate that bilinguals performed similarly to monolinguals, preferring to inspect recent-event targets when exposed to both tenses (referring to a recent vs. future action). Although bilinguals showed a decrease in eye-gaze frequency toward recent-event targets (compared to monolinguals) when listening to future sentences, they inspected the plausible future target more than the recent-event target only during the second noun phrase (see, Fig. 1, B). Bilinguals and monolinguals thus appear to be constrained by similar processing biases in their comprehension of spoken discourse in the presence of visual information.

Figure 1. Average number of fixations in the NP2 word region of the monolingual (A) and bilingual groups (B).



[1] Abashidze & Chambers (2016); [2] Shook, et al. (2014); [3] Kaan et al. (2010); [4] Chabal & Marian (2015)

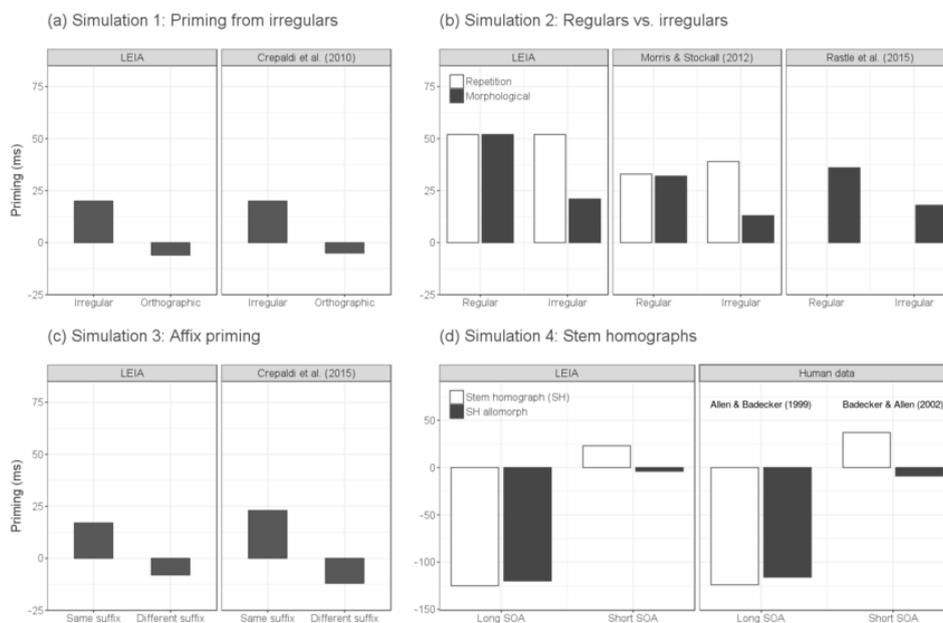
Taking it a level higher: The LEIA model of complex word recognition
João Verissimo, University of Potsdam

Many accounts of the recognition of complex words make use of 'higher' lexico-semantic representational levels, which interact with morphologically decomposed representations at a lower orthographic level (e.g., Crepaldi et al., 2010); however, such theoretical proposals have not yet been formalised into explicit models that can generate quantitative predictions. At the same time, most (localist) computational models of visual word recognition are only defined up to the 'orthographic word-form' level, without syntactic or semantic representations. Here, we aim to bring these two research traditions together and we mitigate their gaps by proposing a novel, implemented computational model of the visual recognition of complex words.

The Lemma-Extended Interactive Activation Model (LEIA) extends the original IA model of visual word recognition (McClelland & Rumelhart, 1981) with a layer of morphosyntactic lemma nodes. The other main tenets of our model are: (a) obligatory morphological decomposition into stems and affixes; (b) structured lemma nodes that may encompass distinct stems; and (c) excitatory and inhibitory links between lemma nodes.

In four sets of (proof of principle) simulations, all with the same parameter values, we show that LEIA produces close quantitative fits to a variety of previously reported morphological priming effects. Specifically, as shown in the figure, we were able to successfully simulate: masked priming from irregular inflections, but weak inhibitory effects from orthographic neighbours (*fell*→*fall* vs. *full*→*fall*) (panel a); masked priming from regular inflections, which was larger than priming from irregulars (*walked*→*walk* vs. *fall*→*fell*) (panel b); masked priming from suffix-sharing pseudowords, but weak inhibition from different suffixes (*sport-er*→*teach-er* vs. *sport-al*→*teach-er*) (panel c); and, finally, strong inhibition from homographic stems (Spanish *cerr-ar* / *cierr-as* 'close inf/2sg'→*cerr-os* 'hills'), which was present for overt primes, but disappeared in masked priming (panel d). Additional simulations showed that this timecourse effect critically changes around 70ms (the typical threshold for masked primes), even though there are no distinct processing stages in LEIA.

The combined results suggest that a higher lemma level, with properties like the ones outlined above, may be an indispensable assumption in a fully specified theory of morphological processing. In addition, our model highlights the appropriateness of interactive activation frameworks for capturing the timecourse of complex word recognition.



INVESTIGATING FACTORS THAT INFLUENCE THE INTERPRETATION OF AMBIGUOUS PHRASES AS LITERAL OR SARCASTIC

Ruth Filik (University of Nottingham), Christina Ralph-Nearman (LIBR) & Rachel Giora (Tel Aviv University)

In the current study we aim to examine factors that may influence the interpretation of ambiguous phrases, specifically, regarding whether they should receive a literal or sarcastic reading. A number of factors are investigated; some relating to properties of the message itself (e.g., valence, punctuation, and negation), and some to properties of the perceiver (e.g., age, personal tendency to use sarcasm, and tendency to engage in indirectly aggressive acts). Findings are discussed in relation to contemporary theories of sarcasm comprehension.

In **Experiment 1**, 164 native English-speakers (aged 18-84) rated 36 ambiguous written utterances (presented in context) in relation to how sarcastic they thought the character was being. Utterances were either superficially positive (e.g., *It was so interesting*) or negative (e.g., *It was so boring*), and were accompanied by a full stop, wink emoticon, or ellipsis (...). Participants also completed a self-report sarcasm survey (Ivanko et al., 2004) and indirect aggression questionnaire (Forrest et al., 2005). Results showed that utterances were rated as most sarcastic when accompanied by a wink, less sarcastic with an ellipsis, and least sarcastic with a full stop (supporting Filik et al., 2016). Comments with a negative valence were rated as more sarcastic than those with a positive valence. Sarcasm rating scores positively correlated with self-reported levels of sarcasm use and tendency to use indirect aggression, and negatively correlated with participant age.

In **Experiment 2**, we wished to further investigate the hypothesis that ambiguous utterances in which negation is used to mitigate a highly positive concept (e.g., *He's not the best lawyer*) are interpreted sarcastically by default (Giora et al., 2015). 162 native English-speakers (aged 18-74) rated 28 negative phrases (e.g., *This isn't the most hygienic restaurant*), regarding whether they conveyed a sarcastic or literal message, before again completing a self-report sarcasm survey and indirect aggression questionnaire. Overall, participants rated materials as sarcastic, supporting the predictions of the Defaultness Hypothesis. Interestingly, results across both Experiments 1 and 2 showed a negative correlation between participant age and self-reported use of sarcasm, which would concur with the finding (in Experiment 1) that the tendency to interpret ambiguous materials sarcastically was negatively correlated with age. These results fit well with recent findings suggesting that older adults have a greater tendency to misinterpret comments that are intended sarcastically (Phillips et al., 2015).

In conclusion, results would indicate that a broad range of factors (both text-based and perceiver-based) can influence the sense of sarcasm that is experienced when reading ambiguous stimuli, offering support for constraint-satisfaction approaches (e.g., Campbell & Katz, 2012), and suggesting further 'constraints' that need to be taken into consideration.

Campbell, J. D., & Katz, A. N. (2012). Are there necessary conditions for inducing a sense of sarcastic irony? *Discourse Processes*, 49, 459-480.

Filik, R., Turcan, A., Thompson, D., Harvey, N., Davies, H., & Turner, A. (2016). Sarcasm and emoticons: Comprehension and emotional impact. *Quarterly Journal of Experimental Psychology*, 69, 2130-2146.

Forrest, S., Eatough, V., & Shevlin, M. (2005). Measuring adult indirect aggression: The development and psychometric assessment of the indirect aggression scales. *Aggressive Behavior*, 31, 84-97.

Giora, R., Givoni, S., & Fein, O. (2015). Defaultness reigns: The case of sarcasm. *Metaphor and Symbol*, 30, 290-313.

Ivanko, S. L., Pexman, P. M., & Olineck, K. M. (2004). How sarcastic are you? Individual differences and verbal irony. *Journal of Language and Social Psychology*, 23, 244-271.

Phillips, L. H., Allen, R., Bull, R., Hering, A., Kliegel, M., & Channon, S. (2015). Older adults have difficulty in decoding sarcasm. *Developmental Psychology*, 51, 1840-1852.

BILINGUAL'S REFERENTIAL CHOICE IN COGNITIVELY DEMANDING SITUATIONS

Carla Contemori & Iva Ivanova (University of Texas at El Paso)
 ccontemori@utep.edu

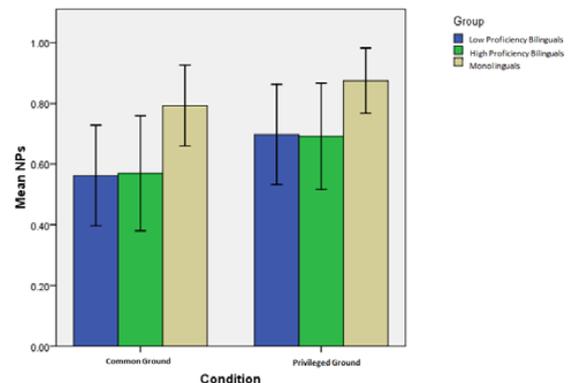
Appropriate references are a prerequisite for successful communication, but monolinguals and bilinguals differ in their choice of referential expressions. Native English speakers use attenuated forms (pronouns) when the referent is the topic of conversation, and more explicit forms (full noun phrases) when introducing new entities in the discourse (Arnold & Griffin, 2007). However, even unbalanced but highly-proficient bilinguals may over-use pronominal forms in a null-subject second language (L2) (Contemori & Dussias, 2016; Belletti, Bennati, & Sorace, 2007). Such residual indeterminacy in L2 referential choice could be due to an increased need for cognitive resources when computing interface structures between syntax and pragmatics as are referential expressions (Interface Hypothesis, Sorace, 2011). Such structures are less likely to be successfully used by bilinguals than structures without this interface because speaking an L2 is overall cognitively costly and might also require cognitive resources to prevent between-language interference.

The present study tests the ability of the Interface Hypothesis to explain bilingual referential choice by examining the production of pronouns and noun phrases in unbalanced Spanish-English bilinguals in common and privileged ground (i.e., when the preceding discourse is shared by speaker and listener or only known to the speaker, Fukumura & vanGompel, 2012). The privileged ground condition is potentially more cognitively effortful, because the speaker has to consider the addressee's discourse model and choose a more explicit referring expression (i.e., a noun phrase). According to the Interface Hypothesis, in privileged ground bilinguals in their L2 may be more likely to differ from monolinguals by producing fewer specified references (i.e., noun phrases), relative to common ground.

English monolinguals and low- and high-proficiency Spanish-English bilinguals participated in a story-telling task in English (bilinguals' L2). Participants saw two pictures of a male and a female character performing different actions, then heard a two-sentence description of the first picture, and then produced descriptions of the second picture to a confederate. The second picture cued references to the character that was salient (e.g., *the boy*) or non-salient (e.g., *the girl*) in the preceding discourse. The second context sentence (e.g., *The boy got really annoyed*) was either presented to both participant and confederate (common ground condition) or only to the participant (privileged ground condition).

Participants produced more noun phrases in the privileged ground than in the common ground condition ($\beta=-0.4$, $SD=0.1$, $t=-2.313$, $p<0.02$), and monolinguals produced more noun phrases than low-proficiency ($\beta=0.20$, $SD=0.09$, $t=2.172$, $p<0.03$) and high-proficiency bilinguals ($\beta=0.19$, $SD=0.09$, $t=2.138$, $p<0.03$). However, the difference between bilinguals and monolinguals was similar in the common and privileged-ground conditions (no interaction between group and condition, $\beta=0.08$, $SD=0.1$, $t=-0.515$, $p=0.6$).

These results reveal that both monolinguals and bilinguals are sensitive to privileged ground, and tend to give more explicit referring expressions when the information is not shared with the addressee. Also, as in prior studies, bilinguals tended to use fewer explicit references (noun phrases) than monolinguals, possibly because of a difficulty to evaluate discourse salience which resulted in the use of expressions more economical for the speaker (Contemori & Dussias, 2016). However, the difference in referential use between bilinguals and monolinguals was similar in common and privileged ground, inconsistent with the Interface Hypothesis. In an on-going experiment, we further test the Interface Hypothesis in a picture-description task under verbal and non-verbal cognitive load. A greater difference in referential use between bilinguals and monolinguals under load would support the Interface Hypothesis.



ON THE RELATIONSHIP BETWEEN LEXICAL PROCESSING SPEED AND VOCABULARY IN TODDLERS

Seamus Donnelly and Evan Kidd (seamus.donnelly@anu.edu.au)

Much research shows that the speed of early lexical processing, as measured by the Looking While Listening Task, predicts vocabulary size both concurrently and retrospectively (Fernald, Perfors & Marchman, 2006, Marchman & Fernald, 2012). In this task, children see two images (e.g., a ball and a car) and hear a sentence such as *Look at the ball*. An infant's reaction time (RT) is defined as the average time it takes to look to target image in those trials where the participant was looking at the distracter image prior to the onset of the target word. The measure is thought to reflect lexical processing efficiency; however, the interpretation of average reaction times on this task and the nature of its relationship with vocabulary size remains unclear. There are two possible interpretations of average RTs on this task. First, it is possible that RTs on this task reflects some centralized lexical (or cognitive) processing speeds and that every word represents a sample of other that underlying speed. Second, it is possible that average RTs are an emergent consequence of a number of *potentially independent* word-specific processing speeds. These two theoretical perspective correspond to two classes of structural equation models--effect and causal indicator models, respectively—and could, therefore, be compared using these models. Moreover, while there are many studies showing correlations between vocabulary size and lexical processing speed, very little research has processing speed predicts growth in vocabulary over and above prior vocabulary and whether vocabulary predicts processing speed over and above prior vocabulary. In this talk, we address these two questions using a subsample ($N = 91$) of participants from the Canberra Longitudinal Child Language Project, a longitudinal study of individual differences in language processing. We administered the MacArthur-Bates Communicative Development index and Looking While Listening task at 18, 21 and 24 months. To determine whether lexical processing speed was better modelled as a central processing speed or a collection of word-specific processing speeds, we fit effect- and causal indicator models to the LWL task at 18 months. Both models fit the data adequately ($RMSEA_{effect} = .00$, $RMSEA_{causal} = .00$), and a nested vanishing tetrad test did not provide evidence for a preference of either of the models ($\chi^2(27) = 6.91$, $p = .99$). These results suggest that both of these conceptualizations of lexical processing speed are plausible. Next we tested whether vocabulary and lexical processing speed independently predicted growth in one another. We fit the path analysis illustrated in Figure 1A. As can be seen in Figure 1B, RTs at 18 months predicted vocabulary at 21 months over and above vocabulary at 18 months, and vocabulary at 18 months predicted RTs at 21 months over and above RTs at 18 months. However, these relationships disappeared in the 21 to 24 month interval. This suggests that the two variables are initially reciprocally causally related but become dissociated across development.

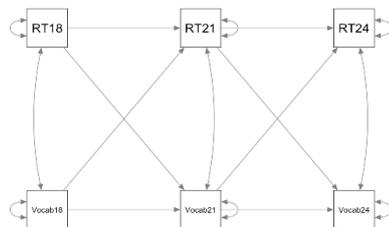


Figure 1A: Path Analysis Specification. Straight arrows represent causal paths. Curved Arrows represent correlations.

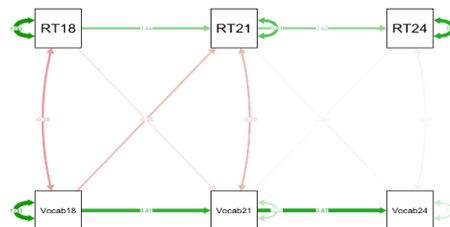


Figure 1B: Path Analysis with Parameter Estimates. Edge width represents the magnitude of the relevant coefficient.

Thursday Poster.86

THE “PRODUCTION P2” EFFECT PRIMARILY REFLECTS TRAINING IN PICTURE NAMING

Agata Wolna (Jagiellonian University), Jakub Szewczyk (Jagiellonian University),
Patrycja Kałamała (Jagiellonian University), Jonas Walther (Jagiellonian University),
Zofia Wodniecka (Jagiellonian University)

agatawuwu@gmail.com

Naming pictures in L1 is considerably slowed down when it is preceded by naming pictures in L2. This phenomenon is known as the L2 after-effect¹. On the ERP level, it is accompanied by the N300 component (more negative ERPs for L1-after-L2 naming in the 250-350ms time-window²).

However, the mechanism underlying the L2 after-effect is unknown. On the one hand, it can be language-specific, i.e. reflect cross-linguistic interference between lexical units that are activated in L2 and L1 blocks. On the other hand, the mechanism can be domain-general, reflecting a task change between L1 and L2 naming.

To adjudicate between these two possibilities, we designed an ERP study consisting of three pairs of blocks: L1 naming after L1 naming (L1-after-L1), L1 naming after L2 naming (L1-after-L2), and L1 naming after a non-linguistic task (L1-after-NLT). The order of the two last pairs of blocks was counter-balanced across participants. Overall, 33 participants named 250 unique pictures. The pictures were rotated across participants and across the 5 blocks of picture naming (the sixth block was the NLT).

We hypothesized that if the after-effects are domain-general, then the NLT should affect subsequent L1 naming in the similar way as L2 naming, i.e. lead to longer naming latencies and more negative N300 component, compared to the baseline condition (L1-after-L1). Conversely, if the after-effects are language-specific, we should observe the N300 component in the L1-after-L2 condition only.

On the behavioural level, we observed longer naming latencies in the L1-after-L2, than in the two remaining types of blocks, indicating that the L2 after-effect is language-specific. The ERP data brought a surprising effect: we observed a large fronto-central modulation in the 150-250 ms time-window (i.e. preceding the N300 effect), which resembled the “production P2” effect reported in the literature^{3,4}. The strongest predictor of its amplitude was the trial number: Initially the effect was strongly negative and it became more positive with each subsequent trial. We interpret this effect as training in picture naming. This decrease in “production P2” amplitude did not occur in the L1-after-NLT block, suggesting that the change of task disrupted the training effect. Finally, the amplitude of the component was more negative in the L1-after-L2 block than in the remaining blocks (keeping other effects constant), implying that the “production P2” effect also has a lexical component. Overall, the unexpected findings suggest that: 1) the “production P2” effect reflects both lexical effects and training in picture naming; 2) the “production P2” effect is dissociable from naming latencies; 3) to reliably measure the P2 and the N300 effects, the order of all blocks must be fully counterbalanced.

References: 1) Branzi et al. (2014). *Neuropsychologia*, 52, 102–116. 2) Wodniecka et al., (under review) <https://osf.io/vcsyn/> 3) Strijkers et al (2010). *Cerebral Cortex*, 20(4), 912–928. 4) Strijkers et al. (2011) *Journal of Memory and Language*, 65(4), 345–362

Thursday Poster.87

**PERCEPTUAL PRIMING AND SYNTACTIC CHOICE IN RUSSIAN
LANGUAGE: MULTIMODAL STUDY**

Mikhail Pokhoday, Yury Shtyrov and Andriy Myachykov
(mikhail.pokhoday@gmail.com)

In a fully developed production system, perception provides an input of information about the event, attention foregrounds relevant/important information for the conceptual analysis, and subsequent language production mechanisms collaborate to generate speech (Levelt, 1989). A part of this complex process is the necessity to select between simultaneously available syntactic alternatives. For example, English language provides several options that can describe the same visual event, e.g., an officer chasing a burglar. These minimally include (1) The officer is chasing the burglar and (2) The burglar is (being) chased by the officer. These active- and passive-voice alternatives differ in assigning object and subject roles to agent (officer) and patient (burglar). Existing evidence suggests that the system responsible for assigning the grammatical roles is sensitive to the distribution of the speaker's attention within the described scene (Tomlin & Myachykov, 2015, for a recent review). Specifically, a speaker of English is more likely to choose a passive-voice frame when her attention is directed to the patient of the described event and she is more likely to use an active-voice frame when the agent is in her attentional focus (e.g., Myachykov, et al., 2012). While this and other studies indicate a regular interplay between attention and syntactic choice, they also exclusively used variants of the visual cueing paradigm (Posner, 1980). As a result, the reported link between attention and syntactic choice cannot be generalised beyond the visual modality. A more ecologically valid proposal needs to take into account a multi-modal nature of attention.

Here, we report results of a series of sentence production experiments, in which Russian native speakers described visually presented transitive events (e.g. kick("pinat"), chase ("presledovat'/ubegat"). In half of the trials the agent appeared on the left and in the other half – on the right. Speakers' attention to the referents was manipulated by means of lateral cues. In Experiment 1 by visual cue (a red circle); in Experiment 2 – auditory (beep played monaurally); in Experiment 3 – motor (participants were prompted to press a left or a right key depending on the color of the central fixation cross). Hence, the Cued Referent (Agent/Patient) was crossed with the Cue Type (Visual, Auditory, Motor). The proportion of the sentences where the cued patient referent was put in the sentence before agent was the dependent variable. In Experiment 1 we registered a main effect of visual cue location – patient has been chosen as a starting point in the sentence more often when he had been cued: $X^2(1) = 4.15, p=.042$. Also, there was a main effect of event orientation – Russian speakers produced more patient-first sentence when the patient was on the left in the picture: $X^2(1) = 3.91, p=.048$. There were however no interaction of those factors. In Experiment 2 there was no effect of auditory cue, but there was a strong effect of event orientation with more patient first structures produced when the action on the picture was right-to-left: $X^2(1) = 5.23, p=.022$. Data of Experiment 3 is now collected and will be reported. Overall these results as well as English language experiments suggest an existence of a hierarchy in effects of modality of primes on syntactic choice with an interesting addition that Russian speakers tend to be more affected by event orientation than their English speaking counterparts.

References:

- Gleitman, L.,R., January,D., Nappa,R. and Trueswell, J.C. (2007). On give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544-569.
- Myachykov, A. and Tomlin, R. S. (2014). Attention and salience. In Dabrowska, E., & Divjak, D. (Eds.). (2015). *Handbook of Cognitive Linguistics*(Vol. 39). Walter de Gruyter GmbH & Co KG.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, 1191(1), 182-200.
- Tomlin, R. S. (1995). Focal attention, voice, and word order: an experimental, cross-linguistic study. *Word order in discourse*, 517-554.

HEAD DIRECTIONALITY INTERACTS WITH DEPENDENCY LENGTH

Himanshu Yadav¹ (JNU, New Delhi), Ashwini Vaidya² & Samar Husain³ (IIT Delhi)
¹yadavhimanshu059@gmail.com, ²ashwini.vaidya@gmail.com, ³samar@hss.iitd.ac.in

Dependency length minimization has been claimed to be a crosslinguistic processing constraint (Futrell et al., 2015; Liu et al., 2017). According to this hypothesis, a head and its dependent are more likely to be close than further away due to working memory constraints. Contrary to the dependency minimization hypothesis, considerable research (e.g., Levy and Keller, 2013) has shown that in head-final languages, head-dependent length can be long. This research points to parser adaptability based on certain typological properties of a language (Vasishth et al., 2011; Levy et al., 2013). This ‘adaptability hypothesis’ predicts that there should be an interaction between dependency length and head directionality. Recent crosslinguistic corpus work (Shukla et al., 2018) has shown that this interaction is indeed observable in the case of crossing dependencies. The current work investigates whether dependency length is determined by the interaction of the dependency direction and the dominant head-directionality in the language for all dependency types.

We choose 12 languages (Head-final:6, Head-initial:6) annotated with Universal Dependencies (Nivre et al., 2016). We classify dependencies as (a) Head-F or (b) Head-P.

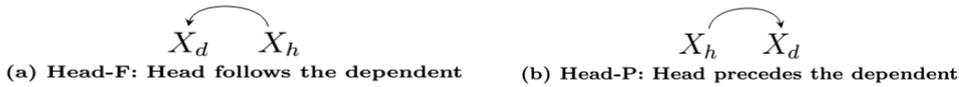
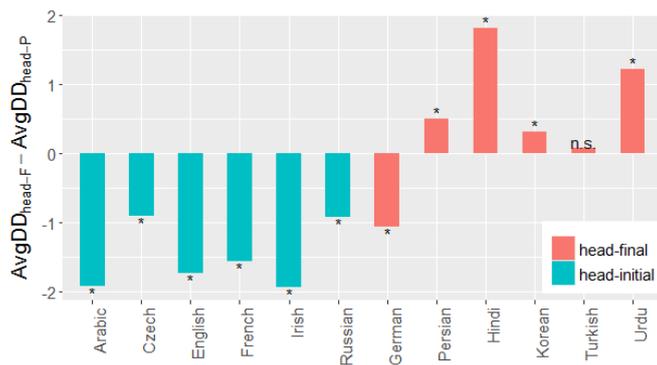


Figure 1: Dependency based on head-dependent configuration. X_d = Dependent, X_h = Head. The corpora chosen for the 12 languages came from news, blog or wiki genre. Total number of sentences in each language was > 3500 (except Irish). We computed the average dependency distance (AvgDD) for both Head-P and Head-F configurations in each language using (1).

$$(1) \text{AvgDD} = \frac{1}{n} \sum_{i=1}^n DD_i \quad \begin{array}{l} n = \text{total number of dependencies.} \\ DD_i = \text{dependency distance for the } i^{\text{th}} \text{ dependency,} \end{array}$$

The adaptability hypothesis predicts that AvgDD for a particular syntactic dependency should interact with the dominant head directionality of a language, i.e., AvgDD for Head-F should be more than Head-P in head-final languages and vice-versa for head-initial languages. We test this hypothesis for all syntactic dependencies where the proportion of Head-P and Head-F configurations are comparable.

The results are consistent with this hypothesis, difference in AvgDD between Head-F and Head-P cases was positive for head-final languages and negative for head-initial languages (Figure 2) ($p < 0.05$, except 1 using Wilcoxon test). Length of a dependency is larger if its head direction aligns with the dominant head directionality of a language (OV/VO). Crucially, this holds even after controlling the Head-F vs Head-P proportion difference for syntactic dependencies. This suggests that the dominant head directionality of a language determines the dependency length across such dependency types.



This work points to parser adaptability with respect to length of a dependency based on certain typological properties of a language (cf. Vasishth et al., 2011; Levy et al., 2013). While average head-dependent length is being minimized cross linguistically (Futrell et al., 2015), our work shows that dependency length is also being influenced by the typological properties of a language. || References: Futrell et al. (2015) PNAS; Liu et al., (2017) Life of Phy Reviews; Levy and Keller, (2013) JML; Vasishth et al. (2011) LCP; Levy et al. (2013) JML; Shukla et al. (2018) CUNY ||

THE DISCOURSE BEHAVIOUR OF WEAK DEFINITES IN GERMAN

Frederike Weeber & Klaus von Heusinger (University of Cologne)

fweeber1@uni-koeln.de

We present empirical evidence on the discourse behavior of weak definites, such as *go to the doctor*. We show (i) that weak definites are less accessible for discourse anaphora than (regular) indefinites and definites and (ii) that this effect is driven by two factors: the local context and the morphological marking at the noun phrase (NP). Both these factors seem to independently contribute to the weak referential reading of weak definites.

Weak definites are definite NPs that, unlike regular definites, do not imply global uniqueness of its referent. They always take narrow scope and, again unlike regular definites but also unlike indefinites, express enriched meanings (Aguilar Guevara & Zwarts, 2010; Carlson et al., 2006). Importantly, weak definites seem to have limited ability to establish a discourse referent. While this claim has been widely accepted in the more theoretical literature, there has been little empirical evidence.

German has a strong and a weak form of the definite article (Schwarz, 2009), and the latter merges with specific preceding prepositions. While regular definites may or may not merge with a preposition, weak definites always do so. German then provides a good test case for contrasting weak definites with both regular indefinites and definites. To that end, we conducted a comprehension and a production study and asked whether weak definites indeed have a weaker potential to establish a discourse referent than regular indefinites and definites. In both experiments, we took discourse anaphora to reflect referent accessibility and thus to reflect the establishment of a discourse referent.

In Expt1, a visual world eye tracking study, 20 participants (ps) listened to stories like the one in (1, from German). They consisted of a context sentence, a sentence introducing two human referents, and a subsequent target sentence that included an ambiguous personal pronoun. The subject of the second sentence was a proper name and the object appeared inside a goal PP to make the weak definite article visible. During story presentation, four pictures appeared onscreen: the subject (*Thomas*), the critical object (*lawyer*), and two unrelated object distractors. Mean fixation times time-locked to the onset of the pronoun showed that ps looked more to the object picture when the PP hosted an indefinite than when it hosted a weak definite. The regression model including the weak vs. indefinite distinction better explained the data than the null model $X^2(1) = 4.75, p = .029$. We also found more looks to the subject picture in the weak than the indefinite condition, $X^2(1) = 2.99, p = .084$.

In Expt2, a story continuation task, we asked 90 ps to complete short stories in a natural way. We used the stories from Expt1 with two exceptions: First, we included strong definite NPs in addition to the weak definite and indefinite NPs and, second, we tested weak and non-weak, i.e. regular contexts, cf. (2). Our data show that overall ps mentioned the referent inside the PP less often in weak than in regular contexts, in particular when it was introduced by the weak definite article. Our regression analyses with re-mentions as dependent measure revealed a main effect of context, $b = .51, SE = .19, z = 2.7, p = .007$, as well as a significant difference between weak definites and strong definites on the one hand, $b = .35, SE = .13, z = 2.69, p = .007$, and weak definites and regular indefinites on the other, $b = .33, SE = .13, z = 2.5, p = .013$, across contexts.

- (1) *The breach of the contract caused immense damage. Thomas went to the weak to a lawyer. Because this was during the summer holidays, he had plenty of time.*
- (2) *The breach of the contract caused immense damage. Thomas complained to the lawyer. Because this was during the summer holidays, _____.*

In sum, our data support the view that weak definites have a limited capacity to establish discourse referents and are therefore weaker competitors in ambiguous pronoun resolution than indefinite NPs (Expt1). Also, weak referential readings of weak definites are independently driven by both weak contexts and their specific morphological marking (Expt2).

Aguilar Guevara, A. & J. Zwarts. 2010. *Weak definites and reference to kinds*. In SALT 20, 179-196.; Carlson, G., R. Sussman, N. Klein & M. Tanenhaus. 2006. *Weak definite noun phrases*. In Proceedings of NELS 36, 179-196. Amherst, MA: GLSA.; Schwarz, F. 2009. *Two types of definites in natural language*. Univ. of Massachusetts Amherst dissertation.

Poster Presentations - Abstracts Friday

Poster Abstracts Friday, Sept 07, 2018

Friday Poster.1

(EVENT) SEMANTIC CUES FOR PRONOUN REALIZATION

Semra Kizilkaya (University of Cologne)

semra.kizilkaya@uni-koeln.de

Pronominal paradigms of many German dialects differentiate full and clitic pronouns. While the alternation of pronominal forms has mostly been described from a syntactic and pragmatic point of view, the influence of verbal semantic factors like event structure is understudied. I will present evidence from Bavarian illustrating that event semantics significantly affects the acceptability of full pronouns.

In Bavarian, the third person singular masculine (3SG.M) direct object pronoun can take two morphological forms, the full pronoun *eam* and the clitic pronoun *=n*. An alternation of both forms is only possible in the position adjacent to the finite verb. While the unstressed clitic *=n* is seen as the unmarked form of reference, the stressable full pronoun *eam* is generally associated with markedness or a focus-inducing function (Altmann 1984, Weiß 1999). In recent literature, an additional semantic restriction is postulated for the full pronoun, according to which *eam* can only refer to animate entities (Stangel 2015, Weiß 2016).

Findings of a pilot corpus study suggested that *eam* can also refer to inanimates, but only in combination with verbs whose lexical semantics select high degrees of affectedness. This led to the hypothesis that affectedness has an effect on the acceptability of the full pronoun *eam*. Affectedness is used to describe a two-dimensional event semantics, where (i) a theme participant undergoes change and (ii) a scale participant measures the change (Beavers 2011: 335). See type verbs are underspecified for affectedness (1a), whereas *throw* type verbs obligatorily entail change in their theme arguments, thus an affected event structure (1b).

To test the hypothesis, an acceptability judgement task was carried out. It consisted of eight conditions in a 2 x 2 x 2 factorial design, whereby the animacy of the antecedents (animate vs. inanimate) as well as verb type (affected (*throw* type) vs. non-affected (*see* type)) were manipulated. For each condition, 3 critical items were constructed and manipulated for type of anaphoric expression (full pronoun vs. clitic pronoun). The 24 critical items, in addition to 20 fillers, were distributed across two balanced lists and rated by 30 native speakers of Bavarian on each list (60 participants in total; mean age: 41 years, range: 18-81 years). The statistics was analyzed in R, using *lme4* (Bates et al. 2014) to perform a linear mixed effects (LME) analysis with animacy, affectedness and pronominal form as fixed effects, and participants and items as random effects.

The results show a main effect for affectedness with inanimates: Full pronouns are rated significantly better in affected event structural contexts (*throw* type verbs) in comparison to non-affected contexts (*see* type verbs) ($b = -0.5779$, $SE = 0.1984$, $t = -2.912$). The analysis shows a three way interaction of animacy x affectedness x pronominal form ($b = -0.89805$, $SE = 0.36847$, $t = -2.437$) and testifies full pronouns to be an (unmarked) alternative to clitics in affected verbal semantic contexts.

The results confirm the hypothesis that affectedness and therefore event semantics has an effect on pronominal realization in Bavarian. Crucially, Elman, Kehler & Rohde (2006) had already shown pronoun interpretation to be sensitive to verbal properties like the degree of affectedness. To understand and unfold the function of affectedness for pronominal form, further cross-linguistic studies are necessary.

- (1) a. I hob=*n/eam* gseng. to see, no entailed change in theme argument
'I saw him'
- b. I hob=*n/eam* hi-gschmissn. to throw down, entailed change in theme
'I threw him down' argument, scale: location

Selected References

Elman, Jeffrey L., Andrew Kehler and Hannah Rohde. 2006. Event structure and discourse coherence biases in pronoun interpretation. Proceedings of the Annual Meeting of the Cognitive Science Society. • Stangel, Ursula. 2015. Form and function of reflexives in Austrian varieties of Bavarian ('Form und Funktion der Reflexiva in österreichischen Varietäten des Bairischen'). Stuttgart: Franz Steiner. • Weiß, Helmut. 1998. Syntax des Bairischen. Tübingen: Niemeyer.

Friday Poster.2

WERNICKE AREA STIMULATION DIFFERENTIALLY AFFECTS ACQUISITION OF NOVEL CONCRETE AND ABSTRACT SEMANTICS

Diana Kurmakaeva, Nadezhda Mkrtychian, Daria Gnedykh, Evgenii Blagoveschenskii, Svetlana Kostromina (St. Petersburg State University) & Yury Shtyrov (Aarhus University, St. Petersburg State University)

diana.s-pb@mail.ru

A body of research suggests different mechanisms for the storage and processing of concrete and abstracts knowledge. There is a behavioural evidence that concrete concepts have an advantage over the abstract ones in verbal processing and learning (Holcomb, Kounios, Anderson, & West, 1999; Strain, Patterson, & Seidenberg, 1995). It is usually explained in the dual-coding framework (Paivio, 1990) as an advantage for linguistic items that can be coded both verbally and referentially (e.g. visually), or by a higher context availability for concrete concepts as opposed to abstract knowledge (Schwanenflugel & Stowe, 1989). Concrete words are processed faster than abstract ones (Kroll & Merves, 1986; Mestres-Missé, Münte, & Rodriguez-Fornells, 2014; Palmer, MacGregor, & Havelka, 2013). Furthermore, studies of semantic memory impairments suggested different neurophysiological mechanisms for storing these two types of semantics (Wang, Conder, Blitzer, & Shinkareva, 2010). However, how exactly concrete and, especially, abstract representations emerge, and what neural mechanisms underpin this process, remains obscure, since most studies focus on investigating the nature of existing presentations rather than their development. Here, we aimed to fill this gap and investigate the acquisition of new concrete and abstract concepts. Furthermore, to assess the causal role of the core language comprehension systems in semantic acquisition, we modulated the activity in Wernicke's area, one of key semantic-processing hubs, by applying cathodal transcranial direct current stimulation (tDCS).

A sample of 48 right-handed monolingual Russian speakers was divided into cathodal tDCS group and control (sham) group. Following 15 minutes of real/sham stimulation applied at CP5 scalp location, both groups underwent a contextual word learning session, in which the meaning of 20 novel words had to be inferred from a sequence of 5 sentences, equalised for length and presented visually word by word. All novel word forms had tri-syllabic 8-letter structure, were matched for various parameters and rotated across conditions/meanings; known words and unlearned word forms served as a control. For novel concrete words, we used rare, unusual or ancient objects (e.g., "medieval device for trapping lice"). Novel abstract concepts were borrowed from other cultures, with a requirement that no equivalent exists in Russian language (e.g. "the awkward feeling of waiting for the green light when everyone crosses on the red"). These were validated as unfamiliar in a separate rating study. After the training, we assessed the acquisition using a free-form definition task, in which the subjects had to (1) match the new word with its meaning and (2) describe the new words/concepts in their own terms. Four experts scored free-form definitions, and the coherence of their ratings was assessed with W-Kendall coefficient. The accuracy was compared between conditions (using Wilcoxon signed-rank test) and groups (using Wilcoxon rank-sum test). In order to address both immediate learning outcomes and consolidation/retention processes, the assessments were done both after the training session (Day 1) and on the following day (Day 2).

Both groups successfully acquired both types of novel concepts following just 5 presentations. Whereas there was a trend for a reduced performance on the Day 2 as compared with Day 1, it diverged between the stimulation conditions. In the sham group, accuracy of semantic matching of abstract words decreased on Day 2. In the cathodal group, however, the accuracy of semantic matching on Day 2 decreased only for concrete but not for abstract words. A comparison between cathodal and sham groups revealed overall better performance of cathodal group in both semantic matching and definition accuracy for the abstract words on Day 2. The results suggest rapid efficient systems for contextual acquisition of concrete and abstract semantics, and a more prominent role of Wernicke's area in consolidating/retaining abstract knowledge.

Supported by RF Government grant contract No.14.W03.31.0010.

Friday Poster.3

AGE DIFFERENCES IN THE USE OF SYNTACTIC AND SEMANTIC ASSOCIATIONS DURING SENTENCE PROCESSING

Caroline Beese (Max Planck Institute for Human Cognitive and Brain Sciences), Markus Werkle-Bergner (Max Planck Institute for Human Development), Ulman Lindenberger (Max Planck Institute for Human Development), Angela D. Friederici (Max Planck Institute for Human Cognitive and Brain Sciences), & Lars Meyer (Max Planck Institute for Human Cognitive and Brain Sciences)

beese@cbs.mpg.de

While sentence processing remains generally well preserved with increasing adult age, difficulties arise when sentence processing taxes verbal working memory. Such difficulties may be related to age differences in the use of syntactic and/or semantic associations to reduce the memory load (Stine-Morrow & Payne 2016). Both syntactic and semantic associations enable the grouping of single words into larger units (Bonhage et al. 2014). Therefore, in this behavioral study, we varied the availability of syntactic and semantic associations. Syntactic associations were made available in sentences, which were contrasted to word lists; semantic associations were made available in meaningful sentences and lists, which were contrasted to pseudoword sentences and lists (Table 1). We evaluated the extent to which older compared to younger adults may differentially use the availability of syntactic or semantic associations to cope with verbal working memory limitations. Eased by the availability of syntactic or semantic associations, participants judged whether the serial order of two words from either sentence or list matched the order in which they were previously encountered. Varying the level of verbal working memory demands, two experiments were conducted. In experiment 1, 8-word stimuli were used; in experiment 2, 11-word stimuli were used, increasing verbal working memory demands. 27 younger (mean age: 26 years) and 26 older adults (mean age: 66 years) participated in experiment 1. Similarly, 26 younger (mean age: 25 years) and 27 older adults (mean age: 64 years) participated in experiment 2. While the effective use of syntactic associations was expected to decrease with age, the use of semantic associations was hypothesized to be enhanced for older compared to younger adults (Stine-Morrow & Payne 2016). The results showed that, only when verbal working memory demands were high (i.e., in experiment 2 when using longer stimuli), the use of syntactic regularities was indeed compromised at old age (experiment 2, sentence structure x age group interaction, $F(1,51) = 9.88, p < .01$), while the benefit of semantic information for sentence processing was comparable across age groups (experiment 2, semantic information x age group interaction, $p > .05$). In light of the reduced use of syntactic associations, our findings may suggest that semantic information processing may become relatively more important for successful sentence processing with advancing adult age, possibly inducing a syntactic-to-semantic-processing strategy shift.

Table 1: Experiment 1, example of stimulus material

		SYNTAX	
		+	-
MEANING	+	der Opa verdarb die Suppe mit dem Salz <i>the granddad ruined the soup with the salt</i>	der dem Suppe mit Opa Salz die verdarb <i>the the soup with granddad salt the ruined</i>
	-	der Apo verword die Junne mit dem Sohr <i>the Apo verword the Junne with the Sohr</i>	der dem Junne mit Apo Sohr die verword <i>the the Junne with Apo Sohr the verword</i>

References

- Stine-Morrow, E. A., & Payne, B. R. (2016). Age Differences in Language Segmentation. *Experimental Aging Research*, 42(1), 83-96.
- Bonhage, C. E., Fiebach, C. J., Bahlmann, J., & Mueller, J. L. (2014). Brain Signature of Working Memory for Sentence Structure: Enriched Encoding and Facilitated Maintenance. *Journal of Cognitive Neuroscience*, 26(8), 1654-1671.

Friday Poster.4

**ASSOCIATION OF SPEECH PERCEPTION AND PRODUCTION IN 2-MONTH-OLDS:
RELATING EVEN-RELATED BRAIN POTENTIAL AND VOCAL REACTIVITY MEASURES**

Gesa Schaadt (University of Leipzig, Max Institute for Human Cognitive and Brain Sciences),
Angela D. Friederici (Max Institute for Human Cognitive and Brain Sciences), Hellmuth Obrig
(University of Leipzig, Max Institute for Human Cognitive and Brain Sciences), & Claudia
Männel (University of Leipzig, Max Institute for Human Cognitive and Brain Sciences)

schaadt@cbs.mpg.de

Perceptual and expressive phonological abilities are key features for success in language development and a functional connection between speech perception and production has been postulated. In line with this assumption, it has been shown that babbling – a form of vocalization – shapes speech processing in 10-month-olds [1]. Precursors of babbling (e.g., imitation of mouth movements, vocalization) already develop around the second month of life [2], but the association of speech perception and production (i.e., vocalization) has not been investigated during this early developmental period.

In the present study, we investigated speech perception and production in 2-month-olds. For speech perception, the Mismatch Response (MMR) was measured in a multi-feature paradigm [3] with four deviant stimulus categories, namely consonant (/ga/), vowel (/bu/), pitch (F0; /ba+), and vowel length changes (/ba:/) that were compared to the standard stimulus /ba/. For speech production, we used the subscale *vocal reactivity* of the parental *Infant Behavior Questionnaire*, defined as the amount of infants' vocalization exhibited in daily activities [4]. Our data ($N=25$) reveal significant positive MMRs for all deviant categories, typically observed in infants at that age. Importantly, we found a negative correlation ($r = -.38$, $p < .03$) between the MMR to vowel changes and vocal reactivity, but no correlation between the MMR to the other deviant stimulus categories and vocal reactivity. Thus, a more negative MMR to vowel changes was associated with infants' higher amount of vocalization. That the MMR to vowel changes, but not to, for example, consonant changes was associated with vocal reactivity, might be explained by findings showing that the perception and production of vowels emerges earlier in development, compared to the perception and production of consonants [5]. Our results suggest that speech perception and production are shaping each other already at an early age. Moreover, the transition from a positive to a negative polarity of the MMR, with negative MMRs indicating more mature responses [6], might be influenced by infants' expressive abilities.

References:

- [1] DePaolis et al. (2013). *Infant Behav Dev*, 36, 642–649.
- [2] Henning et al. (2005). *Infant Behav Dev*, 28, 519–536.
- [3] Näätänen et al. (2004). *Clin Neurophysiol*, 115, 140–144.
- [4] Garstein & Rothbart (2003). *Infant Behav Dev*, 26, 64–86.
- [5] Selby et al. (2000). *Clin Linguist Phon*, 14, 255–265.
- [6] He et al. (2009). *Eur J Neurosci*, 29, 861–867.

ERP INDICES OF ENCODING EFFECTS IN WH-DEPENDENCY PROCESSING

Sergio López-Sancio (UPV/EHU) & Ellen Lau (UMD)
 sergio.lopez-sancio@ehu.eus

Encoding elaborate representations facilitates memory performance in both explicit memory recall and recognition tasks [1]. In sentence comprehension, results from self-paced reading studies using filler-gap dependencies (FGD) found that encoding a syntactically and semantically complex filler leads to faster reading times at the verb [2, 3]. These studies propose that encoding more cues from the filler (a) increases its activation level in memory and (b) makes it more distinct, reducing retrieval similarity-based interference at the verb. The current ERP study investigates how encoding complex fillers impacts online FGD maintenance and filler retrieval at the verb. We also teased apart the contribution of syntactic and semantic complexity. Our results suggest that syntactic complexity plays a central role in facilitating integration of the filler with the verb.

In our experiment, native speakers of English read sentences presented at a rate of 500 ms per word, followed by a yes/no comprehension question. We manipulated the complexity of the *wh*-filler (1), resulting in three *wh*-conditions: a bare *wh*-filler (*who*), a two-word syntactically more complex *wh*-filler (*which person*) and a two-word syntactically and semantically more complex *wh*-filler (*which-NOUN*). A *that*-clause with no FGD was used as the baseline. We focused on two ERP components elicited in FGD processing: a late positivity (P600) at the verb associated with integration cost [4], and a sustained anterior negativity (SAN) throughout the dependency linked to memory maintenance cost [5]. Based on previous work, more complex fillers were predicted to: (a) reduce integration difficulty at the subcategorizing verb, eliciting a smaller P600; and (b) incur in higher memory load during the maintenance stage, eliciting a larger SAN than simple fillers.

At the verb (n=37), *wh*-conditions elicited a P600 with respect to the baseline, which resulted in a significant main effect of condition (500-700 ms: $F_{(3,108)}=3.34, p=0.02$). Its amplitude was modulated by syntactic complexity: syntactically complex fillers (WHICH-N and WHICH-PERSON) produced a smaller effect than simple fillers (WHO). Pairwise comparisons confirmed a main effect between WHO and THAT (500-700 ms: $F_{(1,36)}=9.48, p=0.004$; 700-900 ms: $F_{(1,36)}=5.03, p=0.03$) and between WHICH-N and THAT (500-700 ms: $F_{(1,36)}=4.24, p=0.05$). Crucially, no difference was found between WHICH-N and WHICH-PERSON ($F_{(1,36)} < 1$). Unfortunately, we could not replicate the SAN found in previous studies.

Our results support the idea that encoding complex representations in memory facilitates retrieval in sentence comprehension. Additionally, complex fillers are longer, and the parser has more time to encode its representation in memory, which may facilitate its subsequent retrieval. The role of semantic cues of the filler and the underlying nature of SAN effects in previous work need to be further addressed.

(1) The manager knew...

- ... **which waiter** the new owner of the coffee shop would fire after the scandal. WHICH-N
- ... **which person** the new owner of the coffee shop would fire after the scandal. WHICH-PERSON
- ... **who** the new owner of the coffee shop would fire after the scandal. WHO
- ... **that** the new owner of the coffee shop would fire the waiter after the scandal. THAT

Fig. 1. Grand average ERP response at the verb.

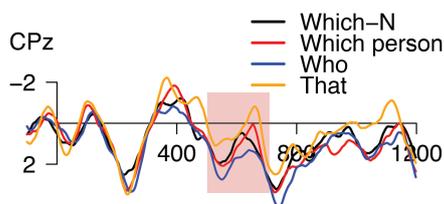
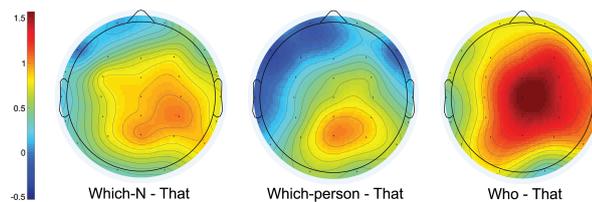


Fig. 2. Topographic maps based on average voltage differences between *wh*-filler and control (THAT) conditions between 500-700 ms.



[1] Bradshaw, G. L., & Anderson, J. R. (1982). *J Verb Learn Verb Be*, 21(2), 165-174. | [2] Hofmeister, P. (2011). *Lang Cognitive Proc*, 26(3), 376-405. | [3] Hofmeister, P., & Vasishth, S. (2014). *Front Psychol*, 5, 1237. | [4] Kaan, E. et al. (2000). *Lang Cognitive Proc*, 15(2), 159-201. | [5] King, J. W., & Kutas, M. (1995). *J Cognitive Neurosci*, 7(3), 376-395.

Friday Poster.6

INDIVIDUAL (NON-)VARIABILITY OF PROSODIC CUE PRODUCTION IN COORDINATE STRUCTURES

Clara Huttenlauch, Carola de Beer, Sandra Hanne, Isabell Wartenburger (University of Potsdam)

huttenlauch@uni-potsdam.de

This work was funded by the Deutsche Forschungsgemeinschaft (DFG), Collaborative Research Centre SFB 1287, Project B01.

Background: In language production, structurally ambiguous coordinates like, for instance name sequences, can be disambiguated or grouped by using a variety of prosodic cues [1]. The type of prosodic cue being produced and whether cues occur in isolation or in combination is a matter of variability induced, for instance, by the speakers themselves (see [2]) and by contextual factors (e.g., the speaker's interlocutor [3]). Previous data show that prosodic cues are increased in utterances directed to speaker-dissimilar interlocutors (e.g. [3]). Yet, the effect of different interlocutors on prosodic cue production in structurally ambiguous coordinates is still unclear.

Aim: The current study aims to investigate intra- and inter-individual variability in prosodic cues used for grouping of coordinate name structures, such as (1) and (2). Specifically, we address the question of whether and how speakers adapt their use of prosodic cues when addressing interlocutors who differ from them in age.

- (1) Name1 and Name2 and Name3
- (2) (Name1 and Name2) and Name3

Method: Stimuli consisted of a sequence of three disyllabic, trochaic German names that were coordinated by *und* ("and"). Six different name sequences were constructed, each in two conditions grouping either the first two names together (2) or without internal grouping (1). Prosodic cue production in coordinates was investigated by means of a task similar to a referential communication task with three contexts each with a female interlocutor: (a) a young adult, (b) a child, (c) an elderly person. Contexts were presented in blocks. At the beginning of each block, the interlocutor presented themselves in a short video clip. Stimuli were presented visually on screen with brackets to indicate the targeted prosodic grouping. Each visual stimulus was preceded by an auditorily presented context question (*Wer kommt?* "Who is coming?") produced by the context-specific interlocutor. Participants (n=16 native German adults) were asked to produce the name sequences in a way that would allow the interlocutor to group the names in the intended way as accurate and fast as possible. Recordings took place in a sound-attenuated booth.

Results: Preliminary results of 11 participants' productions show that speakers mark the grouping of coordinates like (2) using an increased pitch range of the rise on name2, a (longer) pause at the end of the first group (i.e., at the position of the bracket), and by lengthening of the final vowel of name2 compared to the condition without internal grouping (1). On an individual level, however, speakers vary in whether and how they modulate the strength of prosodic cues across contexts: some speakers use a higher pitch range on name2 to indicate the grouping when addressing the child and/or the elderly adult compared to addressing the young adult while others do not show such a pattern. The discussion will also consider additional variables (e.g. speaking-rate) and cue trading relations. In addition, we will focus on non-varying patterns in order to get a better understanding of the limits of individual variability, that is those patterns which are indispensable for a successful communication.

[1] Kentner, G., & Féry, C. 2013. A new approach to prosodic grouping. *The Linguistic Review*, 30(2), 277–311.

[2] Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, 61, 71–92.

[3] Biersack, S., Kempe, V., & Knapton, L. 2005. Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. *Proceedings of Eurospeech*, (Lisbon), 2401–2404.

Friday Poster.7

GERMAN DEMONSTRATIVES ARE SENSITIVE TO PERSPECTIVE-TAKING

Stefan Hinterwimmer (University of Cologne) & Umesh Patil (University of Cologne)
stefan.hinterwimmer@uni-koeln.de

German demonstrative pronouns (DPros) have been shown to avoid maximally prominent referents as their antecedent when prominence is rendered through topicality or agentivity. Hinterwimmer and Bosch (2016, 2017) propose that when the proposition denoted by a sentence containing a DPro is evaluated from the narrator's perspective even the topical referent becomes available as an antecedent of the DPro. We report an acceptability rating study and an eye-tracking study in the visual-world paradigm that test their proposal about the influence of perspective-taking on DPros' antecedent preference.

In expt.1 (n=85) we conducted an acceptability rating study with sentences as in (1). The first sentence, same across six conditions, established an individual referred to by a proper name as topic. The second sentence in *Narr-judg* conditions (a-b) clearly expressed an evaluation of the topical referent by the narrator, as indicated by the semantic content and a switch from past tense to present. In *Top-judg* conditions (c-d), the second sentence clearly expressed a thought of the topical referent in Free Indirect Discourse mode, as indicated by the semantic content, an interjection and a deictic expression that could only be interpreted with respect to the topical referent's perspective. Finally, the *Neut* conditions (e-f) were neutral continuations of the previous sentence. These three continuations types referred to the topic with either a personal pronoun (PPro) or a DPro giving rise to six conditions. Effectively, in *Narr-judg* conditions, 'Emil' is prominent in terms of topicality, while the narrator is prominent in terms of perspective-taking. In contrast, in *Top-judg* conditions, 'Emil' is prominent both in terms of topicality and perspective-taking, and in *Neut*, 'Emil' is prominent because in the absence of an overt perspective-taker, topicality is the only factor influencing prominence. The participants were asked to judge whether the story openings such as in (1) sound native or not. The rating showed significant interaction between the *Narr-judg* and *Neut* conditions, which was driven by two DPro conditions such that the DPro sentences from *Narr-judg* type were rated more acceptable than the DPro sentences from the *Neut* type. There was no interaction between the *Top-judg* and *Neut* conditions. The PPro sentences were rated equally in all three sentence types and higher than the DPro sentences.

Expt. 2 (n=48) was an eye-tracking study in the visual-world paradigm with discourses as in (2). Sentence 1 and 2 established a male individual as topic. In sentence 3, it was either referred to by a personal pronoun providing a neutral continuation, or by an epithet providing an evaluation of the topical referent from narrator's perspective. The discourse also introduced another human masculine referent and two non-human referents as distractors. The DPro, occurred in the complement clause of the third sentence. The display showed these four referents together with an unmentioned distractor object. Gaze frequencies showed that, after the onset of the DPro, when the topical referent was referred to by an epithet, it was significantly more preferred than when it was referred to by a personal pronoun. We take the results from the two experiments as evidence that perspective-taking increases the prominence status of the narrator and hence even the topical referent becomes available as an antecedent of the DPro.

- (1) Sentence 1: When Emil wanted to drive a nail into the wall, he hit his thumb.
Sentence 2: [a-b: *Narr-judg*] **He / DPro** really has no craftsmanship at all.
[c-d: *Top-judg*] Ouch, **He / DPro** didn't need that at all today.
[e-f: *Neut*] **He / DPro** called a friend first to ask for help.
- (2) Sentences 1 and 2: Good news. The policeman has just parked the motorcycle and talks to the photographer.
Sentence 3: **He / the_nice_sergeant** has just told the photographer who is here because of the kangaroos that **DPro** has won the lottery.

Friday Poster.8

CAN GERMAN DEMONSTRATIVE PRONOUNS REALLY BE BOUND?

Umesh Patil (University of Cologne) & Stefan Hinterwimmer (University of Cologne)

upatil@uni-koeln.de

Hinterwimmer and Brocher (2016; to appear) have provided empirical evidence for the claim made by Hinterwimmer (2015) that German demonstrative pronouns of the *der/die/das* paradigm (DPros) can be bound, which is contrary to the claim in Wiltschko (1998). Using self-paced reading and acceptability rating studies, they show that DPros can be bound not only by clause mate referential expressions, but also by universally quantified DPs if the respective binders are not grammatical subjects but rather direct or indirect objects. Consequently, they concluded that DPros receive bound variable-like interpretations in the same way as personal pronouns (PPros), with their anti-subject bias being a result of their status as the marked pronoun variant (see Patel-Grosz and Grosz 2017 for discussion). However, it is well-known that universal quantifiers can not only bind PPros in standard binding configurations requiring c-command, but also allow for so-called *telescoping* across sentence boundaries (for example, in (1) below the DP headed by the universal quantifier 'Each' doesn't c-command any pronoun in the second sentence); in contrast, negative quantifiers only allow standard binding and no telescoping (for example, (1) below with 'No' alternative).

(1) Each / #No degree candidate walked up to the stage. He took his diploma from the Dean and returned to his seat. (Roberts 1989)

Hence, in order to conclude that DPros give rise to bound variable-like interpretations in the same way as PPros, one would have to show that there is no relevant difference between DPros and PPros with respect to reading times in sentences where the only available binder is a negative quantifier: They should be read equally fast, and just as fast as in parallel sentences with universally quantified DPs. On the other hand, if DPros are read slower than PPros in sentences with negative quantifiers, while they don't differ with universal quantifiers, this would provide evidence that DPros are not bound in the same way as personal pronouns.

In order to test these predictions we conducted a self-paced reading study (n=46, 24 items, 48 fillers) in which participants were presented with target sentences such as in (2a-d). These target sentences were preceded by an introductory sentence, which was same across four conditions, to set up a sound context for the use of these two types of quantifier. In each case, only the quantificational object DP (headed by quantifiers *jeden* or *keinen*) was in principle available as binder, due to matching gender features.

We found no significant difference in reading times between (2a), (2b) and (2c) in the target or the spillover region. But there was a slowdown in the spillover region for (2d), where the only available binder for the DPro was a negative quantifier. This effect was small (~20ms) but significant ($t > 2.3$). Our results therefore suggest that DPros may not be bound in the standard way, but rather by a more constrained mechanism such as telescoping.

(2) **Prelude:** In der Grundschule, in der die Lehrerin arbeitete, wurde auch eine Hausaufgabenbetreuung angeboten.

In the elementary school in which the teacher_FEM worked, after-school homework supervision was offered.

[a-b] Die Lehrerin lobte **jeden** Jungen, der fleißig war, vor **seiner / dessen** Klasse, weil die anderen Kinder sich daran ein Beispiel nehmen konnten.

[c-d] Die Lehrerin lobte **keinen** Jungen, der fleißig war, vor **seiner / dessen** Klasse, weil die anderen Kinder sich gleichbehandelt fühlen sollten.

*The teacher_FEM praised **every/no** boy who was diligent in front of **his / his_{DPro}** class, because the other children should feel treated equal.*

Friday Poster.9

REFERENCE TO QUANTIFIED EXPRESSIONS IN SWEDISH: AN ERP STUDY

Fredrik Heinat (Linnæus University) & Eva Klingvall (Lund University)

fredrik.heinat@lnu.se

We present the results from an Event Related Potentials (ERP) study on the processing of anaphoric reference to quantified expressions (QEs) in Swedish. QEs pick out proportions of possible members of some set for which a property holds. In (1a) and (1b), for example, some or few members of the set of students attended the lecture.

- (1) a. Some students attended the lecture.
b. Few students attended the lecture.
- (2) a. They found it very interesting.
b. They stayed at home instead.

Some and *few* differ in polarity: *some* is positive (upward entailing) while *few* is negative (downward entailing) (Peters and Westerståhl, 2006) and this is of importance when referring back to the QE using anaphoric expressions. The sentence in (1a) is naturally followed by (2a), which is about the students attending the lecture (the reference set, REFSET). The sentence in (1b), in contrast, is naturally followed by (2b), which is about the students *not* attending the lecture (the complement set, COMPSET) (e.g. Moxey and Sanford, 1987). While (1b) can in fact be followed either by (2a) or (2b), (1a), cannot be followed by (2b).

Filik et al. (2011) is one of few studies of anaphoric reference to QEs in English using online measures (ERP). They report results for positive and negative QEs separately. Each type of QE shows REFSET and COMPSET effects, as described above, on the disambiguating word. A larger N400 for COMPSET vs. REFSET continuations for positive QEs, and the opposite for negative QEs. However, they do not report any results for the contrast between positive and negative QEs in the COMPSET condition. Since this is a very important condition and since it is known that QEs differ across languages (Nouwen, 2010; Tsai et al., 2014), we investigated this issue for Swedish.

160 experimental items of four sentences each were manipulated along two dimensions: polarity (positive vs negative quantifier, *några* vs *få* in (3)), and set (REFSET vs COMPSET targeting disambiguating adjective, *duktiga* vs *dåliga* in (3)). The quantifiers included were: *några* ('some'), *få* ('few'), *många* ('many'), *inte många* ('not many'), *alla* ('all'), *inga* ('no'), *nästan alla* ('almost all'), *inte alla* ('not all').

- (3) Några/Få studenter skrev bra på tentan igår och att de_{CW} var så duktiga/dåliga_{CW} förbryllade professorn.
some/few students wrote well on the exam yesterday and that they were so good/bad confused the professor.

There were four lists with 40 sentences from each condition. Each participant (29 in total, results reported below based on the first 13) only saw one sentence from each item, but saw all types of manipulation. In total, each participant read 400 sentences (160 test items, 240 fillers).

Unlike Filik et al. (2011) we found that positive QEs showed a pronounced positivity over the central region (FCZ, CZ, CPZ, PZ) in the COMPSET condition relative to negative QEs, in the P600 time span (500–800 ms) after the onset of the critical word (the disambiguating adjective, 'bad'). A linear mixed effects model analysis (LmerTest) showed a highly significant main effect of polarity in the central region and the P600 time span above. We interpret this to mean that for positive QEs, a new discourse referent needs to be introduced following COMPSET reference, while for negative QEs this discourse referent is already available (Burkhardt, 2007).

ANTECEDENT RETRIEVAL FOR REFERENTIAL AND DONKEY PRONOUNS

Dave Kush (NTNU Trondheim) & Ragnhild Eik (NTNU Trondheim)
dave.kush@ntnu.no

We investigated whether the processing of distinct anaphoric dependencies employs different mechanisms for antecedent retrieval. [1] proposed that the parser uses the same retrieval mechanism for all grammatical antecedents, regardless of their type (*referential* or *quantificational*), their position, or the dependency that they ultimately establish with a pronoun. On the other hand, [2] suggested that at least some quantificational NPs (QPs) – in particular those that scope over, but do not c-command, a pronoun (NoCC-QPs) – are accessed using a mechanism distinct from those used for referential NP or c-commanding QP antecedents. This hypothesis was based on their finding that sensitivity to gender mismatch manipulations seemed to vary as a function of antecedent type: Manipulating gender-match between a pronoun and a referential NP/c-commanding QP antecedent resulted in reliable gender-mismatch effects (GMMEs [3]), but gender-mismatch effects were absent when match between a pronoun and a NoCC-QP antecedent was manipulated. [2] hypothesized that relations between pronouns and NoCC-QPs might be analyzed as *D-Type* anaphora, and further conjectured that other instances of D-type anaphora would be processed similarly. This predicts that other D-type anaphora should display the same insensitivity to gender-match manipulations as NoCC-QPs in [2]. We tested this prediction by comparing the processing of *donkey pronouns* (often given a D-type analysis: e.g [4]), with standard referential pronouns. **Design:** In two SPR experiments (both Ns=32) native-speakers of Norwegian read sentences in which a critical pronoun was the object of the matrix verb. Sentences also contained an RC-internal indefinite *antecedent NP* that did not c-command the pronoun. Test sentences followed a 2×2 design: GENDERMATCH varied whether the critical pronoun matched the antecedent NP in grammatical gender. RELATION varied whether the indefinite NP could establish a *Coreferential* or *Donkey* dependency with the pronoun. In the *Coreferential* conditions the main subject was definite (*the dad*). In the *Donkey* conditions the subject was quantificational (*every dad*). In **Experiment 1**, pronouns (*it*_[NEUT]/*it*_[MASC/FEM]) and antecedent NPs were inanimate. In **Experiment 2**, pronouns and antecedent NPs were animate. An English equivalent of an experimental item from Experiment 2 is given in (1). **Predictions:** We expected a GMME in the two *Referential* conditions: RTs should be shorter after a matching pronoun than after a mismatching pronoun. The presence or absence of a similar GMME was of interest in the donkey items. If Donkey anaphora are processed similarly to the NoCC constructions from Moulton & Han (2017), there should be no GMME in the *Donkey* sentences, yielding a GENDERMATCH × RELATION interaction. A GMME in *Donkey* sentences would provide evidence that D-type anaphors do not access their antecedents using a different route from that used for coreference. **Analysis & Results:** All RTs were log-transformed and residualized before analysis using maximal LMEMs [5]. In Experiment 1 there was a marginally significant GENDERMATCH×RELATIONTYPE interaction at the pronoun ($t = 1.90$) indicating a larger GMME in the referential items ($t = 2.00$) than in the donkey items ($t < 1$), followed by a significant main effect of GENDERMATCH at the post-pronoun region ($t = 2.78$). In Experiment 2 we failed to observe the interaction at the pronoun: Immediate GMMEs were observed in both the *Coreferential* and the *Donkey* sentence pairs at the pronoun and post-pronoun regions ($t_s = 2.23, 5.70$, respectively). The results suggest that antecedents for D-type and referential pronouns are retrieved in the same manner. They also suggest that the absence of a GMME in [2] cannot be tied to a D-type analysis *per se*.

(1) **REFERENTIAL**

The dad [who has a daughter in the soccer league] has to drive her/*him to games ...

DONKEY

Every dad [who has a daughter in the soccer league] has to drive her/*him to games ...

[1] Kush, D., Lidz, J. & Phillips, C. (2015). *JML*; [2] Moulton, K. & Han, C-H. (in press). *Language*; [3] van Gompel, R & Liversedge, S. (2003). *JEP:LMC*; [4] Elbourne, P. (2005). *Situations & Individuals*; [5] Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J. (2013). *JML*.

CHILDREN WITH SLI CAN USE NUMBER AGREEMENT IN OBJECT-INITIAL SENTENCES TO OVERCOME THEIR DIFFICULTIES WITH CASE MARKING

Maja Stegenwallner-Schütz (University of Potsdam) & Flavia Adani (University of Potsdam)
stegenwa@uni-potsdam.de

Object-initial declarative sentences (OVS) in German are poorly understood by typically developing (TD) preschoolers, when case marking essentially indicates correct thematic role assignment (OVS-C, cf. Table, Schipke et al., 2012). However, number agreement can enhance the accuracy of OVS in typically developing children, when OVS sentences display number and case marking. The structural intervention account (e.g., Belletti et al., 2012) predicts a facilitation when the different number marking on the subject and the object NPs assist the integration of subject-verb agreement relation. In this study, we investigated the effect of case and number marking on OVS comprehension accuracy in a group of children who are known to have difficulties with morphological features, namely children with Specific Language Impairment (SLI). Hence, they are predicted to perform poorly on OVS-C sentences. We test whether number agreement with plural or singular subjects enhances comprehension accuracy of case-marked OVS sentences, as it has been shown for TD children. Monolingual German-speaking children with SLI (N=27; mean age: 6 years, 9 months) and their age-matched TD controls participated in an auditory sentence-picture matching task. Test sentences were OVS sentences, displaying either case only (OVS-C) or case and number marking. As for the latter, the number of the subject (and the agreeing verb) differed from the object: It could be either singular or plural. The table reports an overview of the conditions and associated accuracy scores for each group of participants:

Condition		SLI	TD
OVS-C	e.g., Dem Kind winkt der Mann <i>the.SG.DAT child waves.SG the.SG.NOM man</i> 'The man waves to the child'	37%	78%
OVS-C+sg/pl	e.g., Der Oma winken die Polizisten <i>the.SG.DAT grandma wave.PL the policemen</i> 'The policemen wave to the grandma'	56%	90%
OVS-C+pl/sg	e.g., Den Mädchen gratuliert die Mutter <i>the.PL.DAT girls congratulates.SG the mother</i> 'The mother congratulates the girls'	39%	81%

Note. DAT: dative case; NOM: nominative case; SG: singular; PL: plural

Children with SLI performed overall less accurately than the TD group (GLMM, $p < 0.001$). Both groups similarly showed greater accuracy of OVS-C+sg/pl than OVS-C sentences ($p < 0.05$), but such difference was not evident between OVS-C+pl/sg and OVS-C sentences ($p = 0.78$). The study shows that plural features can facilitate OVS comprehension in children with SLI, similarly to TD children. This is in line with the structural intervention account, which predicts enhanced performance driven by features that enter the subject-verb agreement relation, such as number in German. However, a number facilitation was only attested in the OVS-C+sg/pl condition, i.e., sentences with a singular object and plural subject. This number asymmetry may be driven by the parser's tendency to overlook plural features on the object NP, as in OVS-C+pl/sg sentences (cf. Patson & Husband, 2016, for number misinterpretations). These results indicate that German-speaking children (with and without SLI) can use number features and in particular, the plural marker of the subject and verb to enhance comprehension.

Selected references:

- Belletti, A., Friedmann, N., Brunato, D., & Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122, 1053–1069.
- Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, 15, 762–774.

REFERENTIAL OVERSPECIFICATION: FROM EGOCENTRICITY TO RATIONALITY

Elli Tourtouri, Les Sikos & Matthew Crocker (Saarland University)
elli@coli.uni-saarland.de

Contra the Gricean Maxim of Quantity (Grice 1975), recent studies indicate that speakers' referential expressions often contain more information than strictly necessary for target identification (e.g., Koolen et al. 2011). Such overspecifications may be driven by egocentric concerns (minimise speaker's cognitive effort, cf. Keysar et al. 1998) or audience design (maximise listener comprehension). That is, egocentric speakers may overspecify targets regardless of attribute informativity, while rational speakers should exploit redundant attributes primarily to reduce referential entropy (i.e., number of distractors) more efficiently, thus lowering cognitive processing effort for the comprehender (Tourtouri et al. 2017).

We investigated the use of overspecification in a referential communication task. Speakers and Listeners saw displays of identical objects that could differ in their spatial arrangement. Following a 2 second preview, one object was identified as the target on the Speaker's display only. The Speaker's task was to ask whether the target was on the left or right side of the Listener's display. Using a 2x3 design, we manipulated the Necessary Adjective (i.e., whether colour or pattern was required to specify the target) and the Entropy Reduction Advantage of each adjective (i.e., whether colour or pattern reduced entropy more, or equally). The dependent measure was the overspecification rate per condition. We predicted that if speakers are egocentric they should produce overspecifications systematically. In contrast, if speakers are rational, more overspecifications should be found when the redundant adjective reduced referential entropy more than the necessary adjective (Fig.1b), relative to conditions in which the necessary adjective reduced entropy more (Fig. 1a) or when both adjectives reduced entropy equally (Fig.1c).

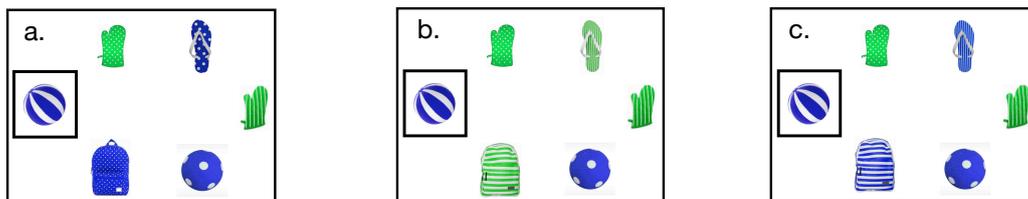


Figure 1. Sample stimuli for the Pattern Necessary conditions (i.e., "striped" is required to identify the target). Although a colour adjective (e.g., "blue") is strictly redundant, it differs in the degree to which it reduces referential entropy across Entropy Reduction Advantage conditions: in (a) colour reduces entropy from 6 to 4 referents, and in (b) from 6 to 2 referents. In (c) pattern and colour reduce entropy equally, from 6 to 4 referents. Black box indicates target object (seen by Speaker only). Colour Necessary conditions contained complimentary manipulations.

Based on their use of overspecification, we identified three groups of speakers (Fig.2). One group (N=16, Fig.2A) fully specified the target more than 80% of the time, thus minimising speaker effort, matching the egocentricity predictions. A second group (N=10, Fig.2B) used colour redundantly more than 80% of the time. This strategy is low-cost for both speakers and listeners, as colour is a salient property that can guide visual search and ease target identification. The remainder of speakers (N=16, Fig.2C) used a redundant colour or pattern adjective more often when it reduced referential entropy more than the necessary adjective, matching the rational account and suggesting that referential overspecification may be a rational strategy. These results highlight individual differences in language production, and indicate that speakers' choice to use redundant modifiers may be driven by different factors that can vary from egocentricity to rationality.

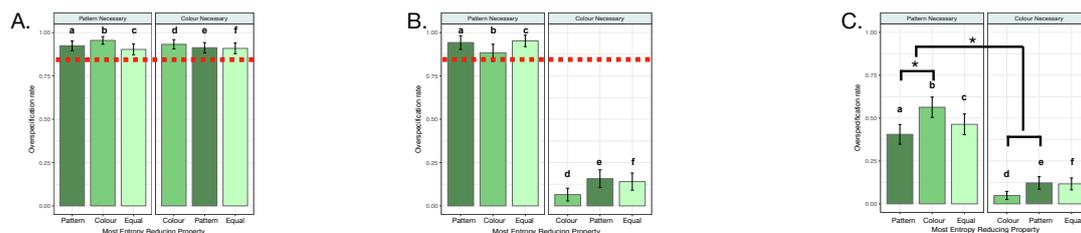


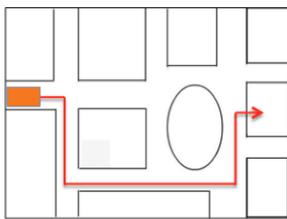
Figure 2. Overspecification rate per condition for each speaker group. Group A fully specified the target more than 80% of the time (red line). Group B overspecified for colour (d-f) more than 80% of the time. Group C overspecified pattern more often when it reduced entropy more than the necessary colour adjective (b), and colour more often when it was more entropy-reducing than the necessary pattern adjective (e).

LANGUAGE-INDUCED EFFECTS ON EVENT MEMORY

Yaqi Wang (University of York), Gareth Gaskell (University of York), Silvia Gennari (University of York)

yw1308@york.ac.uk

How do we represent the duration of past events that we have conceptualized through language? Prior research on memory and language suggests that describing stimuli during encoding may lead to distorted memories consistent with the linguistic concepts (Feist & Gentner, 2007, Lupyan, 2008). Interactive encoding accounts, for example, argue that top-down linguistic features may augment bottom-up stimulus features. Yet little is known about how we encode dynamic events. Here we examine whether language-induced event encoding shapes event duration recollection, and specifically, whether language modulates event encoding, retrieval or consolidation processes during sleep, since sleep is known to integrate conceptual and episodic features in memory (Walker & Stickgold, 2010).



We constructed 21 cartoon-like animations varying duration (3 to 9 sec). Each animation was paired with two phrases implying either fast or slow motion, which fitted the animation equally well (see figure: arrow indicates the rectangle's motion path).

SLOW PHRASE: *grandma taking the bus to the hospital*

FAST PHRASE: *an ambulance taking someone to the hospital*

In five experiments, participants first studied all 21 animations, each paired with one of the phrases. Later they mentally replayed each animation as accurately as they could when prompted with a cue. The duration of this replay was measured using a button press to mark the beginning and end.

We manipulated the number of study viewings (one vs. three), the type of cue used to prompt mental replays (a frame vs. a phrase), and the post-study interval (12 waking hours vs. 12 sleep hours). We reason that if language modulates how the animation is first encoded, as suggested by interactive-encoding accounts, mental replays should be longer for animations described by **slow** phrases than for those described by **fast** phrases, irrespective of cue type or time passed after study. Alternatively, language may influence memory only when phrases prompt retrieval, or after consolidation during sleep.

Experiments 1 and 2 used a video frame to prompt mental replay soon after studying the stimuli once or three times in random order. Despite repeated study, there was no effect of language in reproduced duration, though more study improved duration accuracy. This suggests that conceptual features of the phrase and episodic event details were not combined. Verbal and visual information may have been kept separate in memory.

Experiments 3 and 4 also varied the amount of study but used phrases instead of video frames to prompt mental replay. In both experiments, animations described by **slow** phrases were replayed for longer than those described by **fast** phrases (mixed-effects models: $\chi^2(1)=8.05$, $p=0.005$; $\chi^2(1)=10.03$, $p=0.001$). Accuracy also improved with more study. As in Experiments 1 and 2, replay duration was associated with stimulus duration, suggesting that episodic features were merged with the phrases' conceptual features, making replays slightly longer or shorter but still consistent with the stimuli. Therefore, language shaped memory representations at retrieval when events were accessed through the associated phrase.

Experiment 5 again used video frames to prompt replay but tested participants after 12 hours including sleep (overnight) or wake (same day). A language effect was not observed for the same-day group, but was found after sleep ($\chi^2(1)=4.02$, $p=0.043$). This suggests that even in the absence of a linguistic cue to recollection, the conceptual features of the phrases were integrated with episodic event memories during sleep.

Overall, these results suggest that language modulates event duration recollection only when language mediates retrieval (phrase cue) or after sleep-dependent consolidation if visual cues are used. These findings contradict interactive encoding accounts but are consistent with con-

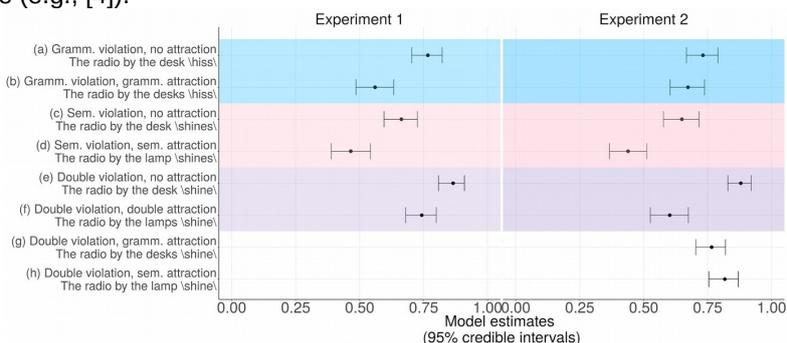
SEMANTIC ATTRACTION IN SENTENCE PROCESSING

Anna Laurinavichyute (University of Potsdam; Higher School of Economics, Moscow) & Titus von der Malsburg (University of Potsdam)
anna.laurinavichyute@uni-potsdam.de

Cunnings and Sturt [1] demonstrated that sentences with implausible subject-verb combinations (*The letter ... was shattered.*) were read faster in the presence of an intervening noun matching the semantics of the verb (*the cup*). This effect resembles agreement attraction where the verb erroneously agrees morphosyntactically with an intervening non-subject noun. This is surprising given that agreement attraction is typically assumed to be caused by derailed morphosyntactic processing [2]. If we find that “attraction” effects exist also in other domains, this would suggest that the mechanisms underlying attraction effects might be of a more general nature than is usually believed.

We employed an experimental paradigm that has previously been used to study agreement attraction in sentence production. In two single-trial online experiments, participants were asked to memorize a verb, press a button to see a sentence fragment, and then to decide as quickly as possible whether the verb is a possible continuation of that sentence fragment (see Fig. 1). We tested 25 item sets in which the verb could mismatch the subject's grammatical number and/or meaning (type of *violation*: grammatical, semantic, double). Also, the verb could match or mismatch the attractor in number and/or meaning (type of *attraction*: none, grammatical, semantic, double). We tested classic agreement attraction (more errors in the presence of a number-matching attractor, cond. *b* vs. *a*), its semantic analog (more errors in the presence of a semantically matching attractor, *d* vs. *c*), and double attraction (errors due to an attractor matching the verb's number and semantics, *f* vs. *e*). In Exp. 1 (N=1072), we replicated the classic agreement attraction effect (Est.=-1.02, CrI:[-1.66, -0.39]), and found a similarly large effect of semantic attraction (Est.=-1.01, CrI:[-1.74, -0.35]). Double attraction (Est.=-0.64, CrI:[-1.19, -0.1]) was smaller than the sum of the two previous effects, potentially due to the easier-to-spot double violation, so conditions *g* and *h* were used as additional baselines in Exp. 2. In Exp. 2 (N=1426), the semantic (Est.=-1.12, CrI:[-1.98, -0.27]), but not the agreement attraction effect was replicated, and the effect of double attraction was as big as the sum of the semantic and grammatical attraction effects in double violation cases (*g* vs. *e*, *h* vs. *e*).

Our results suggest that semantic features can elicit effects similar to classic agreement attraction even when the agreement is completely intact (cf. [3]). The mechanisms underlying agreement attraction may therefore not be limited to the morphosyntactic domain but of a more general nature (e.g., [4]).



[1] Cunnings, I., Sturt, P. (2017) Retrieval interference and sentence interpretation. CUNY 2017. [2] Eberhard, K.M., Cutting, J.C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. Psychol Rev, 112(3), 531. [3] Thornton R., MacDonald M.C.(2003). Plausibility and grammatical agreement, JML 48(4). [4] Lewis R., Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. Cog sci 29.3: 375-419.

Friday Poster.15

KNOWLEDGE OF ACADEMIC WORDS PREDICTED BY A COMBINATION OF TESTS, DESPITE COGNATE INFLATION

Breno Silva (University of Warsaw), Agnieszka Otwinowska (University of Warsaw)

For experimental purposes, proficiency in English is often assessed by the LexTALE (Lemhöfer & Broersma, 2012). However, it focuses on general vocabulary knowledge, not on English academic words. Academic vocabulary is different from general vocabulary in that it is difficult to learn incidentally (Vidal, 2010; Knoch et al., 2015) because it is typically abstract and morphologically complex (Corson, 1997). Assessing the knowledge of academic words for educational and experimental purposes (e.g. research on reading with advanced learners) is highly problematic, too. This is because word lists such as Coxhead's (2000) Academic Word List (AWL), created to facilitate instruction and assessment, include a high proportion of cognates, which are more easily recognised than non-cognates, leading to a mechanism called "cognate inflation", i.e. an overestimation of participants' results due to the presence of cognates. Single tests do not suffice to overcome the problem because they are too prone to cognate inflation; also, this inflation affects various L1s differently, rendering test scores distorted and incomparable among learners of different nationalities. Thus, new, fast ways of academic vocabulary assessment are needed, less sensitive to cognate inflation effects. To this end, we investigated whether a combination of results from a recognition and a recall vocabulary test could be used.

We assessed the receptive vocabulary knowledge of 106 Polish learners of English (B2/C1 level) via two tests: 1. a computerized version of the Vocabulary Size Test (VST; Nation & Beglar, 2007) learners' lexical size (i.e., recognition of up to 14,000 word families), 2. a tailor-made Yes/No Academic Vocabulary Test (AVT), derived from a corpus of applied linguistics texts (167,634 tokens) and similar in format to LexTALE (Lemhöfer & Broersma, 2012). The AVT comprised 600 lemmatized items (300 noncognates, 105 Polish-English cognates, 195 plausible nonwords) corresponding to 308 AWL families selected based on Coxhead's (2000) criteria and using software available at Lextutor (Cobb, 2018). The items, matched for frequency (Brysbaert & New, 2009), concreteness (Brysbaert et al., 2014), length and parts of speech, were divided into 3 equivalent test versions. Effectively, each participant took the VST (140 items) and one version of the AVT (200 items: 100 non-cognates, 35 cognates, 65 nonwords). The order of testing was counterbalanced, and each version of the AVT was distributed to a similar proportion of learners.

We first checked for cognate inflation in both tests. In both the VST and AVT, cognates were known significantly better than non-cognates, which is a sign of cognate inflation. Next, we asked whether the VST score could predict participants' academic vocabulary knowledge, and if so, whether there was a threshold on the VST that differentiates participants with regards to their knowledge of academic words. Since a regression analysis revealed that the VST predicted students' performance on the AVT (adjusted $R^2=.385$), we ran cluster analyses (hierarchical and K-means) combining both tests' scores to find the VST threshold that best predicted successful performance in the AVT. The results showed that 89.65% of learners who scored at or above the 9900 VST threshold also mastered receptive academic vocabulary (operationalized by the AVT score of 88.89% or higher; for discussion see Schmitt et al., 2001). We argue, therefore, that since a threshold has been found in the VST, this test can now be reliably used to predict English academic vocabulary knowledge, and hence is a useful tool for assessment purposes in the Polish context. We also argue that assessing academic word knowledge should follow similar statistical procedures for each language (or at minimum for each context) separately due to differences in the proportion of cognates across languages.

ELECTROPHYSIOLOGICAL CORRELATES OF IMPLICIT AND EXPLICIT ACQUISITION OF NOVEL WORDS

Olga Shcherbakova (St. Petersburg State University), Alexander Kirsanov (St. Petersburg State University), Elizaveta Nikiforova (St. Petersburg State University), Margarita Filippova (St. Petersburg State University), Evgeny Blagoveshchensky (St. Petersburg State University), Yury Shtyrov (St. Petersburg State University, Aarhus University)
o.shcherbakova@spbu.ru

Successful acquisition of new words plays an essential role in human communication; however, underlying mechanisms of this crucial skill are poorly understood. A range of recent results suggests that learning strategies applied during word acquisition can be either implicit (the learner is not directly aware of what is to be learned, e.g. the so-called “fast mapping” through context) or explicit (e.g., direct instruction, when one is informed of the knowledge to be acquired). It has been suggested that these strategies differ in terms of their brain bases (e.g., stronger reliance on neocortical language systems in fast mapping than explicit encoding: Merhav et al. 2015), but neural correlates of these learning types remain elusive (Atir-Sharon et al. 2015). Here, we address this question by comparing event-related potentials elicited while learning new wordforms either implicitly in a fast-mapping (FM) fashion or through explicit encoding (EE), with the two tasks being maximally matched for their auditory and visual features as well as the manner of presentation.

We designed a set of audio-visual stimuli that combined images of one familiar and one novel object with auditorily presented questions. Under the FM condition, the meaning of a new wordform could be inferred from the context (e.g., ‘Is BYZ striped?’, with 2 images presented simultaneously, only one of them (novel) striped). Under the EE condition, the novel image and the wordform corresponded explicitly, and the task was to learn this correspondence (e.g. ‘This is BYZ. Will you remember it?’). All wordforms were balanced for length, bigram and diphone frequency, and counterbalanced across experimental conditions. Ten novel words were presented in each condition to 12 healthy right-handed Russian-speaking volunteers; as a control condition, familiar words and images were used in identical presentation settings. All conditions were pseudo-randomised in a single sequence. We recorded ERPs elicited by passive listening of wordforms and compared them before and after the training session.

Behavioral results (accuracy and reaction time in free recall, recognition and semantic word-picture matching tasks) showed that our paradigm ensured equally successful learning of new wordforms in both FM and EE conditions, with no significant differences between the two. Auditory ERPs peaked at ~170, 250 and 520 ms. All three peaks indicated clear learning effects, with ERPs elicited by familiar and novel wordforms diverging before training but becoming highly similar following the learning session. Moreover, the topography of ERPs effects differed significantly between the FM and EE conditions in the early negativity arising at ~170 ms. For wordforms presented under EE condition, this topography displayed right-lateralised changes, whereas for those presented in the FM condition, learning-related dynamics was distributed more centrally/left-lateralised. Later peaks showed a number of further strategy-related differences between ERP topographies elicited by the novel wordforms after the training. Analysis of cortical activity sources using LORETA suggested that the learning dynamics was underpinned by differentially lateralised activity in anterior temporal lobes of the two hemispheres as well as parietal and frontal areas.

The results support the existence of two different neural learning mechanisms for word acquisition (Dollaghan 1985; Merhav et al. 2015). In spite of the highly similar mode of presentation and tight control over stimulus features, which led to equal behavioural performance and learning outcomes, the neurophysiological activity underpinning this performance diverged between EE and FM conditions and involved different neural generators in the two cerebral hemispheres. These results may inform development of targeted learning strategies for use in education and in amelioration of language impairments.

Supported by the grant of the Government of Russian Federation № 14.W03.31.0010.

IMPLICIT CAUSALITY BIASES AND THEMATIC ROLES IN ASL VERBS

Anne Therese Frederiksen & Rachel I. Mayberry (University of California, San Diego)
a.t.frederiksen@ucsd.edu

Cross-linguistic research has shown that thematic structure of verbs predicts their implicit causality (IC) bias, with stimulus-experiencer (SE) verbs eliciting NP1-biases, experiencer-stimulus (ES) verbs eliciting NP2-biases, etc. (Rudolph & Försterling, 1997; Ferstl et al, 2011, Hartshorne & Snedeker, 2012). However, SE verbs are assumed not to exist in sign languages (Kegl, 1990; Meir et al, 2007; Oomen 2017), either because they are used as one-place predicates (Edge & Herrmann, 1977), or because they are interpreted as ES verbs (Winston, 2013). This begs the question of how IC biases in sign languages are distributed, as the strongest and most consistent subject-biases in documented spoken languages are found in SE verbs (Goikoetxea et al., 2008; Ferstl et al., 2011). A lack of SE verbs would imply a different relationship between thematic roles and biases, or a different distribution of IC verbs in signed compared to spoken languages.

As the first of its kind, the present paper provides norming results for IC biases in a large number of American Sign Language (ASL) verbs. We examined the thematic roles in potential SE verbs, and analyzed how they relate to IC biases. We used a sentence completion paradigm to test this. As no published list exists of thematic roles in ASL verbs, we selected verbs that were categorized as SE (e.g. *inspire*), ES (e.g. *love*), agent-patient (e.g. *kick*), or agent-evocator (e.g. *blame*) in English. We first assessed through a small-scale acceptability study that these verbs could be used as transitives in ASL. Following this, native signers (N=8) completed sentence fragments (n=239) of the type 'NP V NP WHY?', 'NP V NP, because ...'. Their responses were coded for whether the subject or object from the fragment was mentioned as the subject in their free continuation. Calculating a bias score for each verb showed that 88 verbs were biased towards the subject, and 116 towards the object. A comparison between logistic mixed effects analyses with verbs and subjects as random effects and with/without thematic role as the fixed effect confirmed thematic role as a significant predictor for the next mentioned referent ($\chi^2(3) = 30.054, p < 0.001$), as expected from prior research. However, pairwise comparisons between the different levels of the fixed effect revealed that SE verbs only differed significantly from ES verbs, not agent-evocator verbs. This suggests that there is no clear subject-bias in the assumed SE verbs as a whole in ASL. Instead, a detailed analysis revealed that verbs within this category range from strongly subject-biased (e.g. CHARM, FLATTER, EMBARRASS) to strongly object-biased (e.g. COMFORT, ANNOY, INTEREST). Analyzing the context of the sentence completions revealed that the majority of verbs with object-bias were used with ES structure instead of the expected SE structure, and vice versa.

In conclusion, the results of the present study provide norming data on implicit causality biases in ASL verbs. This forms the basis for future psycholinguistic studies of ASL. We further show that thematic role predicts IC bias in (at least one) signed language as well as spoken language, suggesting that this principle is universal across not only language but also modality. Our results do not support previous claims that the stimulus-experiencer category does not exist in ASL; rather, many potential SE verbs are in fact acceptable as transitives, but vary in whether they are interpreted as ES or ES verbs. This accounts for the relatively low occurrence of IC verbs with subject-bias, compared to those with object-bias in our norming data, something that researchers should take into account in designing experimental studies in ASL.

Rudolph, U., & Försterling, F. (1997). *Psychol Bull.* Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). *Behav Res Methods.* Hartshorne, J. K., & Snedeker, J. (2013). *Lang Cogn Process.* Goikoetxea, E., Pascual, G., & Acha, J. (2008). *Behav Res Methods.* Kegl, J. A. (1990), *Sign Language Research: Theoretical Issues.* Meir, I., Padden, C. A., Aronoff, M., & Sandler, W. (2007). *J of Ling.* Oomen, M. (2017). *SL&L.* Edge, V. & Herrmann, L. (1977), *On the other hand.* Winston, C. (2013). MA Thesis. Purdue University.

ANIMACY-DRIVEN EXPECTATIONS IN NORWEGIAN RELATIVE CLAUSE PROCESSING

Dave Kush (NTNU Trondheim) & Ragnhild Eik (NTNU Trondheim)

dave.kush@ntnu.no

Subject relative clauses (SRCs) are often easier to comprehend than object RCs (ORCs) [1-3]. It has been argued that this SRC advantage can be attributed, in part, to expectation: upon encountering the left-edge of an RC, incremental parsers expect subject RC continuations and experience difficulty when those expectations are violated [4,5]. Some work suggests that ORCs are less difficult to parse when they have *inanimate* heads than *animate* heads. If expectations underlie RC-advantages, it is possible that animate heads engender expectations for SRC continuations, while inanimate heads lead to expectations for an ORC (see [8]). We investigated whether we could find evidence for such expectations in Norwegian.

Experiment 1 (SPR, N=32). Participants read the Norwegian equivalents of RC-containing sentences like (1) and (2). We manipulated RC-head animacy and whether the RC was an SRC or an ORC. In SRC sentences an adverb (*recently*) followed the complementizer and served as the first cue to the SRC parse. Such adverbs allowed us to decouple effects of structural expectation and integration: participants could identify the type of RC they were processing *before* encountering the RC-internal verb. If animate heads license an expectation for an SRC, RTs at the adverb should be lower in the *Animate-SRC* condition than in the *Inanimate-SRC* condition. If inanimate heads lead to expectations for ORCs, we expect more difficulty at the RC subject in the *Animate-ORC* condition than in the *Inanimate-ORC* condition. **Analysis & Results:** We compared residualized logRTs at the first word in each RC (*the* v. *recently*). SRCs were read more quickly than ORCs ($t = 2.30$, maximal LMEM), qualified by a RC-TYPE x ANIMACY interaction ($t = 2.89$). SRCs were read more quickly than ORCs when the head was *animate* ($t = 3.69$), but there was no effect of RC-TYPE with *inanimate* heads ($t < 1$). A similar interaction was observed at the following adverb (*recently*). **Experiment 2 (Corpus Search).** We checked whether head animacy impacts the probability of SRC- and ORC- continuations in written text, as it has been proposed that parsing expectations reflect distributional patterns in language users' experience. We extracted all RCs with overt relative pronouns from the parsed *NorGramBank Newspaper* corpus [10], and annotated the RCs for type and head animacy. Results of this process, shown in **Table 1**, show that SRCs are far more likely than ORCs, but even more so after animate heads than after inanimate heads ($p(\text{SRC}|\text{animate}) \sim 0.98$ v. $p(\text{SRC}|\text{inanimate}) \sim 0.85$). This is broadly consistent with the idea that experiential factors can play a role in parsing expectations, as argued by many [3,4,7,8]. **Experiment 3 (Sentence Completion, N=68).** We also checked whether expectations in the SPR task would be reflected in production preferences, as these preferences have been argued to serve as reliable proxies for expectation in comprehension [7,9]. We had Norwegian-speaking adults provide completions for sentences based on (1). Participants exhibited an overwhelming preference for subject RC completions (88% v. 9% direct object completions), irrespective of head animacy, but non-subject completions were more frequent with inanimate heads (79/390, ~20%) than with animate heads (14/406, ~3%). **Take Home.** SPR results indicate that animate heads license strong syntactic expectations for an SRC in comprehension, but expectations following inanimate heads are equivocal, in line with [8]. Expectations do not straightforwardly align quantitative estimates of participant experience from corpus search and production preferences.

- (1) The researcher trusted the {assistants/results} that *recently* had ... [SRC]
 (2) The researcher trusted the {assistants/results} that *the thorough group* had... [ORC]

References: [1] King, J. & Just., M.A. (1991). *JML*; [2] Traxler, M. Morris, M.K. & Seely. R.E. (2002). *JML*; [3] Mak, W.M., Vonk, W. & Schriefers, H. (2002). *JML*; [4] Roland, D., Dick, F. & Elman, J.L. 2007. *JML*; [5] Staub, A. 2010. *Cognition*; [6] Traxler, M., Williams, R.S., Blozis, S.A., & Morris, R.K. 2005. *JML*; [7] Gennari, S.P. & MacDonald, M.C. 2008. *JML*; [8] Wagers, M. & Pendelton, E. (2015). Proceedings of WCCFL 33. [9] Gennari, S. & MacDonald, M.E. (2009). *Cognition*. [10] [11] Rosén, V. et al. 2010. LREC2012

Friday Poster.19

FREQUENCY EFFECTS OF MULTIWORD SEQUENCES IN MANDARIN CHINESE
Ching Chu Sun (University of Tübingen), Peter Hendrix (University of Tübingen),
Harald Baayen (University of Tübingen)
ching-chu.sun@uni-tuebingen.de

At the word level, the effect of frequency is perhaps the most well-documented effect in the psycholinguistic literature. Recently, frequency effects above the word level have been observed as well. Arnon and Snider (2010), for instance, observed that phrasal decision latencies were shorter for high frequency phrases such as “all over the place” as compared to low frequency phrases such as “all over the city”. (see also Siyanova-Chanturia, Conklin and Van Heuven, 2011; Bannard & Matthews 2008).

Here, we present the results of two studies that extend previous work on frequency effects for multi-word sequences in two ways. First, the current studies are the first to report frequency effects of multiword sequences for a non-alphabetic language: Mandarin Chinese. Second, whereas previous studies primarily reported frequency effects of multiword sequences in language comprehension, the focus of the current work is on frequency effects of multiword sequences in language production.

First, we carried out a phrase reading experiment in Mandarin Chinese. Whereas previous studies contrasted low frequency phrases with carefully selected high frequency phrases, the multiword sequences in the current experiment were randomly selected trigrams from a large-scale corpus of Mandarin Chinese: the Simplified Chinese Corpus of Webpages (SCCoW; Shaoul, Sun & Ma, 2016). We analyzed the experimental data analyzed with generalized-additive mixed-effect models (GAMMs; Wood 2006, Wood 2011).

The analyses revealed significant effects of trigram frequency across the predictor range for both naming latencies ($\chi^2 = 22.256$, $p < 0.001$) and acoustic durations ($\chi^2 = 37.520$, $p < 0.001$). Response times were shorter for high frequency multiword sequences, as were acoustic durations. The effects of trigram frequency existed over and above the effects of unigram frequencies, bigram frequencies, and phonological length (i.e., the number of phonological segments in a trigram).

Second, we extracted the acoustic durations of word trigrams from a corpus of spontaneous speech in Mandarin Chinese: the Taiwan Mandarin corpus of spontaneous speech (Fon, 2004). As was the case for the experimental data, GAMM analyses revealed a robust effect of trigram frequency on acoustic durations ($F = 42.667$, $p < 0.001$), independent of the effects of component bigram and unigram frequencies and the effect of phonological length. The effect of trigram frequency was qualitatively similar to the effect of trigram frequency in the experimental data, with shorter pronunciation durations for more frequent word trigrams (cf. Arnon & Cohen Priva, 2013 for comparable findings in English).

The results of the work reported here demonstrate that the frequency of multiword sequences influences lexical processing not only in speech comprehension, but also in speech production. In addition, we established that frequency effects of multiword sequences are not limited to alphabetical languages, but also exist in the character-based language Mandarin Chinese. The observed pattern of results furthermore indicates that language users are aware of the combinatorial properties of words.

At least two possible interpretations of this awareness exist. First, in theories like data-oriented parsing (Bod, 2006) and memory-based learning (Daelemans & Van den Bosch, 2005), at least some multiword sequences are stored as a whole in the mental lexicon. Given the fact that we observed trigram frequency effects across the trigram frequency range, this would require the storage of hundreds of millions of multiword sequences. Alternatively, knowledge about combinatorial properties of words is stored in association patterns between word-level representations (Baayen, Hendrix & Ramscar, 2013). Such an interpretation, we argue, fits more straightforwardly with the current findings.

Friday Poster.20

**PRESERVATION OF PHONOLOGICAL CONTRAST DOES NOT BLOCK BUT DOES
ATTENUATE PHONETIC IMITATION**

Václav Jonáš Podlipský (Palacký University Olomouc), Šárka Šimáčková (Palacký University Olomouc) & Filip Smolík (Academy of Sciences of the Czech Republic)
vaclav.j.podlipsky@upol.cz

Spontaneous phonetic imitation (or convergence), i.e. automatic adoption of pronunciation features of recently perceived speech, has been well documented, e.g. [1-3]. Previous research showed that imitation is constrained by the interlocutors' social [4, 5] and possibly also linguistic closeness [6, 7]. Inspired by Nielsen's [8] results of a shadowing study (repetition of prerecorded isolated words), this study tests the hypothesis that imitation is constrained by the preservation of contrast between phonological categories, so that phonetic realization shifted towards a category boundary is less likely imitated.

Czech, a vowel-quantity language, contrasts short /u/ and long /u:/, and also prevoiced (negative-VOT) /d/ with unaspirated /t/. We prepared Czech words with reduced prevoicing in /d/ (a shift towards /t/), extended prevoicing in /d/, reduced duration of /u:/ (a shift towards /u/), and extended duration of /u:/. **Experiment 1**, with 18 Czech listeners, tested AX discrimination of the manipulated against the original naturally-produced words to assess the perceptual magnitude of the manipulations. Results showed that (i) the salience of the manipulations was low, i.e. the manipulated segments were within the range of phonetic variability (as intended); (ii) manipulations of /d/ prevoicing duration were relatively less salient than those of vowel duration; and (iii) the reductions (i.e. shifts towards the contrasting categories) were slightly more noticeable than the extensions (in line with 'categorical perception' [9]). **Experiment 2** assessed how close the reductions shifted the segments towards the boundaries with their phonological counterparts. Another 18 Czech listeners heard the words with only the /u:/ and the /d/ (plus a third of the following vowel) retained and the rest low-pass-filtered, and indicated their certainty of having heard /u/ vs /u:/ and /t/ vs /d/ on a continuous scale. We found that (i) both /d/ reduction and /u:/ reduction led to listeners' decreased certainty, while (ii) the /u:/ reductions resulted in shifts closer to the /u/-/u:/ boundary than did the prevoicing reduction to the /t/-/d/ boundary, possibly due to the differing salience (cf. Exp 1) or differing perceptual-cue weight for the respective sound contrast. **Experiment 3**, the actual imitation experiment with a pretest-shadowing-posttest design, had 16 new native Czech participants. We indeed found asymmetrical imitation patterns for reductions vs extensions. For /d/, only extended prevoicing was imitated, not reduced prevoicing, even though this direction of manipulation was the slightly more salient one (cf. Exp. 1). For /u:/, extension was imitated, and so was reduction (being more salient than /d/ reduction, cf. Exp. 1 and 2); however, (i) the degree of imitation was much smaller for /u:/ reduction than for its extension, and (ii) imitation transferred from shadowing to posttest only for /u:/ extension, not for reduction.

To conclude, we found that contrast preservation, as a potential mechanism of maintaining relative stability of phonological representations against diachronic sound change, does not necessarily completely preclude imitation (which may thus actuate sound change, e.g. [10]): /u:/ reduction (which shifted the long vowel to its short counterpart) was imitated to some extent. At the same time though, since the shifts away from contrasting categories (extensions) were clearly more likely to be imitated than the shifts towards them (reductions), contrast preservation does seem to attenuate the degree of phonetic imitation.

References

- | | |
|---------------------------------------------------------------|----------------------------------------------------------------|
| [1] Gambi & Pickering (2013). <i>Front Psychol</i> , 4, 340. | [6] Kim et al. (2011). <i>Lab Phonol</i> , 2(1), 125–156. |
| [2] Nguyen & Delvaux (2015). <i>J Phonetics</i> , 53, 46–54. | [7] Olmstead et al. (2013). <i>Front Psychol</i> , 4, 475. |
| [3] Pardo (2013). <i>Front Psychol</i> , 4, 559. | [8] Nielsen (2011). <i>J Phonetics</i> , 39(2), 132–142. |
| [4] Pardo (2006). <i>J Acoust Soc Am</i> , 119(4), 2382–2393. | [9] Repp (1984). <i>Speech Lang</i> , 10, 243–335. |
| [5] Babel (2012). <i>J Phonetics</i> , 40(1), 177–189. | [10] Stevens & Harrington (2014). <i>Loquens</i> , 1(1), e003. |

COHERENCE AND FINITENESS EFFECTS IN EXTRACTION FROM ADJUNCT ISLANDS IN ENGLISH

Christiane Müller, Damon Tutunjian & Anna-Lena Wiklund (Lund University)
christiane.muller@nordlund.lu.se

Adjunct clauses are standardly treated as strong islands for extraction. However, a closer look suggests that extraction from adjunct clauses is possible in English, provided that a coherence relation (e.g., a causal, as opposed to a purely temporal relation) holds between the events referred to by the matrix and the adjunct clause, and provided that the adjunct is non-finite (Truswell, 2011). Observations like these raise important questions regarding filler-gap association in island domains. If coherence and finiteness matter for extraction from adjunct clauses, then this would call into question claims that filler-gap association is suspended in island domains (e.g., Traxler & Pickering, 1996). The current study investigates how coherence and finiteness affect the acceptability and the real-time processing of adjunct island extraction in English.

In Experiment 1, we collected acceptability judgments for 40 sentences in which we manipulated the telicity of the matrix verb such that sentential coherence was either augmented (coherent, 1a) or impeded (non-coherent, 1b) and the finiteness of the adjunct clause ([finite/non-finite] in 1). Seventy-two monolingual native English speakers (rotated over four lists) rated the sentences using a 7-point scale. Linear mixed models were used to analyze z-scored responses. As expected, coherence improved extraction ($\beta = 0.190$, $t = 4.814$, $p < .001$). Furthermore, a finiteness by coherence interaction was found, indicating that non-finiteness benefits only coherent structures ($\beta = 0.058$, $t = 1.925$, $p = .05$).

Experiment 2 used self-paced reading (48 monolingual native English speakers over four lists) to investigate the online processing of a set of 40 sentences similar to, but partially modified from what was used in Experiment 1 (2). A linear mixed models analysis of log residualised reading times revealed a main effect of coherence, with marginally faster reading times for coherent sentences at the matrix verb ($\beta = -0.014$, $t = -1.759$, $p = .09$) and significantly faster times for coherent at the adverb ($\beta = -0.339$, $t = -28.329$, $p < .001$), the embedded verb ($\beta = -0.014$, $t = -2.232$, $p < .05$) and the wrap-up region ($\beta = -0.022$, $t = -2.517$, $p < .05$). At the embedded verb, we also found a main effect of finiteness (non-finite faster than finite, $\beta = -0.019$, $t = -2.962$, $p < .05$). Moreover, at the wrap-up region, we observed a trending interaction between coherence and finiteness ($\beta = 0.011$, $t = 1.625$, $p = .10$) which was driven by an emergent finiteness effect in pairwise comparisons for coherent structures (non-finite faster than finite, $p < .05$) but not for non-coherent structures ($p = 0.80$). Taken together, this suggests that coherence and an absence of finiteness facilitate processing and dependency formation in adjunct islands.

Our finding that coherence and finiteness have an impact on the acceptability of extraction from adjuncts as well as on the processing of such structures at the point of filler integration suggests that gap assignment need not be suspended in adjunct clauses. One possibility is that gap assignment is supported in the presence of a cue (e.g., telicity of the matrix verb) for a tighter semantic relation (coherence) between the adjunct and the matrix. If lack of integration is taken to be the defining characteristic of island effects, this suggests that coherent adjuncts are not strong islands, but are permeable to some degree.

- (1) a. coherent | non-finite/finite
Which beer did he almost stumble [after chugging / after he chugged]?
b. non-coherent | non-finite/finite
Which beer did he stroll a little [after chugging / after he chugged]?
- (2) a. coherent | non-finite/finite
Which beer did he stumble immediately [after chugging/after he chugged] last night?
b. non-coherent | non-finite/finite
Which beer did he stroll a bit [after chugging/after he chugged] last night?

References: Traxler & Pickering (1996). *JML*, 35, 454-475 • Truswell (2011). *Events, phrases and questions*. Oxford: Oxford University Press.

Friday Poster.22

'GEDOWNLOADET' OR 'DOWNGELOADET'? PARTICIPLE FORMATION OF MULTIMORPHEMIC ENGLISH LOAN VERBS IN GERMAN

Britta Schulte (Universität Potsdam) & Ulrike Freywald (Universität Potsdam)
bschulte@uni-potsdam.de

Participle formation in German multimorphemic verbs such as *anmerken* 'to remark' or *brandmarken* 'to brand', in addition to the attachment of a participle affix, involves attaching the syllable *ge-* to the verb stem. Such verbs can be divided into three subgroups, depending on whether the *ge-* occurs in word-initial position (e.g. *brandmarken* – *gebrandmarkt* 'brand – branded'), word-internal position (e.g. *anmerken* – *angemerkt* 'remark – remarked'), or whether its position is variable (e.g. *bauchlanden* – *gebauchlandet* / *bauchgelandet* 'belly-land – belly-landed'). The interesting question that arises when thinking about loan verb integration is into which category loan verbs construed as multimorphemic are integrated.

The present study investigates the cognitive mechanisms underlying participle formation in multimorphemic English loan verbs such as *downloaden* 'to download' or *babysitten* 'to baby-sit' and therefore how they are integrated into the above-mentioned subgroups of German participles. For the participle forms of such words, the position of the *ge-* is currently optional, with both *ge-initial* (e.g. *gedownloadet*) and *ge-internal* (e.g. *downgeloadet*) frequently occurring in the language (e.g. Hausmann, 2006). This raises the question of how a speaker decides whether to use the *ge-initial* or *ge-internal* form during production.

21 native speakers of German (mean age 26.4, 10 female) participated in a production priming experiment based on a cloze task. Target stimuli consisted of short sentence fragments with blanks, which the participants had to fill in with the participle form of a given loan verb (e.g. *downloaden*). Each target was preceded by a short prime sentence containing either (1) a participle of a multimorphemic verb with the *ge-* in word-initial position (e.g. *gebauchlandet* 'belly-landed'), (2) the participle of the same verb with the *ge-* in word-internal position (e.g. *bauchgelandet* 'belly-landed'), or (3) the participle of a non-*ge-*verb (e.g. *verstorben* 'passed away'), which served as a baseline. If participle formation for English multimorphemic loan verbs is based on the activation of morphological rules, prior activation of the *ge-initial* rule during processing of a *ge-initial* prime should lead to an increased number of *ge-initial* forms when producing the target participle. If, in contrast, participle formation is instead based on the frequency of a particular form, prior exposure to a different participle with a *ge-* in an initial position should not affect which form is chosen.

The results show significantly more *ge-initial* target completions following *ge-initial* primes than following *ge-internal* or baseline primes. As the primes were otherwise completely unrelated to the target verb, this suggests that participants indeed activated a morphological rule (which contained information about the position of the *ge-*) when processing the prime. This rule was still active during completion of the target, which led to a priming effect. We conclude that, at least for these specific multimorphemic loan words, participle forms do not possess separate entries in the mental lexicon, but that the production of such participles instead involves the application of abstract morphological rules which determine the position of the *ge-*.

**VARIATION IN FRENCH PARTIAL INTERROGATIVES:
SOCIAL MEANING AS A KEY FACTOR TO UNDERSTAND SOCIOLINGUISTIC NORM VIOLATIONS**
- Gabriel THIBERGE & Barbara HEMFORTH (LLF, CNRS, Université Paris-Diderot, LabexEFL) -
Contact: gtac@tuta.io

French partial interrogatives show considerable variation (Coveney, 2011). The *wh*-element can be in a declarative argumental position (in situ, IS, “Ils mangent où ?”/“They eat where?”). It can also be in a fronted position, with subject-verb inversion, FINV (“Où mangent-ils ?”/ “Where eat they?”) or without inversion, F (“Où ils mangent ?”/“Where they eat?”). Syntactic constraints have been used to explain the differences in frequency of use for those variants: e.g. fronting is often analysed as structurally more ‘complex’ because of an underlying movement (Jakubowicz, 2011).

From a sociolinguistic view, previous research mostly studied the preferred use of one variant or another by specific social groups (Quillard, 2001). We extend this perspective in the framework of social meaning games (Burnett, 2017), where variation is a tool to socially position oneself in a specific context of interaction. To that end, we ran a series of two acceptability judgment tasks (AJT) and a written matched-guise test (MGT, Lambert, Hodgson, Gardner, & Fillenbaum, 1960).

Experiment 1: We ran an online written AJT (57 participants, 15 items, scale 1-10). It shows that FINV interrogatives are judged significantly better than IS questions (see Figure 1; $p < .001$). This preference is entirely due to judgments of participants older than 30 years (interaction age*question type: $p < .01$; see Figure 2), which argues for a ‘generation gap’ among French speakers with regards to the consideration of all three interrogative variants.

Experiment 2: An online MGT provides a more fine-grained account: the perception of a French speaker’s social persona (e.g. Eckert, 2008; Ochs, 1992) is affected by the syntactic structure he/she uses when producing partial interrogatives. 58 new participants from various social and educational backgrounds were presented with three interviews by a “journalist” that contained only one of the three question types. They judged the journalist on a variety of properties (Figure 3). Focusing on FINV vs. IS constructions, the use of FINV leads to a biased social perception on the receiver’s part. People using FINV structures are thought to be of a higher social status, to have studied for a longer time, or to be older but also to be less relaxed.

Experiment 3: In an online AJT (44 new participants, 30 new items, scale 0-10), we integrated the questions in a formal or informal context. Results confirm the preference for FINV constructions over the two other variants as well as the age factor. Data also show that the preference for FINV constructions is context-dependent, in particular for participants 30 years and older. They judge FINV interrogatives significantly better ($p < 0.02$) in formal contexts than in informal ones (Figure 4).

Our AJT results, in combination with the MGT results, argue for the idea that French native speakers do make a meaningful choice when considering either one of the variants available to them for partial interrogatives. They attach social positioning values on various dimensions to each variant, and AJT judgments reflect these social positioning values. Whether that choice is conscious or not in production remains an open question to be studied in future research.

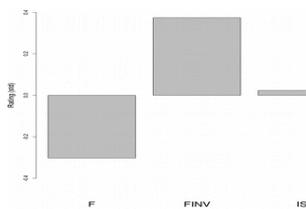


Figure 1: AJT 1 overall results (std.)

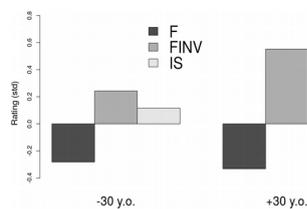


Figure 2: AJT 1 results by age (std.)

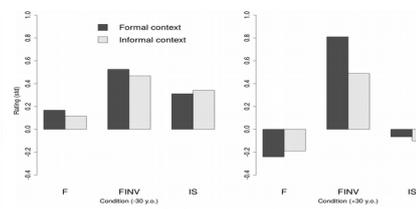


Figure 4: AJT 2 global results (std.)

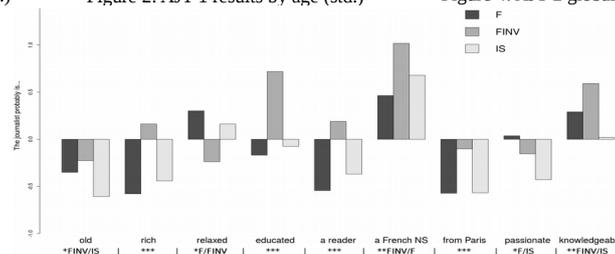


Figure 3: MGT global results (std.)

Friday Poster.24

CONSTRUAL IN LANGUAGE: A VISUAL WORLD APPROACH

Srdan Medimorec, Petar Milin, & Dagmar Divjak (The University of Sheffield)
s.medimorec@sheffield.ac.uk

Language provides a variety of ways to express events. For example, speakers can choose between using the active or passive voice to express the same situation. Lexical and syntactic choices reflect a specific framing of the experience and a certain commitment to how that experience will be communicated between interlocutors. Language can thus be seen as a highlighting device that promotes or demotes the salience of various situational cues; this, in turn, modulates how we attend to, perceive and process information. Cognitive linguists use the theoretical concept of “construal” to account for these alternating ways of expression. Even though construal has the potential to provide insight into the relation between language choice and conceptualization, there is a dearth of empirical research investigating construal phenomena. Furthermore, most attempts to explain linguistic choices by appealing to alternative construals have relied heavily on the analysts' own intuitions about the data.

In the current study we set out to investigate whether differences in construal could trigger different conceptualizations of the situation. We used a variation of the Visual World paradigm to investigate whether linguistic encoding affects the way in which events are perceived, and thus potentially conceived, by speakers. We focused on three different linguistic contrasts, ranging in strength from quite obvious to very subtle semantic manipulations (i.e., different ways in which the relation among the elements in a scene was described): preposition (typical/atypical as in *the screen is on the desk* versus *the desk is under the screen*), voice (active/passive as in *the girl pushed the boy* versus *the boy was pushed by the girl*), and dative (noun-phrase/prepositional-phrase, as in *she gave him a book* versus *she gave a book to him*).

Sixty University of Sheffield students and staff participated in an eye-tracking study run on an EyeLink Portable Duo eye-tracker (SR Research Ltd). In block 1 of the experiment, participants viewed a set of 48 full-coloured photographs depicting naturalistic scenes. Across blocks 2 and 3, participants heard 96 individual sentences describing these 48 different events in different ways (e.g., active/passive). Each sentence was immediately followed by an image from block 1, depicting the event described in the sentence. Images contained 2 or 3 areas of interest (AOI, empirically determined by combining eye-tracker generated heat maps with relevant events described in the corresponding sentences). We fitted a Generalized Additive Mixed Effects model to the relationship between eye movement measures and condition (i.e., free viewing versus different constructions describing the same event). The dependent variables were conditioned on interest areas. They contained an early measure of information uptake, specifically the order of access (calculated using start time of the first fixation in an interest area) and two later measures of information uptake: first run dwell time (i.e., summation of fixations before moving out of an area for the first time) and total dwell time (i.e., summation of the duration of all fixations in an area). We also analysed average pupil size (per interest area).

Our results show that construal interacts with naturalistic scene viewing tendencies in a number of ways. Language increases the cognitive load by adding another source of relevant information, as indexed by increases in pupil diameter for all language-mediated conditions compared to naturalistic scene viewing. Critically, language affects the order in which AOIs are accessed and the time spent viewing the AOIs, but these effects differ among the three linguistic contrasts. While the Preposition alternation shows strong and consistent early and late effects of construal, Voice shows early effects only and Dative shows late effects that are inconsistent with order of mention. The extent to which the informational value or relevance of the scene components is modulated by language thus depends on the semantic strength of the linguistic manipulation.

Learners generate unencountered novel structures drawing on knowledge of a universal

Adam Milton Morgan and Victor S. Ferreira, UC San Diego (adam.milton.morgan@gmail.com)

Adults readily learn new syntactic properties of their L1 (e.g., the “car needs washed” construction^[3]) and new syntactic representations (as in L2 learning). The mechanism of such grammar acquisition is often modeled as abstraction over input. Here we ask: Can adults generate new structures without input over which to abstract? For instance, imagine you are learning a new language with various types of relative clauses (RCs), including subject (SRCs; 4) and direct object (DORCs; 5). By chance, you only gain exposure to DORCs. Can you create a structural representation of a SRC? If so, does that generalization draw on knowledge of your L1, or can you appeal to latent knowledge of general/universal properties of human language?

In **Exp.1** we taught 48 monolingual English speakers to describe pictures with English words but a new grammar for RCs^[3]. English puts RCs after the noun they modify (4-6), but other languages put them before (*preminally*; 1-3). One group ($N=24$) was trained only on prenominal SRCs (1), and another ($N=24$) only on prenominal DORCs (2). After training, we elicited both exposed-type RCs (e.g., SRCs) and unexposed-type RCs (e.g., DORCs). If grammatical learning exclusively relies on abstracting over input, then participants should not be able to produce unexposed structures. If they do generalize, and if they draw on language-specific knowledge (knowledge of English RCs), both groups should be able to produce the untrained structure because English has both SRCs and DORCs. If speakers draw on language-general knowledge, we expect DORC-trainees to generalize to SRCs, but SRC-trainees not to generalize to DORCs. This follows from a typological pattern known as the NP Accessibility Hierarchy^[4] (NPAH): If a language allows one type of RC, e.g., DORCs, it also allows all higher types (SRCs), but not necessarily lower ones (indirect object RCs (IORCs; 6)).

Table 1.	prenominal (e.g., Chinese, Korean)	postnominal (e.g., English, Thai)
subj.RC	(1) a [that <u> </u> _s gave a toy to a dog] girl	(4) a girl [that <u> </u> _s gave a toy to a dog]
dir.obj. RC	(2) a [that a girl gave <u> </u> _{do} to a dog] toy	(5) a toy [that a girl gave <u> </u> _o to a dog]
indir.obj. RC	(3) a [that a girl gave a toy to <u> </u> _{io}] dog	(6) a dog [that a girl gave a toy to <u> </u> _{io}]

SRC-trainees produced very few (6%) DORCs at test; DORC-trainees generalized to SRCs more (19%; $p < .001$). Another group ($N=24$) received a brief explicit grammar lesson on prenominal DORCs and no exposure. At test, these participants did not reliably produce well-formed prenominal RCs (5%), suggesting that the former two groups’ performance was not based on explicit grammatical knowledge. The asymmetrical generalization between groups is inconsistent with the English pattern, but consistent with the NPAH.

Exp. 2 was a replication of Exp.1 in which we prevented participants from using passive voice, successfully encouraging them to rely on language-specific (English) knowledge. At test, SRC-trainees ($N=24$) produced 19% DORCs and DORC-trainees generated 22% SRCs (n.s.).

In **Exp. 3**, we trained participants on a typologically attested, verb-final prenominal RC grammar (e.g., “the [_s to the dog a toy gave] girl.”). 3 groups ($N=24$ each) were trained on subsets of the RC paradigm (SRCs & DORCs, DORCs & IORCs, SRCs & IORCs) and tested on the full paradigm. Because the internal structure of these RCs is unlike English, participants could not exclusively rely on knowledge of English to generate novel RC types. Preliminary data ($N=32/72$) again show a clear asymmetrical pattern in generalization consistent with the NPAH.

In sum, we find that adults are capable of generating previously unencountered syntactic representations, the first such evidence that we are aware of. Further, whatever pressures give rise to the systematic distribution of RCs across languages—be they innate knowledge of UG or constraints imposed by limited general cognitive resources—they can also guide the generation of new structures in adult language learners, consistent with previous work showing that learning biases pattern with linguistic universals.^[1,2] This is problematic for strong versions of empiricist accounts in that learners use more information than can be gleaned from the input.

[1] Christiansen 2000; [2] Fedzechkina et al. 2016; [3] Kaschak 2006; [4] Keenan, Comrie 1979

Friday Poster.26

CASCADING ACTIVATION: EVIDENCE FROM DETERMINER COMPETITION

Anja Riemenschneider (Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin), Christine Mooshammer & Katharina Spalek (Humboldt-Universität zu Berlin).
 anja.riemenschneider@iqb.hu-berlin.de

The notion of cascading activation in spreading activation models is widely agreed upon in the current literature on speech production. However, most studies focus on the activation flow between the lexical and the phonological processing levels. Considerably less attention has been paid to the level of articulation, although there has been evidence for cascading activation onto the latter (cf. Goldrick et al., 2006). In order to gather further insight involving the articulatory level, the verbal picture naming answers from Spalek et al.'s (2010) RT experiment on determiner competition were analysed phonetically. In the experiment, English definite determiners, varying in their phonological realisation between the forms /ðə/ in front of consonants and /ði/ in front of vowels, were produced alongside NPs consisting of an adjective and a noun. In the phonologically consistent condition, both adjective- and noun-onsets were consonants or vowels (e.g. *the orange elephant*), respectively, whereas in the inconsistent condition, the onsets differed (e.g. *the purple elephant*), leading to conflicting activation of the determiner form required by the noun. We hypothesized that in the inconsistent condition, the activation of the non-selected determiner form would cascade onto the articulatory level, causing phonetic distortions in the vowel productions in the direction of the conflicting vowel's quality (first and second formants, representing vowel openness and vowel backness, respectively). To test this hypothesis, we used linear mixed models including a random intercept to account for the variability across subjects, and controlling for gender. We found significant main effects of consistency in the predicted direction for the second formant (F2). For NPs including the adjective with consonant-onset (e.g. *the purple alligator*), the F2 value of consistent items was 51 Hz lower than for inconsistent items (only for male speakers, mean ca. 1480 Hz, $t=-2.4$, $p=0.02$), for those including the adjective with vowel-onset (e.g. *the orange giraffe*), the F2 value of consistent items was 53 Hz higher than for inconsistent items (mean ca. 1980 Hz, $t=2.6$, $p=0.01$). For the first formant (F1), From these results, we conclude that activation of concurrently activated elements on the phonological level affects the level of articulation, supporting language production models that assume cascading activation between processing levels.

Table: Mean values for F2 (in Hz), split by adjective-onset, consistency, and gender

Adjective-onset		consistent	inconsistent	Mean
Vowel	f	2031.55	1979.28	2005.25
	m	1895.72	1890.69	1893.22
Consonant	f	1705.97	1705.44	1705.71
	m	1438.17	1488.94	1464.08

Literature:

- Goldrick, M. and S. E. Blumstein (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and cognitive processes* 21 (6), 649-683.
- Spalek, K., K. Bock, and H. Schriefers (2010). A purple giraffe is faster than a purple elephant: Inconsistent phonology affects determiner selection in English. *Cognition* 114, 123–128.

WHAT DEFINES GRAMMATICAL GENDER OF RUSSIAN EXPRESSIVE NOUNS?

Natalia Chuprasova (Saint-Petersburg University), Varvara Magomedova (SUNY, Stony Brook), & Natalia Slioussar (HSE, Moscow, & Saint-Petersburg University)
slioussar@gmail.com

In this paper, we study grammatical gender of Russian expressive nouns (augmentatives and diminutives). Russian has three genders (M, F and N). The gender of the noun cannot be unambiguously determined from its inflectional affixes, but the majority of nouns ending in *-a/ja* in nominative singular are F, most nouns ending in a consonant are M, and almost all nouns ending in *-o/e* are N. Most previous studies (e.g. Vinogradov 1972; Corbett 1982) assume that expressive nouns preserve the gender of the base. Several authors (Savchuk 2011; Sitchinava 2011) have noted the existence of variation in colloquial Russian, but the extent of this variation and well as the factors that may influence it have not been studied before. Nevertheless, it is clear that the main trigger of the variation is the change of the inflectional (sub)class brought by the suffix, as in (1).

(1) *dom*_M 'house' → *dom-in-a*_{M/F} 'big house'; *kotleta*_F 'cutlet' → *kotlet-išč-e* 'big cutlet'

We conducted an experiment: participants received a list of 7 adjectives and then were presented with expressive nouns one by one. They were asked to pick a matching adjective and pronounce the resulting phrase. While they may have guessed that the experiment was about connotations of different expressive suffixes (adjectives had meanings like 'big', 'small', 'cool' etc.), our goal was to analyze gender agreement on the adjective. Participants were 30 native Russian speakers (17 women), aged 19–30. There were 50 stimuli (real and nonce expressive nouns) and 44 fillers. 3 factors were manipulated: the gender of the base word (M/F), the final segment of the derivate (consonant, *-a/ja*, *-o/e*), and animacy (nouns denoting people / animals / inanimate objects).

1200 answers were recorded. In 614 cases (51%) the genders of the base and the derivate were different. *All* changes were triggered by the final segment of the derivate, but the distribution of the assigned genders was uneven: 47% of consonant-final nouns, 15% nouns ending in *-a/ja* and 38% nouns ending in *-o/e* changed their gender to M, F and N, respectively. The differences were statistically significant (for all analyses, we used mixed effects logistic regression with random slopes and random intercepts by participants and by items). The base gender factor was also significant: 75% M nouns and only 25% F nouns preserved their gender. Animacy did not play a role.

Firstly, these results are noteworthy because gender change with expressive suffixes has not been studied previously. Secondly, they are interesting in the light of the discussion on markedness in the Russian gender system. In terms of frequency, $M < F < N$, but N is used as the grammatical default, e.g. in impersonal sentences. As a result, N is assumed to be unmarked in several formal models (Nevins, 2011; Kramer, 2015), while several experimental studies of noun-adjective agreement assume M to be default (Akhutina et al. 1999, 2001; Romanova, Gor 2017). Rice (2005) argues that on the word level, M is unmarked. Corbett & Fraser (2000) claim that on the phrase level, the default gender is N, and on the word level it is M. Slioussar & Malko (2016) who studied gender agreement attraction found that N behaved as unmarked in production, while M did so in comprehension. They proposed that apart from markedness, features have a property of stability, and M is the most stable gender. Our results are compatible with this suggestion.

Akhutina, T. et al. 2001. *Cortex* 37: 295–326. ◇ Akhutina, T. et al. 1999. *J. Psycholing. Res.*: 28, 695–713. ◇ Corbett, G.G. 1982. *Russ. Linguist.* 6: 197–232. ◇ Corbett, G.G., Fraser, N.M. 2000. Default genders. In: *Gender in Grammar and Cognition*. ◇ Kramer, R. 2015. *The Morphosyntax of Gender*. ◇ Nevins, A. 2011. *Linguist. Inq.* 42: 413–444. ◇ Rice, C. 2005. Optimizing Russian gender. In: *FASL 14*. ◇ Romanova, N., Gor, K. 2017. *Stud. Second Lang. Acq.* 39: 97–128. ◇ Savchuk, S.O. 2011. *Computer linguistics and intellectual technologies* 10: 562–579. ◇ Sitchinava, D.V. 2011. Časti reči (in Russian, 'Parts of speech'). URL:<http://rusgram.ru>. ◇ Slioussar, N., Malko, A. 2016. *Front. Psychol.* 7, article 1651. ◇ Vinogradov, V. V. 1972. *Russkij jazyk* (in Russian, 'Russian language').

Friday Poster.28

A CROSS-LINGUISTIC INVESTIGATION OF RESPONSE TIME DISTRIBUTIONS IN LEXICAL DECISION

Peter Hendrix (University of Tübingen)
peter.hendrix@gmail.com

The analysis of experimental linguistic data typically focuses on the mean of the response variable distribution. A least squares regression of the response times in a lexical decision experiment, for instance, estimates the conditional mean of the response time distribution given one or more lexical predictors, such as the length or frequency of a word. Effects of lexical predictors, however, need not be constant over the response time distribution. The effects of some predictors may primarily influence short reaction times, whereas the effects of other predictors may be more prominent for long reaction times. Furthermore, the qualitative nature of predictor effects may be different for different response times.

Distributional analyses provide more insight into the temporal dynamics of predictor effects in behavioral experiments. The (relative) timing of predictor effects is crucial for the development of psycholinguistic theories and models. Distributional analyses therefore have the potential to uncover valuable information about the nature of lexical processing that is not available through more traditional analysis techniques. Here, we present a cross-linguistic distributional analysis of lexical decision latencies that is based on the principles of time-to-event analysis.

Time-to-event analysis estimates the time at which an event of interest occurs. In lexical decision the event of interest is the “word or non-word” decision. Recently, time-to-event analysis techniques have been developed within the framework of generalized additive models (Hastie & Tibshirani, 1986; Wood 2006; Wood 2011). In particular, piece-wise exponential generalized additive mixed models (henceforth PAMMs; see Bender, Groll & Scheipl, 2018) make it possible to uncover non-linear predictor effects that vary as a function of time. The response variable in the PAMMs reported here is the instantaneous hazard rate: the probability that a “word or non-word” decision comes in at time t , given that no such decision came in prior to time t .

We carried out PAMM analyses of lexical decision latencies in four different languages: English (data from the English Lexicon Project (Balota et al., 2007) and the British Lexicon Project (Keuleers et al., 2012)), French (data from the French Lexicon Project (Ferrand et al., 2010)), Dutch (data from the Dutch Lexicon Project (Keuleers et al., 2010)), and Mandarin Chinese (data from MELD-SCH (Tsang et al., 2017)). The results of the PAMM analyses for the different languages showed a remarkable degree of convergence. Across languages, for instance, hazard rates were higher for high frequency words, primarily for the lower part of the response time distribution.

Two predictors revealed particularly interesting patterns of results that highlight the potential of PAMMs in the context of behavioral psycholinguistic data. First, we found an interesting temporal development of the effect of word length. For the shortest response times, the effect of word length was inhibitory in nature, with decreased probabilities of an instantaneous response for longer words. Later, however, the effect reversed, with higher hazard rates for longer words. This suggests that while the increased visual complexity of longer words initially results in additional processing costs, the increased information provided by these words makes them easier to respond to once visual information uptake has been completed (cf. Ramscar et al., 2014).

Second, across languages we found consistent semantic effects. For each language, we calculated semantic neighborhood density measures on the basis of cosine similarities extracted from distributional semantic models (Mikolov, 2013). As was the case for orthographic neighborhood density measures, the PAMM analyses revealed facilitatory effects of semantic neighborhood density, with an increased instantaneous probability of a response for words that live in denser semantic neighborhoods. The effects of semantic neighborhood density emerged early and were present from the moment responses started coming in. The current findings are consistent with the early semantic effects for compound processing observed by Marelli and Luzatti (2012) in an eye-tracking study and suggest that semantic properties of words influence lexical processing at an early stage.

Friday Poster.29

**THE IMPACT OF INFORMATION STRUCTURE ON LANGUAGE CHANGE: AN
EXPERIMENTAL STUDY**

Shira Tal (Hebrew University), Kenny Smith (University of Edinburgh), Jennifer Culbertson
(University of Edinburgh), Eitan Grossman (Hebrew University) & Inbal Arnon (Hebrew
University)
shira.tal1@mail.huji.ac.il

Many languages exhibit differential case marking (DCM), where marking targets only certain types of arguments. In such systems, atypical arguments (e.g., patients that are high on the animacy scale) are more likely to be marked. Recent work using artificial language learning experiments suggests that DCM is driven by a bias for efficient communication [1]. For example, in a language with variable word order, case marking only animate objects prevents ambiguity in an efficient way, since marking is restricted to events that are potentially ambiguous (e.g., events with two animate participants). However, an alternative account suggests that DCM instead reflects the pragmatics of discourse. Agents of transitive sentences tend to be discursively old (given) while patients tend to be new [2]. Atypical discourse associations (i.e., new agents or given objects) are often marked with additional linguistic material. For instance, in Catalan, only given objects are case marked [3]. Marking of atypical associations between argument structure and information structure—rather than ambiguity avoidance—has therefore been argued to be the main source of DCM [3]. Here we test whether discourse status drives case marking during learning of an artificial language.

As in [1], we taught participants (N=42) an artificial language over four days. The language had variable word order (50% SOV, 50% OSV) and optional case marking on objects (objects were case marked 50% of the time in both word orders). All sentences contained animate agents and patients. We manipulated information structure by presenting either the patient or the agent alone before each sentence. In the discourse-match condition, the agent of the sentence was presented in the preceding context trial (e.g., *fireman...the fireman hugged the mailman*); in the discourse mismatch condition it was the patient (e.g., *mailman...the fireman hugged the mailman*). A bias for marking unusual information structure predicts that participants will restructure the language to use case-marking more for given objects. Interestingly, information structure *did* impact participants' output: OSV was used more than SOV when the patient was given, (Fig. 1; $\beta=0.27\pm 0.09$, $p=0.005$). This mirrors a preference to put given information before new in natural language [4]. As in [1], we found that participants used case marking more for OSV sentences than SOV sentences ($\beta=0.61\pm 0.16$, $p<0.001$). However, participants did not use case marking more with given objects ($p>0.3$). Our results therefore suggest that learners' change their input in response to pragmatic considerations. In follow up work, we will explore whether case marking *is* differentially used to mark unexpected information structure when word order is fixed and therefore cannot be used to indicate information structure.

1. Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS* 109, 17897–17902.
2. Du Bois, J. (1987) The Discourse Basis of Ergativity. *Language* 63: 805-855.
3. Iemmolo, G. (2010) Topicality and differential object marking: evidence from Romance and beyond. *Studies in Language* (34:2) 239-272.5.
4. Arnold, et al. (2013). Information structure: Linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 403–413.

Friday Poster.30

A WORD OR TWO ABOUT NONWORDS

Peter Hendrix (University of Tübingen)
peter.hendrix@gmail.com

The perhaps most basic experimental task in psycholinguistic research is lexical decision. Participants are presented with a sequence of letters and are asked to decide if this sequence of letters is a real word or a nonword. Typically, the interest of researchers is in the response patterns for real words, and the responses to nonwords receive little attention. Here, we take a closer look at the lexical decision latencies for nonwords in a large-scale lexical decision database, the British Lexicon Project (henceforth BLP; Keuleers et al., 2012).

We analyzed the lexical decision latencies for 18,568 words and 10,000 nonwords in the BLP using piece-wise exponential generalized additive models (henceforth PAMs; cf. Bender, Groll and Scheipl, 2018). In the context of lexical decision data, PAMs offer the possibility to predict the time until a participant makes a decision about the lexical status of a letter sequence through instantaneous hazard rates. Instantaneous hazard rates describe the probability that a participant makes this decision at time t , provided that no decision was taken prior to time t . Through the use of PAMs, non-linearities can be modelled in both the predictor dimension and the time dimension. PAMs thus provide rich information about the temporal development of predictor effects.

In addition to a number of well-known predictor effects (i.e., the effects of word length and letter bigram frequency), we observed an interesting and novel frequency effect for nonwords. Traditionally, it is believed that a frequency effect for nonwords is, by definition, impossible, because nonwords do not have a frequency. Google searches in the English language for the nonwords in the BLP, however, revealed that the median Google frequency of the nonwords under investigation in the Google search index is no less than 25,600.

The Google frequency of nonwords proved highly predictive for lexical decision latencies. The PAM analyses revealed an overwhelming frequency effect not only for words, but also for nonwords. This effect was most prominent for the lower part of the response time distribution. Whereas the probability of an instantaneous response was higher for high frequency real words, it was lower for high frequency nonwords. Participants therefore found it harder to “reject” nonwords with a higher Google frequency.

Furthermore, we observed effects of both orthographic and semantic neighborhood density. We defined orthographic neighborhood density as the average Levenshtein distance between a nonword and its closest real word orthographic neighbors. Semantic representation for nonwords are not available through traditional methods. Recent advances in distributional semantics (Bojanowski et al., 2016), however, have made it possible to generate semantic vectors for words that are not in the input data and, by extension, for nonwords. On the basis of the semantic vectors for nonwords, we defined semantic neighborhood density as the average cosine similarity between a nonword and its closest real word semantic neighbors.

For words, we found facilitatory effects of both orthographic and semantic neighborhood density that were relatively stable over time. The greater the number of orthographic or semantic neighbors, the greater the hazard rates. For nonwords, we observed the opposite pattern of results. Hazard rates were lower for nonwords that live in denser orthographic and semantic neighborhoods. Both orthographic and semantic properties of real words thus have a significant impact on lexical processing for nonwords. To our knowledge, this study is the first to report a semantic neighborhood density effect for nonwords.

The results reported here shed new light on nonword processing. First, we observed robust effects with a similar temporal development for both word and nonwords for a number of well-established lexical predictors. Second, the PAM analysis revealed that it is harder to reject nonwords that are more word-like in terms of frequency, orthography, and semantics. The current results suggest that words and nonwords live in the same distributional space and that nonword reading is, to a large extent, guided by the same principles as word reading.

**LEARNING FROM SPEECH SOUNDS' PROBABILITY DISTRIBUTIONS IS
CONSTRAINED BY PRIOR LANGUAGE EXPERIENCE**

Kateřina Chládková (Czech Academy of Sciences & Charles University)
& Šárka Šimáčková (Palacký University)
chladkova@praha.psu.cas.cz

Distributional learning can be understood as inattentive tracking of stimulus probability distributions in the input and has been found to operate in various modalities. Some studies on speech sound learning suggest that listeners exposed to a bimodal distribution of speech sounds along a particular auditory dimension subsequently discriminate a novel contrast along that dimension better than listeners exposed to a unimodal distribution (e.g. Maye & Gerken 2001, *BUCLD 25 Proc.*); yet, several recent studies did not find such distributional training effects (Wanrooij et al. 2014, *PLoS ONE* 9; Wanrooij et al. 2015, *Proc. 18th ICPhS*).

What causes the varying outcomes across previous studies? While adults are sensitive to distributional statistics across modalities (Love, 2003, *Psychon. Bul. & Rev.* 10; Garrido et al., 2016, *npj Sci. of Learning* 1), whether or not this sensitivity results in the formation of new categories or contrasts (i.e. the traditionally expected effects of distributional training) may be subject to factors other than the input statistics, for instance, prior experience. Specifically for speech, adults, unlike infants, already have at least one speech sound system in place – that of their native language. This brings up the question of whether the native phonological system can affect how adults learn from the sounds' statistics. To this end, the present experiment compared distributional speech sound learning in participants with two different language backgrounds and tested whether the learning outcomes are affected by prior language experience.

We tested distributional learning of vowel categories contrasted by duration. Czech and Greek adults, whose native phonologies contain and lack, respectively, the phonological vowel length feature, were exposed to tokens of a novel vowel quality [ɑ] falling into a unimodal or bimodal distribution along the durational dimension. Listeners' sensitivity to duration differences was assessed before and after training to reveal whether and what the listeners learned from the distributional statistics. We hypothesized that if the ability to track probability distributions is affected by the native phonology in that learners can easily reuse their existing native-language contrasts or features, we would find the effects of distributional training to be larger along an already colonized dimension, i.e. in Czech listeners, as opposed to an unused dimension, i.e. in Greek listeners. If, on the other hand, the ability to track speech sounds' distributions is not modulated by the native phonology, the effects of training would be similar for Czech and Greek listeners.

A linear mixed model analysis detected a significant interaction of Language background, Test (pre, post), Training (uni-, bimodal), and Duration difference (identical pair, 15% difference, 30% diff., 45% diff.), $t = -2.205$, $p = 0.028$. The results showed that for the most difficult durational difference overall (15%, ~ JND for duration), two groups of participants improved their discrimination accuracy from pre- to post-test, namely, bimodally-trained Czechs and unimodally-trained Greeks. In conclusion, language background affected how listeners learn from exposure to distributional information. Czechs, who employ the trained auditory dimension to distinguish native vowels and for whom their existing higher-order representations for length contrasts can guide the bottom-up statistical learning, benefited from bimodal exposure and learned to more accurately differentiate between short and long instances of [ɑ]. Reversely, Greek listeners, unfamiliar with the trained dimension and having no mental representations to guide the learning, did not profit from the distributional statistics. Instead, after the unimodal exposure, they showed an unexpected training effect: they sensitized to within-category variation of the frequently presented stimuli.

Our findings suggest that in adults, exposure to sounds' probability distributions may not be sufficient for novel categories or contrasts to be formed unless the listeners have feedback from some already established mental representations. The requirement of this form of implicit feedback indicates that, at least in the domain of speech, distributional learning may not be available as a purely unsupervised mechanism throughout lifetime.

**EXPLICIT ENCODING AND FAST MAPPING OF NOVEL CONCRETE WORDS:
BEHAVIOURAL EVIDENCE OF EQUAL EFFICIENCY**

Alexander Kirsanov (St. Petersburg State University), Olga Shcherbakova (St. Petersburg State University), Elizaveta Nikiforova (St. Petersburg State University), Margarita Filippova (St. Petersburg State University), Evgeny Blagoveshchensky (St. Petersburg State University), Yury Shtyrov (Aarhus University, St. Petersburg State University)

The ability to acquire new words plays a defining role in our communicative behavior but the mechanisms underlying language acquisition still remain obscure. One unresolved question is the putative existence of two different learning mechanisms: a body of evidence suggests that while one learning strategy (so-called "explicit encoding", EE) relies on direct instruction and conscious learning effort, the other mechanism ("fast mapping", FM) is based on implicit knowledge acquisition from context by deduction or inference (Carey & Bartlett, 1978). Furthermore, it has been argued that they have different efficiency, with the FM route being suggested as being more rapid, ensuring a faster less effortful learning. However, this advantage remains arguable, with some studies failing to replicate it. Moreover, in most previous studies, the learning materials and the paradigms were not balanced between the two regimes, which may have confounded the learning effects.

To overcome these difficulties and scrutinize the EE-FM differences, we have designed a carefully balanced experimental design, where two learning protocols were maximally matched and carefully controlled. The learning task included images of novel items which were fully balanced for their basic visual features. These were combined with auditorily presented questions containing novel names of new objects, which were also made as similar as possible across the two conditions. In the EE condition, a participant was directly introduced to the target object (Here is LYN. Will you remember it?), whereas the FM condition required the use of contextual information (Does FER have ears? Answer 'yes' or 'no'). Ten new items were presented in each condition, each appearing only ten times using different instances of graphical objects and different auditory questions every time. All stimuli were psycholinguistically balanced and novel words were rotated across conditions and objects. Familiar real words with matching pictures as well as novel untrained pseudowords served as control conditions. The two types of trials were pseudorandomly mixed in a single training session. To test the outcomes, accuracies and reaction times were measured in three tests after the learning block: free recall, cued recognition and semantic word-picture matching.

All three tests confirmed accuracy and RT advantage for all familiar items in comparison to novel ones. The performance of free recall task was rather low, with participants recalling less than half of all items. Nevertheless, the accuracy of free recall for all trained stimuli was significantly higher than for untrained control stimuli (which also appeared ten times during the session but without any learning task). When cued, the recognition rate for all trained stimuli was also higher than for control items. In the semantic word-picture matching task the number of correct matches significantly exceeded the chance level. The application of a 2x2 factorial ANOVA (Novelty x Learning type) indicated a significant main effect of Novelty across the tasks. Finally, and importantly, no significant differences whatsoever were found between the behavioural outcomes of the two learning regimes.

The results suggest that we are able to quickly encode novel words even under rather stringent conditions, with 20 novel items being learnt within a short session containing only 10 unique presentations of each. Our data also indicate that when stimuli and learning regimes are balanced for their basic visual and auditory feature and the presentation mode, explicit encoding and fast mapping conditions produce similar behavioural outcomes. This is in line with some recent studies failing to find FM advantage in older individuals suggested in previous literature. The question still remains whether the brain mechanisms, which underpin this successful learning performance of the two learning strategies, may nevertheless diverge as suggested by some previous studies.

Supported by the RF Government grant contract № 14.W03.31.0010.

L2 SPEAKERS ARE NOT MORE RATIONAL THAN L1 SPEAKERS WHEN IT COMES TO LOSS AVERSION

Zoe Schlueter, Chris Cummins & Antonella Sorace (University of Edinburgh)
zoe.schlueter@ed.ac.uk

Recent research suggests that people are less prone to irrational decision-making in their second language (L2) than in their first language (L1) (Costa et al. 2014, Keysar et al. 2012). This “foreign language effect” has been attributed to lower emotional weight associated with the L2, reducing the impact of decision-making heuristics in L2 reasoning. However, the materials in these experiments are susceptible to multiple interpretations and a decision that is irrational under one interpretation is perfectly rational under another interpretation (as shown in monolinguals by Mandel 2013). Therefore, the difference between L1 and L2 speakers might be due to subtle differences between native and non-native interpretations affected by proficiency. We explore this in two experiments in which we investigate loss aversion bias in native Spanish speakers and we do not replicate the finding that L2 users in general are “more rational” in their decision making than L1 speakers. Instead, our results are consistent with the hypothesis that loss aversion bias depends on the interpretation of the materials and is affected by proficiency in the L2.

In both experiments presented here, we use two distinct versions of the classic Asian Disease Problem (ADP), one with the standard disease scenario and one with an economic scenario (following Costa et al. 2014, who found the “foreign language effect” in both). Participants are presented with a situation (e.g. 600 people will die from a disease) and have to choose between two options with equal expected value that are either framed in terms of **gain** (Medicine A: 200 people will be saved (*safe option*); Medicine B: 1/3 chance that 600 people will be saved and 2/3 chance that no one will be saved (*risky option*)) or in terms of **loss** (Medicine A: 400 people will die (*safe option*); Medicine B: 1/3 chance that no one will die and 2/3 chance that 600 people will die (*risky option*)). In previous research, native speakers were found to have a loss aversion bias, reflected in a higher proportion of safe choices in the gain compared to the loss frame (framing effect), but this effect was reduced in L2 users. However, although the options are supposed to be extensionally equivalent across frames, the optimal choice from a utilitarian perspective depends on whether the number term is interpreted as exact, upper-bounded or lower-bounded (Mandel 2013).

In Experiment 1, we find a significant framing effect not only for L1 speakers (N = 48) but also for L2 speakers (N = 47): logit mixed models show that both groups are more likely to choose the safe option when it is framed in terms of gain rather than loss (L1: $p = .04$; L2: $p < .01$). Importantly, there is no significant interaction between framing and language group ($p = .86$). However, the framing effect is stronger for more proficient L2 participants ($p < .05$), while emotional weight of the L2 (calculated from ratings of the emotional weight of swear/taboo words and terms of endearment) does not affect the framing effect ($p = .49$). Experiment 2 uses a novel manipulation in which the numbers are chosen to promote exact interpretations while still making the equivalence of expected values transparent (e.g. 633 instead of 600). Under an exact interpretation, the options really are extensionally equivalent across frames, which should reduce the framing effect and its interaction with L2 proficiency. Like in Experiment 1, L1 speakers (N = 46) and L2 speakers (N = 43) exhibit the same patterns (framing x language: $p = .55$). Although we find an overall framing effect ($p < .01$), this is driven by the economic scenario (frame x scenario: $p = .01$). Pairwise comparisons show that for both language groups the framing effect is significant for the economic scenario (L1: $p < .05$; L2: $p < .001$) but not the disease scenario (L1: $p = .77$; L2: $p = .98$), and neither L2 proficiency nor emotional weight affect the framing effect. We posit that the much stronger framing effect in the economic scenario is due to the materials (e.g. 633,000) admitting inexact interpretations while those in the disease scenario (e.g. 633) induce exact interpretations. However, participants’ self-reported interpretations vary in both scenarios, a point we aim to address in future work. Together, the data from these two experiments show that high proficiency L2 speakers behave like L1 speakers in respect to loss aversion bias, both with materials susceptible to pragmatic enrichment and with materials inducing exact interpretations.

AGREE TO DISAGREE: AGREEMENT ATTRACTION EFFECTS IN RESUMPTIVE PRONOUNS

Mor Ovadia, Maayan Keshev & Aya Meltzer-Asscher (Tel-Aviv University)
maayankeshev@mail.tau.ac.il

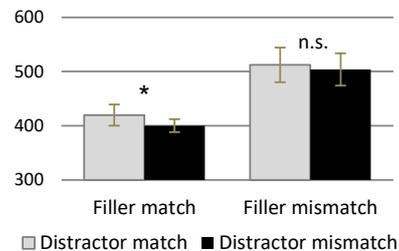
Agreement attraction effects, namely interference induced by a structurally irrelevant antecedent, are often reported in subject-verb dependencies [1, 2]. However, the occurrence of agreement attraction in pronoun-antecedent relations is debated. This study sets out to investigate agreement attraction effects in resumptive pronouns (RPs), pronouns which appear at the tail of unbounded dependencies, where a gap would otherwise appear. RPs can potentially constitute a strong retrieval cue for dependency formation, by indicating the filler's agreement. However, given the robustness of the active-filler strategy [3,4], it is not clear whether such retrieval is needed. In fact, if the pronoun initiates its own search, this may induce interference from other matching antecedents, and agreement attraction effects may arise. Alternatively, if the dependency is fully formed at the verb, and/or if the filler, being active, wins over any competing antecedent, these effects are not expected.

We conducted a self-paced reading experiment (34 participants, 32 experimental sets, 48 grammatical filler sentences) in Hebrew. In this language, RPs are obligatory in oblique positions. Materials were comprised of relative clauses resumed by obligatory RPs, of the structure in (1). We manipulated the gender features of the filler and the intervening NP (underlined in 1), forming four conditions: two grammatical (RP matching the filler and the distractor, or only the filler) and two ungrammatical conditions (RP matching only the distractor, or neither antecedent).

(1) *pitamu et ha-kupa'i(t) še-ha-menahel(et) hoda/hodeta*
we.fired acc the-cashier-M/F that-the-manager-M/F admitted-M/F
*še-ha-lakoxot hitloneno **ale'a** mispar peamim beavar.*
that-the-costumers complained **about.her** number of times before.

'We fired the cashier (M/F) that the manager (M/F) admitted that costumers complained about her a few of times before'

A mixed-effects regression revealed no effects on the RP. Results at the spillover region revealed a main effect for the match between filler and RP ($p = .0001$) such that grammatical sentences were read faster. The interaction was not significant ($p = .11$). However, pairwise comparisons revealed inhibitory interference (slowdown for matching distractors) in grammatical ($p = .02$), but not in ungrammatical sentences ($p = .45$).



The results suggest that the processing of RPs is (partly) vulnerable to interference from non-filler NPs. There are two possible explanations for this effect. First, it could be that the parser admits the option of regular pronominal binding for the RP, and thus considers the distractor as an antecedent. Alternatively, it could be suggested that (similarity-based) interference is established at the encoding of the NPs [5].

[1] Nicol et al. 1997. Subject-verb agreement processes in comprehension. *JML*, 36(4); [2] Wagers et al, 2009. Agreement attraction in comprehension: Representations and processes. *JML*, 61(2); [3] Sussman & Sedivy, 2003. The time-course of processing syntactic dependencies: Evidence from eye movements. *LCN*, 18(2); [4] Wagers & Phillips, 2014. Going the distance: memory and control processes in active dependency construction. *QJEP*, 67(7); [5] Villata et al. 2018. Encoding and Retrieval Interference in Sentence Comprehension: Evidence from Agreement. *Front. Psychol.* 9.

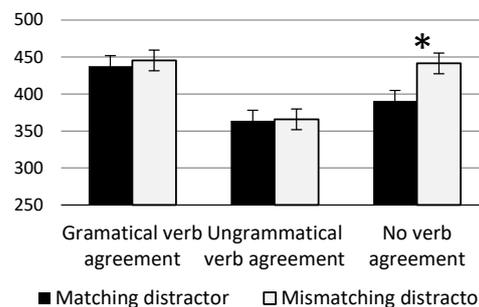
SUBJECT-VERB AGREEMENT AFFECTS THE PROCESSING OF A SUBSEQUENT REFLEXIVE PRONOUN

Maayan Keshev & Aya Meltzer-Asscher (Tel-Aviv University)
maayankeshev@mail.tau.ac.il

Can subject-verb agreement disrupt subsequent retrieval of the subject (e.g. by a reflexive pronoun)? Does the parser use verbal agreement as a shortcut in verifying subject-reflexive agreement? To investigate these questions, this study examines the effect of the verb's features on the processing of a subsequent reflexive, in a language which marks gender agreement on each of these elements. We conducted a self-paced reading experiment in Hebrew (72 native speakers; 30 sets, 60 grammatical fillers). We use the reflexive's sensitivity to agreement features carried by a structurally-irrelevant antecedent [1-2], as a marker for an attempt at retrieving the subject (using reflexives in picture NPs, which were recently found to induce agreement attraction in Hebrew [3]). The experiment therefore tests the distractor's effect on an ungrammatical reflexive. We manipulate the predicate preceding the anaphor to either match the subject (1a), match the reflexive (1b), or bear no agreement marking (1c).

- (1) a. *ha-soxen šel {ha-saxkanit/ha-saxkan} hixnis tmunot šel **acma** le-xadar ha-halbaša*
the-agent of {the-actress/the-actor} put.m pictures of **herself** to-room the-dressing
b. *ha-soxen šel {ha-saxkanit/ha-saxkan} hixnisa tmunot šel **acma** le-xadar ha-halbaša*
the-agent of {the-actress/the-actor} put.f pictures of **herself** to-room the-dressing
'the agent of the famous {actress/actor} has put pictures of herself into the dressing room'.
c. *la-soxen šel {ha-saxkanit/ha-saxkan} yeš tmunot šel **acma** be-xadar ha-halbaša*
to.the-agent of {the-actress/the-actor} BE pictures of **herself** in-room the-dressing
'the agent of the famous {actress/actor} has pictures of herself in the dressing room'.

No significant effects were observed on the pronoun. Results at the spillover region revealed that when the predicate did not bear agreement cues (1c), attraction was observed: RTs following the mismatching reflexive were faster when a structurally-irrelevant antecedent agreed with it ($p = .03$). When the verb matched the subject (1a), a significant attraction was not observed ($p = .15$), but the interaction with the "no verb" conditions (1c) was also non-significant ($p = .17$). Finally, when the verb matched the reflexive (1b), the distractor's agreement (i.e whether it matched the reflexive and the verb) did not affect the pronoun's processing ($p = .81$; interaction with 1c: $p = 0.03$). Moreover, RTs in the mismatching version of this case were faster than other mismatching conditions (both $p < .001$).



The results suggest that agreement marking on the verb can block agreement attraction effects. There are two possible ways in which this influence can arise. First, it could be that the parser first checks reflexive-verb agreement and retrieves the subject only when such cues are unavailable (resulting in agreement attraction in 1c). Under this account, when the verb bears phi-features, the reflexive's processing reflects mismatch (in 1a) or match (in 1b) detection (regardless of the subject's/distractor's features). Alternatively, it could be that when the verb mismatches the subject, the subject's representation is altered to agree with it. This leads to facilitation in the processing of an ungrammatical reflexive following an ungrammatical verb (regardless of the distractor). We find the results in the grammatical verb conditions inconclusive, due to the null effects, and thus currently cannot decide between the two options.

[1] Parker & Phillips, (2017). Reflexive attraction in comprehension is selective. JML 94. [2] Sloggett & Dillon, (2018). Person blocking in reflexive processing: when "I" matter more than "them". Talk at CUNY 31. [3] Keshev, et al., (2018). Objects and self-portraits: Contrasts in the processing of reflexive pronouns across languages and structures. Poster at CUNY 31.

AWARENESS OF LINGUISTIC COMPETENCE INFLUENCES STRUCTURAL PRIMING

Christina S. Kim & Gloria Chamorro (University of Kent)

c.s.kim@kent.ac.uk

Structural priming (Bock 1986; Pickering & Branigan 1998) has been useful for probing abstract syntactic representations due to its implicit nature (Bock & Griffin 2000), which has led some researchers to liken it to implicit procedural learning (Chang et al 2006). We also know from cross-linguistic priming in bilinguals that structural priming targets representations that are not language-specific (Hartsuiker et al 2004; Bernolet et al 2007). The current study asks whether such implicit effects are modulated by higher-level awareness about a speaker’s linguistic competence. We focus on conversations between L1 and L2 speakers, where the latter has incomplete linguistic competence, and may be influenced by their L1.

Participants played a picture-matching game which involved taking turns with another speaker to describe scenes depicting ditransitive events (e.g. “Harry showed Hermione the painting”) using the verb provided. Participants heard a recorded voice and were told to make their descriptions to be maximally clear to the speaker in the recording, who would have to perform the matching task using the participant’s recorded descriptions. Speaker type (native English speaker/NS, non-native Spanish-accented speaker/NNS) was manipulated between subjects. For two groups, the speaker stayed the same across two experimental blocks. For two additional groups, the speaker changed; this was in order to assess whether priming from one speaker transferred to dialogue with another speaker. In one case, the NNS in block 1 was followed by a different Spanish-accented speaker in block 2. In the other, the NNS was followed by a NS. The four conditions are shown below:

	NS-no change	NNS-no change	NNS-speaker change	NNS-lang change
Block1	NS1	NNS1	NNS1	NNS1
Block2	NS1	NNS1	NNS2	NS1

All of the recordings used double object (DO) structures. Half were acceptable in English; half used verbs that do not participate in the dative alternation in English, creating anomalous sentences (e.g. “Hermione described Ron the monument”).

Responses were coded as DO, PD (prepositional dative), or other. Unaggregated responses from blocks 1 and 2 were fitted with separate mixed-effects regression models predicting DO responses, with subject and item included as random effects. The block 1 model included verb type (alternating, non-alternating), trial number, and speaker type (NS, NNS), and all two-way interactions as predictors. Fixed effects were removed from the model using stepwise model comparison if they did not improve model fit or were collinear with other model terms. Unsurprisingly, alternating verbs were more effective primes than non-alternating ones ($\beta=.15, p<.05$). However, this advantage was stronger for NNS than NS ($\beta=.31, p<.0001$); in other words, participants produced more anomalous DO structures when interacting with a NS who also used those structures (see also Ivanova et al 2012).

The block 2 model included verb type, trial number, speaker type, speaker change (whether block 2 speaker was new), and verb overlap (whether the verb appeared in block 1). There was a verb type-speaker change interaction, with a greater advantage for alternating verbs when the speaker had changed ($\beta=.12, p<.0001$); this suggests that priming of atypical verb usage did not carry over to a new speaker. There was a marginal interaction of verb overlap and speaker type ($\beta=.036, p=.08$): for NNS, verbs used in block 1 were better primes than new verbs. Our findings suggest that the awareness that non-native speakers lack full linguistic competence may limit how much their usage of atypical structures affects a native speaker’s own usage. To the extent that we see a lexical boost effect (Cleland & Pickering 2003) with non-alternating verbs (old verbs were numerically better primes than new ones in all conditions), it resembles the improved acceptability observed after repeated exposure to ungrammatical sentences (e.g. Luka & Barsalou 2005).

Friday Poster.37

CAN L2 SPEAKERS ACQUIRE NEW MORPHOLOGICAL DISTINCTIONS? EVIDENCE FROM TEMPORAL MORPHOLOGICAL PRODUCTION IN MANDARIN SPEAKERS OF ENGLISH

Qingyuan Gardner, Holly Branigan & Vicky Chondrogianni (University of Edinburgh)

qgardner@ed.ac.uk

Adult second language (L2) learners produce inflectional markings inconsistently, especially if their native language (L1) does not use inflectional morphology, e.g. Mandarin Chinese (Lardiere, 1998). We report an experiment that set out to discriminate between 3 possible sources for inflectional errors in L2 production. First, under the assumption that speakers only process abstract information relevant to the language produced at the level of conceptualisation (Levelt, 1989), the L1 conceptualisation mechanism may not process L2 relevant information. Second, at the level of encoding and retrieval, newly acquired L2 morphological representations may be inconsistently retrieved due to insufficient activation; such retrieval might further be affected by morphological complexity. Finally, inflectional errors in spoken production might be susceptible to articulatory difficulties (e.g., difficulty in articulating –s and –ed inflections). Evidence from native Mandarin speakers' spoken production of English temporal morphology argues against a conceptualisation account but is less informative with respect to discriminating encoding / retrieval vs articulatory accounts (Gardner, 2017). Therefore, in this experiment, we investigated native Mandarin speakers' production of English temporal morphology in the written modality. Written production should be susceptible to encoding / retrieval difficulties but not articulatory difficulties.

We repeated the scene description task from Gardner (2017) in the written modality to investigate inconsistent inflectional production amongst adult Mandarin learners of English (N=48; IELTS 6.5+, AoA 5+). Participants produced typed descriptions of scenes presented on a computer. We used visual stimuli that included temporal cues (e.g., *Every day, Last week*), singular and plural subjects, and specified transitive verbs to elicit temporal adverbials and the appropriate verb inflections (present habitual 3rd person singular –s, past –ed). Hence, temporal context was manipulated alongside subject number.

Strikingly, similar to previous findings for spoken production, Mandarin speakers of English were sensitive to temporal cues in written production, depending on subject number (Accuracy: Present Habitual: singular 50.00% plural 70.01%, Past: singular 73.41% plural 70.67%; $p < .001$). Importantly, although the overall error rate was significantly lower in written production than previously found in spoken production (written: 34.37% vs spoken: 46.88%; $p < .001$), participants' error patterns did not change across modalities. Similar to spoken production, the complex inflection 3rd person singular –s (encoding person, number and tense information) was also more likely to be omitted in written production than the less complex inflection past –ed (encoding only tense information).

These findings show that L2 speakers were sensitive to L2 temporal cues in production given subject number information, providing further evidence against a conceptualisation account. The fact that the written production gave rise to significantly fewer errors than the previous spoken production but did not affect error patterns indicated articulation contributed to overall error rate but not to error type. The fact that the featurally more complex inflection was especially prone to inflectional omission even without articulation argues against a purely articulatory account and supports the idea that complex inflectional representations may not receive sufficient activation during morphological retrieval. Overall, our findings point towards an important processing locus for erroneous L2 inflectional production, irrespective of production modality.

STICKY LABELS: FORMULATION VS RECONCEPTUALISATION EFFORT IN THE USE OF REFERENTIAL PRECEDENTS

Lucia Castillo, Kenny Smith, & Holly P. Branigan (University of Edinburgh)
lucia.castillo@ed.ac.uk

In dialogue, speakers usually converge in the use of the same labels for referents. Brennan and Clark (1996) showed that a speaker will maintain a mutually known label if they keep on speaking to the same partner, even if this label is overspecified in context. In their experiment, participants described a set of images to a partner. In the first part of the experiment there was more than one image from each category, creating the need for specific labels. In the second part, each target image was unique in its category. Speakers who kept the same partner in the second part maintained their overspecified labels (e.g., “the fluffy white dog”), whereas speakers with new partners switched to context-appropriate basic labels (e.g., “the dog”). Brennan and Clark interpreted their results in terms of *conceptual pacts*: tacit agreements between interlocutors to conceptualise referents in particular ways. Other studies have suggested that maintaining labels (or ‘precedents’) might be the outcome of egocentric processes: Although using basic-level labels reduces formulation effort compared to using more specific, longer labels, the reconceptualization process involved in switching to a basic-level label carries a cognitive cost for the speaker.

In two experiments, we investigated the conditions under which speakers would abandon a precedent for a context-appropriate alternative. Using a setup similar to Brennan and Clark’s, we compared participants’ referential choices when playing the second part with a) the Same Partner as in the first part, b) an Overhearer, who had witnessed their interaction in the first part without participating, or c) a New Partner. Exp. 1 used a small image set (12 coloured and monochrome images). In the first part, participants played 3 rounds with 6 colour pictures of objects (2 targets, each with 2 distractors from the same category) and 6 black tangram figures. In the second part, they played 3 rounds with 6 pictures (2 targets + 4 unrelated items) and 6 tangram figures. Participants did not reduce their descriptions to basic-level terms in any condition, instead maintaining their overspecified labels even with a New Partner. This suggests that speakers used an egocentric strategy that did not take into account their partner’s knowledge. We suggest that in this context, where the number of descriptions a speaker needed to produce was low, the reduction in formulation effort of using basic-level labels was outweighed by the reconceptualisation effort of switching from the original (overspecified) labels.

Exp. 2 used a large image set (20 monochrome images; 4 targets/round). In this context, participants’ use of overspecified precedents varied across conditions: They maintained the same overspecified labels if talking to the Same Partner or an Overhearer, but not if talking to a New Partner (66% SP, 75% O, 25% NP, $p < .001$). Participants in the New Partner condition used significantly more basic level terms in the first round of the second part (i.e., round 4; 27% SP, 17% O, 65% NP, $p = .0072$), and reduced the length of their descriptions through rounds 4 to 6 significantly more than participants with the Same Partner ($p = 0.0173$). Round number was also a significant predictor of the use of more basic-level labels and reduced descriptions, across conditions ($p = .001$). Participants in the Overhearer condition behaved similarly to participants in the Same Partner condition, in all of our measures. These results suggest that in a context where speakers had to produce a large number of labels, the reduction in formulation effort of using basic-level labels outweighed the reconceptualisation effort of switching from the original labels to new ones. Overall, our results point to a conservative tendency in the processing of references: Once a precedent has been established, speakers will not switch to a more context-appropriate alternative unless the processing savings are higher than the cost of switching. However, if the precedent is shared with a partner (even if they didn’t participate in its original establishment), the cost of switching is increased as a consequence of an historical discourse record having been established. Our results suggest that when making referential choices, speakers combine both sources of information: their own processing effort, and the history of interaction with a partner.

**WATCH YOUR WORDS:
THE EFFECTS OF VISUAL MANIPULATIONS IN EVENT SCENES ON LANGUAGE
PRODUCTION**

Yulia Esaulova, Sarah Dolscheid, & Martina Penke (yulia.esaulova@uni-koeln.de)

The present study examines visual and conceptual aspects of depicted event scenes as potential sources for systematic variations in scene descriptions. There is some evidence that when speakers' visual attention is directed to the patient and not the agent (e.g., via cueing), they are more likely to produce passive voice sentences (e.g., Gleitmann, January, Nappa & Trueswell, 2007). At the same time, there seems to be a preference for spatial representations of events, such that agents are typically positioned to the left of patients (e.g., Chatterjee, Southwood & Basilico, 1999). This study investigates whether and how these two visual properties may interact affecting language production. In addition, some voice and word order variations have been explained by prominence effects of animacy, where animate entities are more likely to be assigned subject/agent roles as more prominent than inanimate ones (e.g., Lamers & de Swart, 2012). Whether these prominence effects may be modulated by visual manipulations, such as positioning and cueing of the patient, was another question addressed in this study.

Native speakers of German ($N = 44$, mean age = 23.43 years, $SD = 3.01$) were tested in a picture description task while seated in front of a computer screen with an eye-tracker. The pictures depicted scenes where an animate agent performed an action on either an animate or an inanimate patient. Participants were instructed to describe each picture using one sentence. Patients were situated to the right or to the left of agents (Figure 1) and half of them were preceded by a short visual cue (see Figure 2).

The results show that scenes with left- rather than right-positioned patients lead to longer speech onset times ($F(1, 43) = 6.46, p = .015$) and a higher number of passive sentences ($F(1, 43) = 5.48, p = .024$). In addition, passive utterances occurred more often for scenes with animate rather than inanimate patients ($F(1, 43) = 8.41, p = .006$). Cueing of patients did not have an effect on either speech onset times or the number of passives, however, the analyses of eye-tracking patterns revealed more initial saccades to patients after they were cued than when they were not ($t(43) = 4.83, p < .001$).

Our findings demonstrate that visual and conceptual properties of event scenes influence different aspects of language behavior. Both the initiation of utterances and the voice selection were mostly affected by the positioning of patients in event scenes. Possible processes underlying left-agent preferences may involve an alignment between the linear representation of thematic roles and the serial planning of speech, thus relating to agent-first preferences (e.g., Jackendoff, 2002). Moreover, voice selection was sensitive to the animacy of thematic roles, so that more passive utterances were produced for scenes where both arguments were animate. This is in line with prominence theories suggesting that animate entities are more likely to be realized as subjects in sentence-initial positions than inanimate ones. The discussion of findings integrates cognitive and linguistic models relating differences in linguistic output to attention and prominence effects.

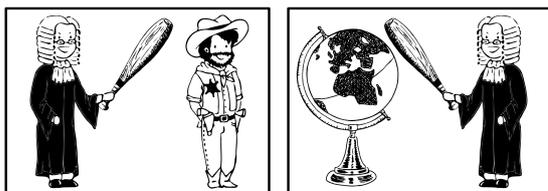


Figure 1.

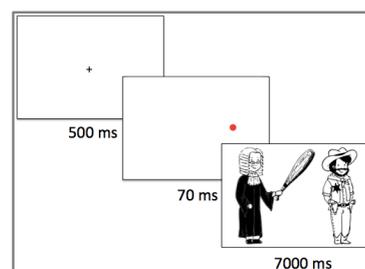


Figure 2.

Friday Poster.40

STAIRS4WORDS: A NEW ADAPTIVE TEST FOR ASSESSING RECEPTIVE VOCABULARY SIZE IN ENGLISH, DUTCH AND GERMAN

Florian Hintz, Suzanne R. Jongman, Marjolijn Dijkhuis, Vera van 't Hoff (all MPI for Psycholinguistics), Markus Damian (University of Bristol), Sascha Schroeder (MPI for Human Development, Georg-August-Universität Göttingen), Marc Brysbaert (Ghent University), James M. McQueen (Radboud University), Antje S. Meyer (MPI for Psycholinguistics)
Florian.Hintz@mpi.nl

Studying language in scientific, clinical or educational settings often requires a precise and reliable estimate of the language users' vocabulary size. However, current vocabulary tests are insufficient for various reasons. For example, most tests employ a fixed order of items, where item difficulty increases linearly and all participants respond to all items. This is often too hard for low-scoring participants or too easy for high-scoring participants. Moreover, many tests involve abilities not directly related to people's knowledge of words (e.g., picture recognition, sentence comprehension). Lastly, item difficulty is often approximated using word frequency, with infrequent words assumed to be more difficult than frequent words. This is problematic because some words that occur rarely in language corpora (e.g. paint brush) are often well known to the majority of the population.

Here, we intend to overcome these problems by developing *Stairs4Words*—a new, quick and intuitive test for assessing individuals' receptive vocabulary size. The test will run online and will be available in three languages: English, Dutch and German. Rather than relying on a fixed item structure, *Stairs4Words* is adaptive, in that item difficulty is tailored to the participant's performance level by implementing a staircase procedure as is often used in psychometric testing. We aim to minimize the involvement of skills unrelated to word knowledge by using a simple task, asking participants to indicate whether or not they know a given target word. Finally, item difficulty is approximated using word prevalence rather than word frequency. Word prevalence refers to the degree to which a word is known by a representative sample of the population and is computed from responses to collections of words representative of the English, Dutch and German language, respectively. Target words were chosen to be equally well known by males and females and individuals from different age ranges. Target words were then grouped in prevalence bands, ascending in difficulty. The test features two tracks: a 'Fast track' and a 'Fine tuning' track. The tracks operate at the back-end and are not noticeable to the participant. Each participant starts on the 'Fast track'. The first target word (presented in written form) is sampled from the prevalence band containing words known to approximately 99% of the population. When the participant indicates they know this word, the following word is sampled from the next, more difficult prevalence band. Non-words are used to prevent yes-biases. After two consecutive 'No' or incorrect responses, they leave the 'Fast track' and enter the 'Fine Tuning'. While the 'Fast track' is aimed at getting the participant to the vicinity of their end score, 'Fine Tuning' is meant to determine their end score by applying the staircase procedure. Specifically, the participant will move to a more difficult band after having responded correctly to four test words (3 existing, 1 non-existing) in a row. In contrast, they immediately move to an easier band after a single error. The test terminates after three consecutive cycles between two neighboring bands. The participant's score is the last band where an entire quadruple was responded to correctly.

We have started piloting the Dutch version of *Stairs4Words* in students with diverse educational backgrounds, recruited from universities and vocational colleges. The 193 participants carried out two subsequent rounds of *Stairs4Words* (using different words in both rounds) and then did the Peabody picture vocabulary test for reference. The correlation between participants' *Stairs4Words* scores in both rounds was $r = .74$. The correlation between the average of both scores and their Peabody scores was $r = .52$. In a related project, we explored whether participants benefit from a bimodal presentation, i.e., auditory and visual presentation, of the target words, as compared to a uni-modal presentation. To that end, we presented university and vocational college students with words from different prevalence bands in three conditions: auditory-only, visual-only and bimodal. The results suggest that overall performance is worst for auditory-only presentation and that there is no clear advantage for bimodal over visual target word presentation. We conclude that visual presentation of the target words is sufficient.

**INDIVIDUAL PARSING STRATEGIES IN COMPLEX VERB-FINAL STRUCTURES.
EVIDENCE FROM MEMORY INTERFERENCE**

Katja Suckow (University of Göttingen) & Jana Häussler (Leipzig University)
katja.suckow@phil.uni-goettingen.de

Multiple center-embedded sentences lacking one of the verbs can be perceived as grammatical. So far, individual differences in susceptibility to this illusion have been neglected. Instead, the debate focused on the underlying mechanisms of this missing-VP effect. Memory accounts attribute this effect either to structural forgetting [1], or erroneous attachment as a result of memory interference [2]. We have conducted a study in German combining the online eye-tracking procedure used in [3] and the sentence design in [2]. Participants (N=48) read sentences with stacked relatives clauses (N=30, plus fillers) and had to answer a post-experiment questionnaire to give feedback. Complete sentences started with a matrix clause followed by a cascade of three verb-final clauses and finished with an adverbial clause:

[matrix clause [$CP_1 \dots [CP_2 \dots [CP_3 \dots VP_3] VP_2] VP_1$] adv clause]

- (1) Es wurde öffentlich, [dass der Kellner/die Kellner, [den/die ausgerechnet der
it became public that the waiter/waiters who just the
Manager, [bei dem das Geld gefunden wurde VP_3 ,] (beleidigt hat) VP_2], (geklagt
manager at who the money found was (insulted has) (litigated
hat/haben), VP_1] nachdem ein Zeuge aufgetaucht war.
has/have) after a witness appeared was.
'It became public that the waiter who only that manager with whom the money was found has insulted has litigated after a witness had appeared.'

Incomplete sentences lacked either VP_2 or VP_1 . Memory accounts claim that the prediction of VP_2 will either be forgotten, [1], or the VP_2 will be integrated high with the wrong subject, [2]. To test this, we varied the number specification of the highest subject (NP1) and the corresponding VP_1 . Under the Structural-Forgetting Hypothesis [1], there should be no number effect; under the High-Attachment-Hypothesis, high attachment of VP_1 should be facilitated by the plural agreement between VP_1 and NP1 making it harder to notice a missing VP_2 .

Reading times of the adverbial clause increased in all ungrammatical conditions. Further comparisons showed that the missing VP_2 condition was harder to detect than the VP_1 condition, indicating a missing VP_2 -effect. In the post-experiment questionnaire, 25 out of 48 participants stated that some sentences were ungrammatical because a part of the sentence was missing. While there was no global main effect of Number, the group that detected the ungrammaticality (thorough readers) showed susceptibility to the number manipulation (the others showed no number effect). The reading times of the thorough readers for the grammatical and the missing- VP_2 sentences were faster in plural conditions compared to the singulars: the plural facilitated the identification of the attachment site. However, this also means that for thorough readers the fast integration of the plural VP_1 with a plural subject can hide that the VP_2 is missing. The plural VP_1 occurs in the linear position of the missing VP_2 , however, because of the plural number marking the VP_1 is likely to be attached high to the plural subject. This would be in agreement with the High-Attachment-Hypothesis in [2], but also other interference accounts of grammaticality illusions involving number agreement, such as intrusion, cf. [4].

[1] Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3), 225–248. [2] Häussler, J., & Bader, M. (2015). An interference account of the missing-VP effect. *Frontiers in Psychology*, 6(766). [3] Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures. *Language and Cognitive Processes*, 25(4), 533–567. [4] Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237.

Friday Poster.42

THE VALIDITY OF THE LEXICAL DECISION TASK FOR BEGINNING READERS

Pauline Schröter (Institute for Educational Quality Improvement (IQB), Humboldt-Universität zu Berlin) & Magdalena Bartelt (Humboldt-Universität zu Berlin)
pauline.schroeter@iqb.hu-berlin.de

One of the most frequently used tasks in research on visual word recognition is lexical decision (LD). By asking participants to decide whether a visually presented letter string is an existing word or not, the underlying assumption is that in order to fulfil this task participants need to check whether it corresponds to an existing lexical representation. Given that in natural language processing nearly all words are recognized, researchers' primary interest is in the speed of lexical access rather than in the precision of the process [1]. Giving priority to analyzing reaction times (RTs) over decision accuracy, however, neglects the impact of the task's decision-making component. Particularly in comparison to tasks that measure lexical access by purely naming words, as in the Naming task, the validity of LD as an index for word recognition instead of word comprehension has been questioned [2]. Especially for research on beginning readers, who have a smaller orthographic lexicon and are more likely to access their larger phonological lexicon to make a decision [3], unobserved costs related to the activation of a word's semantic representation might skew the interpretation of accuracy scores and RTs. For instance, words might be conceived as nonwords not because they were incorrectly decoded, but because their semantic representation was too weak to be retrieved from the mental lexicon in time for the activation threshold to be reached. On the same account, RTs could reflect to a much greater extent the time needed for semantic feedback than for orthographic or phonological access.

The aim of the present study was to investigate the validity of the LD task to measure visual word recognition in children at the very beginning of reading acquisition. Within the frame of the Developmental Lexicon Project [4], 140 German speaking first-graders were audio recorded while performing a standard LD task. Though all participants were instructed to execute the task silently, a large number of them phonologically recoded the stimuli audibly, essentially approaching the LD task as a Naming task. By manually measuring their vocal responses on a trial-by-trial basis, we were able to couple simultaneously collected LD and Naming data from 52 children. Comparing accuracy scores from the two tasks, results show high congruence between LD scores and Naming scores when stimuli were correctly recognized or rejected as words. However, in case of incorrectly rejected words or accepted nonwords, Naming scores reveal that more than 50% of the stimuli were actually read aloud correctly. Differences between LD RTs and Naming offsets on accurate trials demonstrate that 85% of the LD RTs are taken up by the decision-making process. Consistent with studies on the impact of linguistic characteristics on orthographic processing in children, the magnitude of these differences is affected by word frequency and nonword length. Findings point to the crucial role of semantic feedback for LD in children and suggest that in contrast to skilled readers LD results for beginning readers should be interpreted more as a measure of word comprehension rather than of mere word recognition.

[1] Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology*, 3.

[2] Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580.

[3] Van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2012). Lexical decision in children: Sublexical processing or lexical search? *Quarterly Journal of Experimental Psychology*, 65(6), 1214–1228.

[4] Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods*, 49(6), 2183–2203.

CO-SPEECH GESTURES AND CO-PRESENT OBJECTS: ATTENTIONAL ALLOCATION IN REFERENTIAL COMMUNICATION

Raheleh Saryazdi, Craig G. Chambers (University of Toronto)
 raheleh.saryazdi@mail.utoronto.ca

During situated language use, co-speech gestures can enhance listeners' comprehension, which highlights the multimodal nature of communication. But how efficiently can listeners use manual gesture cues when the context requires them to also deploy visual attention to objects located elsewhere in the visual field? Although questions of this type have been explored in the context of deaf signers (where attention to manually-conveyed information is essential, see MacDonald et al., 2018, *DevSci*), little is known about the uptake or use of gesture cues when they occur in a scene region that is near but not overlapping with the region where potential referents are located. In the present study, we examined how “grasp” gestures — a type of iconic gesture typically produced prior to noun onset in utterances such as “Pick up the the candy” — affect real-time referential processing when the object being referred to is also in the listeners' visual field. These grasp gestures, which reflect volumetric properties (e.g., size) should in principle facilitate referential identification because they narrow the set of possible referential candidates. However, listeners' spontaneous tendency to fixate candidate referents may result in little or no attention directed to co-speech gestures, reducing or eliminating the effect of these cues.

We conducted two eye-tracking experiments in which participants watched a video recording of a speaker providing “pick up” instructions referring to objects that were visible to a listener, but whose location was hidden from view of the speaker by a low barrier (making deictic gestures inappropriate). Participants were informed that the speaker was nonetheless familiar with all objects and that she would be reading aloud instructions about which object to pick up on a given trial. Videos were presented on a large-format TV screen and were originally recorded from the perspective of the eventual participant. On half of the critical trials, the speaker produced a grasp gesture that reflected the size/shape of the target object (e.g., candy), which shared onset sounds with a competitor (e.g., candle). Two unrelated objects were also present (see figure). We also manipulated whether the target object was smaller or larger than the other objects (and the corresponding gesture was changed accordingly). The gesture began naturally as “pick up” was uttered, providing a cue in advance of the critical noun. No gesture was produced on the remaining half of critical trials. The participants' task was to click on the denoted object. Filler trials and other features of the task were included to disguise the experimental purpose. We measured listeners' eye movements as the target instruction unfolded. Results from Experiment 1 revealed that, despite an overall low incidence of direct fixations to gestures, listeners were reliably better at differentiating targets from competitors in trials in which gestures were present. This suggests gesture cues were taken up peripherally, in turn providing some benefit for real-time referential identification. The effect, however, was relatively subtle and was most evident when the target was smaller than the competitor. This may be because the correspondingly small grasp aperture is perceived as ruling out a greater number of objects in the scene. In Experiment 2, we added speech-spectrum background noise to examine whether adverse listening environments would boost listeners' attention to gesture cues. Interestingly, the presence of background noise actually reduced listeners' use of gesture cues. This may be because attentional resources had been largely reallocated to the task of listening to a challenging auditory stream. Regardless, this outcome indicates that gesture cues explored here are not re-weighted in a strategic way, perhaps because their utility is never consciously recognized from the outset.

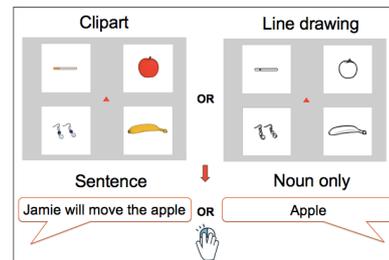


**FACTORS INFLUENCING SEMANTIC COMPETITION
DURING REAL-TIME LANGUAGE PROCESSING**

Raheleh Saryazdi, Craig G. Chambers (University of Toronto)
raheleh.saryazdi@mail.utoronto.ca

The on-line processing of a spoken word is known to activate semantic associates (*hat*→SCARF). In visual world studies, this can be observed via incidental fixations to certain display objects. The magnitude of co-activation effects reflects features of semantic memory, such as the strength of the semantic association in question. In the present study, we explore the influence of other factors beyond semantic relationship themselves. In particular, we study the impact of selected linguistic and visual attributes in visual world contexts, such as the presence/absence of a carrier phrase, word frequency, and the type of visual stimuli. What effect might the presence or absence of sentence context, even a semantically neutral one, have on associate activation? One relevant observation is that English content words occur in multiple grammatical roles, and their specific meanings are narrowed by these roles (e.g., *tie*: N/V, *orange*: N/Adj). Thus, activation patterns found with isolated words may be somewhat unconstrained, whereas in sentences the more restricted meanings imposed by grammatical role information (e.g., noun slot) may entail weaker activation patterns. Further, although frequency is known to affect the processing of target words and phonological neighbours, its effect on the activation of semantic associates is often not considered. E.g., the slower processing of low frequency (LF) target names may allow more time for the eyes to be attracted to semantically associated objects in the visual display. Finally, differences may also arise in relation to type of visual stimuli, such as the black-and-white (BW) line drawings vs. coloured clipart images that are commonly employed in visual world paradigms. Past work in cognitive development and visual cognition has reported differences related to these image types (e.g., Rossion & Pourtois, 2004, *Perception*; Simcock & DeLoache, 2006, *DevPsych*). Drawing on this work, we note that BW line drawings seem more likely to evoke abstract conceptual kinds, whereas clipart images more strongly evoke the idea of specific exemplars due to token-specific surface level detail (colour, texture). Because associate activation effects involve abstract conceptual relations, one possibility is that weaker associate activation effects may occur with clipart images.

Eighty participants were assigned to either the colour clipart or matched BW line drawing conditions (see figure). In two separate blocks, participants heard the target object either within a declarative sentence “*Jamie will move the apple/banana*” (sentence) or on its own “*apple/banana*” (noun only), where the interval between display onset and noun onset was equated in each case. The objects serving as targets vs. category-based semantic competitors were alternated across experimental ‘lists’ (apple/ banana), and target objects’ names varied in frequency. Growth



curve analyses revealed the effects on semantic competitor co-activation as the target word unfolded in time. These analyses showed differential effects of frequency across the carrier phrase manipulation, with greater activation with LF words in the noun-only context. This pattern further interacted with image type, such that the strongest co-activation differences for line drawings vs. clipart (with more activation with line drawings) occurred in the sentence condition with LF words, showing the complex way in which these factors have a combined influence. The present study has both methodological and theoretical implications. On the methodological side, researchers can benefit from the knowledge that certain characteristics of visual and auditory stimuli may boost or reduce the effects of semantic co-activation within or across studies. Thus, these factors should be considered while designing visual world studies of semantic activation. On the theoretical side, the current results suggest a link between the temporal dynamics of lexical processing and the strength of incidental semantic activation. Future studies could explore this link further by including stimuli with greater differences in word frequency or visual attributes (e.g., less optimal orientation).

GERMAN CHILDREN'S PROCESSING OF NON-CANONICAL WORD ORDER SENTENCES: THE ROLE OF TEMPORAL AMBIGUITY

Chiara Boila (University of Potsdam), Tom Fritzsche (University of Potsdam) & Barbara Höhle (University of Potsdam)
 chiara.boila@uni-potsdam.de

Non-canonical object-verb-subject (OVS) sentences as in (1a) and (1b) pose a challenge for children acquiring a language because in order to interpret them correctly full morphosyntactic processing is required as strategy (e.g. agent first). Sensitivity to morphosyntactic dependencies has been found already in children below the age of two years (Franck et al., 2013; Nazzi et al., 2011). Despite this early sensitivity, German learning children have difficulties correctly interpreting OVS sentences until the age of six years (Saueremann, 2016; Schipke et al., 2012). Work on another non-canonical structure (passives) suggests that 5-year-old children's difficulty with non-canonical sentences may result from a temporal ambiguity which requires revision of an initial parse (Huang et al., 2013). The current study investigates the role of temporal ambiguity by manipulating the availability of a reliable case cue on the first argument in OVS sentences in German. Based on previous findings we predict that children who use case marking information to parse non-canonical sentences show higher error rates and more processing difficulties for sentences with temporal ambiguity (1a) compared to those without (1b).

- (1a) *Das Seepferdchen fesselt der Papagei*
 [the sea horse]_{nom/acc} ties [the parrot]_{nom} 'The parrot ties the sea horse'
- (1b) *Den Vogel zieht der Hund*
 [the bird]_{acc} pulls [the dog]_{nom} 'The dog pulls the bird'

In a within-subjects design we presented the two types of non-canonical sentences (1a, 1b) as well as canonical SVO sentences to a group of five-year-old German-learning children in a visual world paradigm (two images with reversed roles, see Fig. 1) while simultaneously tracking their eye movements. So far, 33 children have completed the study (aimed sample size: 48). First results show lower accuracy for sentences containing a temporal ambiguity compared to sentences without (32.6% vs. 59.7%, $p < .001$). In addition to the offline response data we will analyse the eye tracking data in order to address the following main question: Is there an initial preference for a SVO interpretation in temporal ambiguous sentences and if so, can this preference be avoided or reduced by an early unambiguous case cue?

Furthermore, we will discuss individual differences and cognitive control abilities as potential factors on children's processing of non-canonical sentences. Cognitive control abilities were measured using the Flanker task (Eriksen & Eriksen, 1974) and the Task-Switching Paradigm (Jersild, 1927). The Flanker task is used as a measure of inhibitory control and the Task-Switching Paradigm as a measure of how flexibly children switch perspectives when controlling their attention.

Data collection is ongoing and preliminary results will be presented.

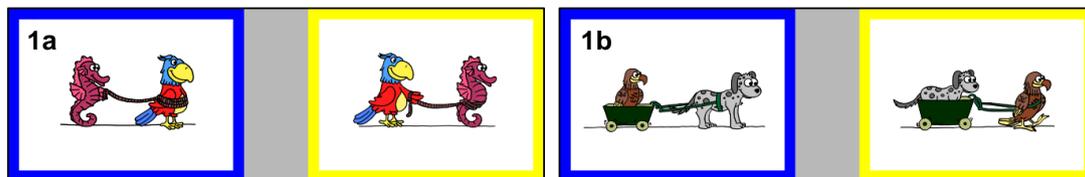


Fig. 1. Two image pair examples for sentence 1a (left panel) and sentence 1b (right panel). Children have to say the colour of the frame from the target image. The correct answer for (1a) is yellow and for (1b) blue.

LOST IN A STORY, DETACHED FROM THE WORDS: ABSORBED READERS ARE LESS SENSITIVE TO WORD CHARACTERISTICS DURING NARRATIVE READING

Roel Willems, Lynn Eekhof, Moniek Kuipers, Xin Gao, Marloes Mak, Emiel van den Hoven, Myrthe Faber, Marloes Mak; Radboud University; University of Basel; University of Freiburg

r.willems@let.ru.nl

Introduction It is well-known that readers are sensitive to word characteristics such as lexical frequency or age of acquisition. It is unclear, however, how differences in sensitivity to word characteristics are related to the reading *experience*. In this study we investigate the relationship between sensitivity to word characteristics and an important aspect of narrative reading, namely the experience of absorption.

Method We combined data from three previously conducted eye-tracking studies in which participants read literary stories under natural reading. Story length differed and participants read in total ~5,000 – 9,000 words each. Stories were annotated on a word-by-word basis for lexical frequency (Keuleers et al., 2010), age of acquisition, concreteness (Brysbaert et al., 2014), orthographic neighborhood size (Marian et al., 2012), and position in the sentence.

Participants filled out the Story World Absorption Scale (SWAS; Kuijpers et al., 2014) for each story. The SWAS is a questionnaire designed to measure absorption during story reading. It consists of statements about the reading experience which are scored on a Likert scale. Participants also indicated their appreciation for each story and filled out the Author Recognition Test (ART) as an indicator of print exposure. In total, data from 171 participants were analyzed. After preprocessing, the combined dataset contained 508,567 data points (2,126 unique words). Data were analyzed using linear mixed modeling in R.

Results We found that overall, participants' eye gaze durations were influenced by lexical frequency, age of acquisition and orthographic neighbourhood size. More important is that participants who scored higher on absorption, were less influenced by lexical frequency. Participants with the weakest coupling between lexical frequency and gaze duration reported the highest absorption scores. The same was true for ART and appreciation: those who were less sensitive to lexical frequency scored higher on print exposure (ART) and appreciation.

Conclusion Our results suggest that more absorbed readers pay less attention to low-level word characteristics. On the contrary, those who enjoyed the stories they read were least sensitive to lexical frequency. This is in line with earlier findings of more skilled readers being less sensitive to word characteristics, as well as to reports of 'mindless reading' in which readers are also less sensitive to word characteristics. In general, our results show that 'inefficient' reading is positively related to reading experience: those who are lost in the story are detached from the words.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, 150, 80-84.

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, 42(3), 643-650.

Kuijpers, M. M., Hakemulder, F, Doicaru, M. M., Tan, E. S. H. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, 4(1), 89-122.

Marian, V., Bartolotti, J., Chabal, S. & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE* 7(8): e43230.

Friday Poster.47

ERP CORRELATES OF SEMANTIC AND SYNTACTIC PROCESSING IN EARLY-DEAF COCHLEAR IMPLANT USERS.

L. Artesini (University of Trento, Italy), F. Pavani (University of Trento, Italy; CRNL, France),
D. Musola (Cooperativa Logogenia[®], Italy), G. N. Frau (Rovereto Hospital, Italy)
& F. Vespignani (University of Trento, Italy)
luca.artesini@unitn.it

Profound deafness can limit language acquisition, impacting on oral and written language processing. Research on the latter in deaf children revealed specific difficulties in syntactic and grammatical abilities, even in individuals who partially recovered hearing through cochlear implants (CI) (Niparko et al., 2010). In agreement with these behavioral results, EEG studies in congenitally deaf adults without CI showed typical N400 responses triggered by semantic incongruities, but no P600 effects in response to syntactic agreement violations (Mehravari et al., 2017).

To evaluate the impact of CI on written sentence processing, we tested 16 early deaf CI users (mean age=20.3 y., SD=9.0 y.; all non-signers; age at implantation: mean=8.7, SD=10.6, range 1-37 y.) and 16 age-matched hearing controls (mean age= 20.25 y.; SD= 8.5 y.), using EEG and behavioral measures (lexical decision, verbal fluency, sentence-picture match, error detection, etc.). To this aim we built 320 sentences: 80 with semantic incongruity (1), 80 with an embedded syntactic agreement violation (2) and 160 correct sentences as a control condition.

1. *Sulla scrivania ho appoggiato una **pecora** bianca.
*On the desk I put a **sheep** white. – [I put a white sheep on the desk.]
2. *Il presidente [n. sing.] **firmano** [v. plur.] l'accordo di pace.
*The president [n. sing.] **sign** [v. plur.] the peace agreement.

In CI users and normal-hearing controls alike, we documented a central distributed N400 (300 – 500 ms) and a centro-parietal P600 (550 – 1050 ms) responses to semantic and syntactic violations, respectively. Thanks to the high number of items per condition, these results are also quantifiable at a single subject level. Single subject analyses allow us to better understand the role of individual differences that are a known characteristic of the deaf-implanted population. An interesting and unexpected ERP effect emerged on frontal sites in an early time window (180-220 ms) where CI users show a larger frontal positivity for syntactic violations than age-matched controls that we interpret in terms of a modulation of the P2 that has been linked to lexical-orthographic processing (Carreiras et al., 2015).

Taken together these results suggest that CI usage can promote hearing-like brain responses to syntactic violations (intact P600) in early deaf participants with possibly greater reliance on orthographic rather than phonological processing in the early stage of detection of the morphosyntactic violation (enhanced P2).

References:

- Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., ... & CDaCI Investigative Team. (2010). Spoken language development in children following cochlear implantation. *Jama*, 303(15), 1498-1506.
- Mehravari, A. S., Emmorey, K., Prat, C. S., Klarman, L., & Osterhout, L. (2017). Brain-based individual difference measures of reading skill in deaf and hearing adults. *Neuropsychologia*, 101, 153-168.
- Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience*, 17(11), 1803-1817.

THE INTERACTION BETWEEN MEMORY RETRIEVAL AND EXPECTATIONS DURING SENTENCE PROCESSING

Luca Campanelli (Haskins Laboratories & The Graduate Center, CUNY),
Julie A. Van Dyke (Haskins Laboratories) & Klara Marton (The Graduate Center, CUNY)
campanelli.l@gmail.com

This study focuses on two classes of explanations for processing difficulty in sentence comprehension: Cue-based retrieval theory [1] and expectation theory [2,3]. According to the cue-based retrieval theory, processing difficulty derives from memory interference from similar words. The expectation theory characterizes processing difficulty in terms of surprisal [2], which corresponds to the total probability of the predicted sentence constructions that are not consistent with the current word. Overall, the two approaches capture different aspects of processing difficulty: the former a "backward-looking" cost, that is, the cost of retrieving and integrating previously processed material with the incoming words; the latter a "forward-looking" cost, that is, the cost of updating or dropping predictions that are incompatible with the current word [4]. There is now agreement that a comprehensive model of sentence processing needs to include both memory retrieval and expectation features in order to explain the full range of data available; however, how these two factors interact remains poorly understood. The present study examined the joint effects of retrieval interference and expectation, considering the possibility that expectations may diminish the damaging effects of retrieval interference by modulating the availability of the target element.

Method. Participants included 36 young adults (ages 22-37; 15 female). The experimental material consisted of object cleft sentences, as illustrated in (1). Memory interference was manipulated using the dual-task self-paced reading paradigm [5]. Interference was determined by whether three words in a memory list learned before reading (e.g., *website-handbag-password*) were (interference) or were not (no interference) plausible objects for the main verb (e.g., *performed/created*; 1a-c). No load conditions had no memory list, removing the possibility for retrieval interference. The effect of expectation was isolated by manipulating cloze probability of the main clause verb, that is, the predictability of the verb in the context of the stimulus sentence. The full design consisted of 6 conditions, with two levels of expectation examined within the interference conditions, but only low expectation within the no interference conditions. The No load conditions included the same sentences, but without the memory list (1).

- (1) a. NoInt-LowExp It was the *dance/* that the *person/* who lived/ in the city/ *performed/* early last month.
b. Int-LowExp It was the *dance/* that the *person/* who lived/ in the city/ *created/* early last month.
c. Int-HighExp It was the *dance/* that the *choreographer/* who lived/ in the city/ *created/* early last month.

Results and Discussion. Average response accuracy to the comprehension questions was above 90% in all conditions, and accuracy for recall of the memory list was on average above 92%. For reading time data at the critical sentence region (main verb; e.g., *created*), no significant effects were found. A different pattern of results was found at the spillover region (e.g., *early last month*). Reading time in the load conditions was faster than that in the no load conditions ($p = .018$), replicating previous results [6]. For the no load conditions, an effect of expectation emerged ($p < .001$), but—as expected—not of interference ($p = .6$), indicating that reading time was faster in the HighExp condition than in the two LowExp conditions. Most importantly, we found significant, or nearly significant, effects of both expectation ($p = .001$) and interference ($p = .054$) in the load conditions, such that reading was slowest in the Int-LowExp condition and fastest in the Int-HighExp condition.

Overall, our findings are consistent with the hypothesis that as the relevance of a specific element increases, the cued information is (probabilistically) pre-activated. If we think of memory retrieval as a gradual accumulation of information in the focus of attention, expectation would exercise its effect via an advance accumulation of evidence before the retrieval is initiated. Such a head start for selection of information would reduce retrieval interference by boosting the availability of the target word relative to its competitors.

References. [1] Lewis, Vasishth, & Van Dyke, 2006; [2] Hale, 2001; [3] Levy, 2008; [4] Demberg & Keller, 2008; [5] Gordon, Hendrick, & Levine, 2002; [6] Van Dyke & McElree, 2006.

DID JOHN CALL HER MOTHER? THE ROLE OF NATIVE-LANGUAGE PROCESSING STRATEGIES IN SECOND-LANGUAGE PRODUCTION ERRORS

Wing-Yee Chow, Esther Jesús Ortiz, Shasha Jin (University College London)
 wingyee.chow@ucl.ac.uk

English possessive adjectives like ‘his’ and ‘her’ are notoriously problematic for second language (L2) learners who speak a Romance language like Spanish and Italian natively [1]. For example, Spanish, French and Catalan learners of English have all been documented to produce errors such as ‘John is calling **her* mother’, where the possessive adjective agrees in gender with the possessee (mother) instead of the possessor (John). This type of errors are commonly attributed to negative L1 transfer [2] because possessive adjectives in English agree only with the possessor, whereas in Romance languages possessive adjectives behave like determiners and agree in gender and number with the possessee noun (Figure 1). In this study we used an elicited production task to investigate the conditions under which such errors occur and examined the role of L1 influence by comparing the performance of Spanish-English bilinguals with that of native English speakers as well as Chinese-English bilinguals, for whom possessive adjectives behave the same way in both their L1 Chinese and their L2 English.



Figure 1. Possessive adjective in English and Spanish

In each trial participants listened to a short narrative about a named character, and were then prompted to answer a question about the narrative by speaking into a microphone (Figure 2). We manipulated the gender match between the possessor (e.g., Tim) and the possessee (e.g., mother) in the narratives, and participants' responses were scored based on whether the possessive adjective in their response had the correct gender; responses without a possessive adjective were excluded. In addition, an end-of-study questionnaire was used to examine bilingual speakers' grammatical knowledge of English possessives. If Spanish-English bilinguals pre-activate the possessee noun prior to the possessive adjective during production, then they should be more likely to use a possessive adjective with an incorrect gender when the possessor and possessee do not match in gender (a ‘gender mismatch’ effect, e.g., John calls **her* mother vs. father) compared to the other groups.

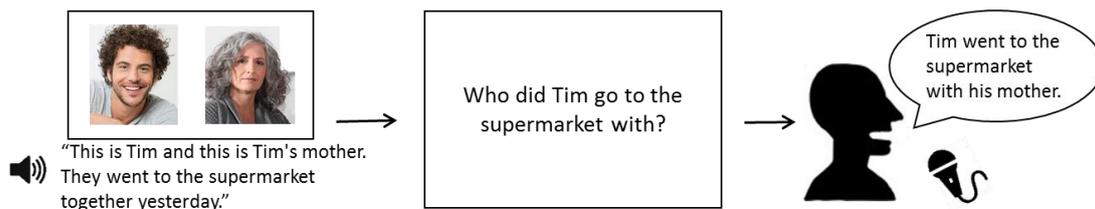


Figure 2. Elicited production task

Results in the elicited production task (n=44 in each group) revealed that while English native speakers performed at ceiling across conditions, Spanish-English bilinguals were significantly less accurate when the possessee and possessor differed in gender (a gender mismatch effect, match=92% vs. mismatch=77%). Crucially, despite their at-ceiling performance in the end-of-study questionnaire, Spanish-English bilinguals showed a significantly larger gender mismatch effect in the production task than Chinese-English bilinguals (97% vs. 93%). Taken together, these results suggest that bilingual speakers' L1 processing mechanisms may impact how they process their L2 in real time even when they have fully mastered the relevant grammatical knowledge in their L2.

References: [1] Lightbown & Spada (1990). *SSLA*, 12, 429–448. [2] Antón-Méndez (2011). *Lang & Cogn*, 14, 318–331. [3] Thierry & Wu (2007). *PNAS*, 104, 12530–12535.

Implicit Causality Affects the Choice of Anaphoric Form

Oliver Bott (Tübingen), Torgrim Solstad (ZAS, Berlin) & Anna Pryslopska (Tübingen)

The choice of anaphoric forms (e.g. *Mary* vs. *she*) depends on a number of factors such as grammatical function, order of mention or topicality (Arnold 2008). For semantic-pragmatic re-mention biases however, which also impact referent salience, recent research has found conflicting results. Thus, Implicit Causality (IC) verbs of S(timulus)-E(xperiencer) (e.g., *fascinate*) and E(xperiencer)-S(timulus) type (e.g., *admire*) display strong preferences for subsequent explanations about the Stimulus argument (Ferstl et al. 2011). Yet, Fukumura & van Gompel (2010; F&vG2010) and Rohde & Kehler (2014) found no effect of IC on anaphoric form. Kehler & Rohde (2013; K&R2013) a.o. thus claim that the production of anaphoric form is dissociated from the likelihood of mention. However, Rosa & Arnold (2017; R&A2017) found that Transfer of Possession (ToP) verbs, with a re-mention bias for goal arguments (e.g., indirect object of *sell* or subject of *buy*) do influence the choice of anaphoric form. R&A2017 speculate that differences in argument structure are behind this discrepancy.

The present study pursued an alternative explanation: Crucial experimental conditions have not been tested for IC verbs yet. Rohde/Kehler (2014), improving on F&vG2010's paradigm, point out that choice of anaphoric form is especially important with two same-gender referents as a strategy to avoid ambiguity. However, unlike F&vG2010 and R&A2017, they did not use a forced-reference paradigm, in which participants are prompted to provide continuations for one particular referent. This is essential when comparing bias-congruent and bias-incongruent continuations. Including all these factors as experimental conditions, we are in a better position to assess the influence of IC on anaphoric form. Importantly, our experiments 2a and 2b allowed for a within subjects comparison of IC and ToP verbs.

German items with 20 SE verbs and 20 ES verbs (+ 40 fillers) were constructed in a 2 (*verb type*) x 2 (*gender ambiguity: same vs. different-gender referents*) design (e.g., *Mary admired/fascinated Peter/Jane because...*). A pretest (N=24) confirmed that the materials were biased above 90% towards subject and object stimuli, respectively. **Experiment 1** (N=32) employed the same method as F&vG2010 and R&A2017 (Exp.3), highlighting the continuation's intended referent (factor *referent focus: subject vs. object*). Anaphoric forms more complex than PERS(onal pronouns) were restricted to object focus continuations (all subject focus conditions $\geq 95\%$ PERS). In the object focus conditions we observed a strong effect of *ambiguity* and a marginal interaction *ambiguity* by *verb type* (GLMER model comparisons: *ambiguity* $\chi^2(1)=15.2$; $p<.001$; interaction $\chi^2(1)=3.4$; $p=.07$). In conditions with different-gender referents, PERS continuations were produced 93.5% of the time for SE items and 89.5% for ES items. In the same-gender conditions, SE items received 51.7% PERS for (bias-incongruent) object continuations as compared to 66.4% PERS for (bias-congruent) object continuations for the ES items. **Experiment 2a** (N=42) examined this marginal *verb type* effect by testing IC items in object focus conditions only. **Experiment 2b** tested 24 ToP items in a 2 (*referent focus*) x 2 (*verb type: subject-goal vs. object-goal*) x 2 (*ambiguity*) design. **Exp. 2a** revealed clear form effects of *gender ambiguity* ($\chi^2(1)=23.1$; $p<.001$) and an effect of IC *verb type* ($\chi^2(1)=6.5$; $p<.05$): In the different-gender conditions PERS were produced 86.3% in bias-congruent ES items and 78.0% for bias-incongruent SE items. In the same-gender conditions, bias-congruent ES items received 62.1% PERS continuations as opposed to only 48.9% PERS for bias-incongruent SE items. The same pattern of effects appeared in the object focus conditions of ToP verbs in **Exp. 2b** (subject focus $>95\%$ PERS; as in Exp. 1). Both *ambiguity* ($\chi^2(1)=14.4$; $p<.001$) and *verb type* ($\chi^2(1)=8.60$; $p<.01$) contributed significantly to the regression analysis with 75.5% PERS continuations (congruent object-goal) and 73.3% (incongruent subject-goal) in the different-gender conditions and 56.9% (object-goal) relative to only 38.8% (subject-goal) in the same-gender conditions.

In sum, the results of our experiments show that – modulated by well-known effects of audience design – referential biases affect reference form production across verb classes, including IC verbs. This finding speaks both against proposals assuming a general dissociation between likelihood of mention and choice of anaphoric form (K&R2013) as well as proposals assuming an interaction with argument structure (as speculated in R&A2017).

**EFFECTS OF REFERENTIAL GAZE IN SPOKEN LANGUAGE COMPREHENSION:
HUMAN SPEAKER GAZE VS. VIRTUAL AGENT LISTENER GAZE**

Eva M. Nunnemann (Bielefeld University), Kirsten Bergmann (Bielefeld University of Applied Sciences), Helene Kreysa (Friedrich Schiller University Jena), Pia Knoeferle (Humboldt University Berlin)
enunnemann@techfak.uni-bielefeld.de

An interlocutor's gaze has rapid effects in face-face communication. Its exploitation has been shown to be beneficial for visual anticipation in joint search tasks [2]. It also helps listeners to understand event roles faster and to quickly anticipate referents [3] and has beneficial effects in offline recall tasks [4]. Studies have shown that people even react to artificial gaze from robots or virtual agents: A virtual agent's gaze helped to enhance interaction in storytelling [1]. As instructor, an agent who inspected the study materials while teaching helped participants to better recall the lesson's content [5]. What has not been answered yet, is, which effect a simultaneous presence of these gaze types has, whether human and virtual agent gaze guide attention in a similar manner and whether the gaze types influence performance in a recall task.

In two studies, we maximally contrasted human speaker gaze and virtual agent listener gaze and then measured their effects on response times, accuracy and eye movements. We manipulated three factors: speaker presence, agent presence and the match between video scene and a subsequent template. Participants watched videos in which – varying with condition - a human speaker and a virtual agent listener jointly inspect a computer screen depicting a static scene with three characters. In experiment 1 the speaker uttered German subject-verb-object sentences describing an interaction between two characters, e.g. 'Der Kellner beglückwünscht den Millionär.' (The waiter congratulates the millionaire), while in experiment 2 the NP2 (e.g. the millionaire) was made unintelligible. Here participants had to rely on the gaze cues to detect the NP2 referent. After each trial a template appeared on screen, schematically representing the characters from the video and an interaction between two of them. Participants verified whether the sentence and the template matched.

Responses in both studies were faster for matches than mismatches. In both experiments, participants were slower to answer for those trials in which the virtual agent was present compared to the conditions when he was not there. The eye movement data for both studies suggest that participants look more to the NP2 referent in those conditions in which speaker gaze was available. Participants did not rely on the agent listener gaze. They do not even exploit it in the condition of experiment 2 in which agent gaze was the only cue available to detect the correct NP2 referent of the unintelligible NP2. This finding suggests that human speaker and virtual agent listener gaze have different effects on a human listener when they are presented simultaneously. Thus, it can inform accounts of situated language processing on how to integrate effects of gaze into language comprehension.

- [1] Bee, N., Wagner, J., Andre, T., Elisabeth and Vogt, Charles, F., Pizzi, D., & Cavazza, M. (2010). Discovering eye gaze behavior during human-agent conversation in an interactive storytelling application. In *International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction* (p. 9).
- [2] Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106, 1465–1477.
- [3] Kreysa, H., & Knoeferle, P. (2011). Effects of speaker gaze on spoken language comprehension: Task matters. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society*.
- [4] Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29, 1045–1060.
- [5] Ruhland, K., Peters, C., Andrist, S., Badler, J. B., Gleicher, M., Mutlu, B., & McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. *Computer Graphics forum*, Vol. 34, No. 6, 299- 326.

Friday Poster.52

THE RELATIONSHIP BETWEEN READING SKILL AND LINGUISTIC PREDICTION: AN INVESTIGATION OF DYSLEXIA

Mohd Taha Norhafizah (UEA), Michelle Yuen (UEA), & Paul E. Engelhardt (UEA)

N.Haji-mohd-taha@uea.ac.uk

There is a positive relationship between reading skills and linguistic prediction, which has been demonstrated in both children and illiterate individuals. In this study, we extend these findings to individuals with reading disability (i.e. dyslexia). Evidence suggests that dyslexia is mostly commonly associated with a specific learning disability affecting skilled reading. However, in this study, we investigated prediction in spoken language in order to assess deficient prediction from more commonly reported issues in dyslexia (i.e. failure to develop fluent reading). Following Huettig and Brower [1], we predicted that dyslexics will exhibit deficits in linguistic prediction even in spoken language. Native adult English speakers with dyslexia ($N=20$) and typically-developing (TD) controls ($N=23$) participated in two experiments, based on the "cloze task", which is the gold-standard psycholinguistic task used to assess prediction. In Experiment 1, participants heard an incomplete sentence and had to complete it as naturally as possible. In Experiment 2, participants heard a complete sentence and had to provide speeded semantic-plausibility judgements.

In Experiment 1, we assessed two DVs. The first was the cloze probability of the word produced and the second was voice onset time (VOT)[2]. If linguistic prediction is significantly diminished in dyslexia, we expected individuals with dyslexia to produce lower cloze probability responses and to have longer VOTs. Independent samples t -tests showed that dyslexics produced significantly lower cloze probability completions $t(41)=2.68, p=.01$ than did TD controls ($M=.39, SD=.04$ vs. $M=.43, SD=.04$, respectively). In contrast, there was no difference in VOT $t(41)=-1.35, p=.19$ (TD: $M=914\text{ms}, SD=251$ & DYS: $M=1057\text{ms}, SD=433$).

In Experiment 2, we used a semantic-plausibility task, in which participants had to respond YES/NO whether the sentence made sense. Three types of sentence completions were used. The first was the highest (cloze) probability word for the sentence, the second was the lowest probability word, and both were based on published norms[3]. Finally, we had a semantically anomalous completion. We further divided our 130 sentences in those that had "high constraint" and those that had "low constraint". This division was again based on published norms. Results of a $2 \times 2 \times 3$ (group: DYS vs. control \times constraint: high vs. low \times completion: high, low, anomalous) mixed model ANOVA showed a significant 3-way interaction. In general, the DYS participants were significantly slower than TD controls. However, in high constraint sentences, we observed that both groups showed significantly faster RTs between the three types of completions (high<low<anomalous), but crucially, in low constraint sentences, the groups patterned differently, the controls again showed significant differences between high and low completions, but the DYS group did not. Thus, participants with dyslexia showed no evidence of being able to predict in less constraining sentences.

Across E1 and E2, we found that dyslexics were "worse" predictors, and thus, we argue that skilled reading is related to linguistic prediction. We can rule out prediction impairments in reading specifically, as both studies utilized spoken language. Ultimately though, the current results present a chicken-and-egg problem--does reading skill lead to better prediction or does better prediction lead to better reading. Given the importance of the role of prediction in language, we think our work is important in convincingly identifying a clinical population that shows deficits in linguistic prediction.

References

- [1] Huettig & Brower (2015). Delayed anticipatory spoken processing in adults with dyslexia – Evidence from eye-tracking. *Dyslexia*, 21(2), 97-122.
- [2] Staub et al. (2015). The effect of lexical predictability on eye movement in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311-321.
- [3] Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.

AN EAR (AND EYE) FOR LANGUAGE: PREDICTORS OF INCIDENTAL AND EXPLICIT FOREIGN LANGUAGE VOCABULARY LEARNING

Marie-Josée Bisson (De Montfort University), Anuenu Baker-Kukona (De Montfort University) & Angelos Lengieris (University of Kent)
marie-josée.bisson@dmu.ac.uk

Much of past research on predictors of foreign language (FL) word learning has focused on explicitly learning a small number of highly repeated items, often pseudowords. It is thought that in order to become fluent in a FL, learners need to acquire a large vocabulary (30 000 words), and this can only realistically be achieved through both explicit and incidental learning activities. Here we therefore systematically investigated predictors of real FL word learning using both incidental and explicit learning situations. We included a large number of items with a few repetitions to reflect how languages are learnt 'in the wild'. The overall aim of our research program is to investigate what makes a good language learner, and what we can do to facilitate learning.

On day one, participants (monolingual native English speakers with no prior knowledge of the FL) first took part in an incidental learning task followed by a carefully matched explicit learning task. Both learning tasks presented participants with auditory and written FL word forms as well as pictures depicting the meaning of the words (8 repetitions in total for each of the 80 FL words). For incidental learning, participants completed a letter-search task on the written word forms, and were unaware of the word learning aspect of the study. For the explicit learning task, participants were instructed to learn the FL words for an upcoming test. On day 2, incidental and explicit FL word learning was assessed through recall and recognition tests. Participants also completed a series of computerised tasks assessing their cognitive abilities (verbal and visuo-spatial short-term and working memory, Flankers and Stroop to measure executive functions, orthographic abilities, phonological abilities along F2, FL and L1 continuum) as well as questionnaires to measure their motivation and confidence to learn a FL. Finally, participants completed a native language vocabulary knowledge test as a control measure.

We calculated composite scores for the memory, phonological abilities and executive functions tasks as well as for the motivation and confidence questionnaires. All other tasks were entered in the model as simple predictors. For the dependent variable, recognition and recall test scores were normalised and averaged together. The regression analysis revealed that, as well as memory and native language vocabulary knowledge, phonological and orthographic abilities were significant predictors of language learning. It seems indeed that a good language learner must have an ear and eye for language. Interestingly, both the abilities to hear subtle differences in non-linguistic (F2) and FL material were important for language learning. In addition, follow-up analyses revealed that phonological abilities, but not memory was crucial for incidental learning, whereas the opposite was true for explicit learning.

The results of the current study demonstrated an important role for phonological and orthographic abilities in language learning. Crucially, recent research has shown that phonological abilities can be trained. Important questions remain: firstly, as to the impact of phonological training on subsequent vocabulary learning and secondly as to whether orthographic abilities can be trained. FL vocabulary learning is an immensely daunting task if one is to achieve fluency. Hence, fine tuning phonological and orthographic abilities to reduce this burden could have a large impact on FL learning over time.

WHY TWO IS NOT ALWAYS BETTER THAN ONE – AN ERP STUDY ON MINIMALITY-BASED AND PROSODIC PREDICTIONS IN GERMAN DISCOURSE PROCESSING

Petra Augurzký (University of Tübingen) & Nadja Schaufller (University of Stuttgart)
 petra.augurzký@uni-tuebingen.de

Recent studies have shown that predictive processing plays an essential role in language comprehension. By reasoning about what a speaker will utter next, comprehenders actively contribute to quick and efficient communication. However, it is still a matter of debate how detailed such predictions are. For instance, the question whether expectancies involve specific form-based (i.e. phonological or prosodic) characteristics is currently an issue of vivid discussion (Nieuwland et al., 2018). As previous research on form-based predictions predominantly focused on the comprehension of single sentences, the present study investigates whether prosodic properties may affect predictive processing on the discourse level.

In two silent-reading ERP studies, we investigated whether prosodic properties interact with predictive processing in contrastive focus environments (Table 1), in which either one (C), or two (A,B) contrastive entities were involved. Most generally, we hypothesized that readers would expect discourses involving a minimal number of entities. Therefore, a given entity should be predicted to follow an already contrastively focused entity in the target sentence. Double contrastive foci (A,B) should thus incur processing cost in terms of an N400 / P600 on NP2 compared to single contrastive foci (C, Schumacher, 2014). In **Exp. 1**, we examined if these expectancies were mediated by a preference for rhythmic alternation. Whereas both A and B involve a double contrast, only A involves an additional rhythmical irregularity in terms of a phonological clash elicited by the presence of directly adjacent pitch accents. In **Exp. 2**, sentences did not involve a pitch accent clash, but NP2s varied as a function of prosodic similarity to the expected given entity: In A, NP2 bears the same word stress as the contextually given NP, and in B, word stress of NP2 differs from the given NP. We thus examined if discourse-based expectancies resulted in the prediction of word stress of an upcoming NP.

Table 1: Sample stimuli of the two ERP studies

	ERP Experiment 1 (ERPs were aggregated from NP2)	ERP Experiment 2 (ERPs were aggregated from NP2)
	Hat Melli gesagt, dass Tobi das Schlagzeug SCHÜ lerinnen gegeben hat? Nein, sie hat gesagt, dass	Hat Melli gesagt, dass Tobi das Klavier SCHÜ lerinnen gegeben hat? Nein, sie hat gesagt, dass
A	a. ...Tobi [das Kla VIER] _{NP1} [LE Hrerinnen] _{NP2} gegeben hat.	a. ...Tobi [das SCHLAG zeug] _{NP1} [LE Hrerinnen] _{NP2} gegeben hat.
B	b. ...Tobi [das Kla VIER] _{NP1} [Stu DENTinnen] _{NP2} gegeben hat.	b. ...Tobi [das SCHLAG zeug] _{NP1} [Stu DENTinnen] _{NP2} gegeben hat.
C	c. ... Tobi [das Kla VIER] _{NP1} [Schü lerinnen] _{NP2} gegeben hat.	c. ... Tobi [das SCHLAG zeug] _{NP1} [Schü lerinnen] _{NP2} gegeben hat.
	<i>Has Melli said that Tobi the drums pupils given has? No, she has said that Tobi the piano a. teachers / b. students / c. pupils given has.</i>	<i>Has Melli said that Tobi the piano pupils given has? No, she has said that Tobi the drums a. teachers / b. students c. pupils given has.</i>
	A = contrastive, clash, B = contrastive, no clash, C = given	A = contrastive, identical stress, B = contrastive, different, C = given

As expected, we found an increased N400 on NP2 for contrastive vs. given discourse entities in both ERP studies, which can be explained by a strong expectancy of an already mentioned entity. However, only in **Exp. 1** did the prosodic manipulation interact with predictive processing: the N400 significantly differed between the two unexpected contrastive NP2s, with a larger N400 for NP2 in condition B vs. condition A. Based on previous sentence production studies, we consider it likely that the phonological expectancy of a rhythmically alternating structure led to the realization of an unaccented NP2 in silent reading, thus rendering NP2 in A phonologically more similar to the expected given NP. By contrast, in **Exp. 2**, the N400 was increased for both unexpected NPs (A,B) alike, with no effects of the word stress manipulation.

In sum, we found strong evidence for an expectancy of minimal discourse in contrastive focus environments: Across studies, a pronounced N400 for contrastive NP2s indicated that readers expected a given entity after already having encountered a contrastive NP. While the presence of a pitch accent clash interacted with this preference (Exp.1), word stress of NP2 (Exp.2) did not affect the N400. The finding that the N400 was not accompanied by a P600 may be explained by the fact that the discourse update did not require an additional meaning shift, as was often involved in previous work (Schumacher, 2014). Our results suggest that the rhythmic preference of a strong-weak alternation elicits expectancies of upcoming lexical items alongside discourse-related predictions, probably due to a decisive role of low-level phonological principles in German even in silent reading. Our studies thus show that readers' expectancies in multiple focus environments are not triggered by a preference for minimal discourse alone, but may also be modulated by low-level suprasegmental demands.

RATES OF SCALAR INFERENCE BEYOND 'SOME' – A CORPUS STUDY

Chao Sun (University College London), Ye Tian (Université Paris Diderot) & Richard Breheny (University College London)

Previous research [1-2] suggests that different scalar expressions give rise to scalar inferences (SIs) at different rates. For instance, [2] used an inference paradigm to establish the variability of inference rates across a wide range of scalar expressions. In [2] the rates of SIs were measured with little context, while [3-4] suggest that participants tend to infer contexts for a single-sentence in terms of QUDs. As a result, different QUDs are likely to be inferred for different scalar terms. Since the availability of SIs is sensitive to QUDs, without controlling for likely QUDs, the diversity pattern found previously might not truly reflect how liable a scalar term is in triggering SIs. Here we argue that this issue could be overcome by measuring SI rates in a corpus-based approach. We collected large samples of sentences containing scalar terms and measured SI rates using the paraphrased task from [5]. Although the observed variation among scalar terms is significantly lower than in [2]'s inference task, we do find the diversity pattern.

Collection of a Twitter corpus: We selected 28 out of 43 scalar expressions from [2]. For each scale, we extracted tweets containing the weak scalar term (min. length = 30 characters). Then we filtered out tweets where scalar expressions appear in environments where the inferences are unavailable or less likely to arise. A word sense disambiguation task was conducted on Amazon MTurk to obtain human annotation on tweets containing polysemous scalar expressions. Considering <old, ancient> for example, 'old' can mean *existing a long time*, which is on the same scale as the core meaning of 'ancient'; but 'old' can also mean *previous*, which could not form a scale with 'ancient'. We excluded texts where the use of the weak scalar term (e.g. old) was annotated as the sense that was irrelevant to its strong scale mate. **Corpus-based paraphrase task:** We ran a paraphrase task based on [5] to measure the rate of SIs triggered by scalar expressions in the corpus. We selected at random 1400 items from a total set of 3075, 50 per scale. Participants (n=550) each judged 28 items, one item per scale.

Results: We found that the frequency of SI varies across scalar expressions, from 27% for <adequate, good> to 86% for <sometimes, always>. These results correlated with the results of [2] ($r=0.81$, $p<.001$), suggesting that, to some extent, the results yield from the inference task based on artificial examples could reflect frequencies of SIs triggered in naturalistic data. However, Levene's test showed that variances of two studies are not equal ($F(1,54)=14.69$, $p<.001$). There is less variation on the paraphrase task. The rates of SIs for quantifiers in the paraphrase task are far lower than rates found in [2] (replicating of [5]'s results), whereas many adjective expressions give rise to SIs more frequently in the paraphrase task.

Discussion: Results bear out our expectation that items used in previous studies impacted on the degree of diversity among scalar terms. Nevertheless, a diversity pattern remains, established by a corpus-based method. Multiple linear regression analyses were conducted to predict the rate of SIs from possible factors explored in [2], including association strength, grammatical class, word frequencies, semantic relatedness, semantic distance, and boundedness. As [2] found with their inference task results, only semantic distance and boundedness are substantial factors.

References: [1] Doran, R. (2009). *International Review of Pragmatics*, 1, 211–248. [2] van Tiel, B. van Miltenburg, E. Zevakhina N. & Geurts, B. (2016), *Journal of Semantics*, 33: 137-175. [3] Breheny, R., Katsos, N., & Williams, J. (2006). *Cognition*, 100(3), 434–463. [4] Bergen, L. & Grodner, D. (2012). *JEP: Learning, Memory, and Cognition* 38(5), 1450-1460 [5] Degen, Judith. 2015. *Semantics and Pragmatics* 8(11). 1–55. [6] Bergen, L., Levy, R., & Goodman, N. D. (2016). *Semantics and Pragmatics*, 9.

ASPECTUAL MAKE-UP REDUCES RELATIVE CLAUSE AVOIDANCE

Miriam Aguilar (Universidade Nova de Lisboa) & Nino Grillo (University of York)
 maguilar@fcs.unl.pt

A strong result in the literature on sentence processing is that the parser consistently avoids a Relative Clause (RC 1a,2a) parse whenever possible (see [1,2] for recent results involving different types of clausal complementation). Restrictive RCs denote properties of individuals: this property is ascribed to the referent of the DP they modify but not to other elements in a contrast set, allowing its unique identification within that set. RC-avoidance is thus reduced or completely eliminated when a contrast set is provided by either felicitous discourse context [3,4,5] or by the presence of focus operators such as *only* [6]. Relatively little attention was paid to factors promoting building a property out of a sentence. We focus on grammatical aspect, and in particular on the availability of a habitual reading of the RC-predicate. We argue that habits are more readily converted into properties than punctual events: *John is a runner* obtains more naturally from the habitual *John (always) runs* than from the progressive *John is running*. In support of this hypothesis, and adding to a growing literature on adaptation [7,8,9,10], we report selective learning effects in the resolution of RC-attachment ambiguities: Spanish speakers more readily adapt to a RC parse when a habitual reading is available.

Background. A number of recent offline and online studies have shown that RC are also avoided whenever the alternative Pseudo Relative (PR, 1b) parse is available [2,11]. RCs containing *imperfective* or *progressive* aspect and appearing in the complement position of perceptual (but not stative) verbs in Spanish can be interpreted as PRs, which denote an event and roughly correspond to an eventive Small Clause in English (2b):

- (1) a. Vi [DP al [NP chico [RC que (siempre) corría]]] b. Vi [PR al chico que corría (*siempre)]
 (2) a. I [V saw [DP the [NP boy [RC that (always) ran]]]] b. I [V saw [SC the boy running (*always)]]

Previous results consistently show that PR-availability has a significant impact in the resolution of RC attachment ambiguity [2]: High Attachment (HA) occurs when PRs are available and Low Attachment (LA) otherwise (provided that other factors are controlled for, e.g. Prosody). We add to this literature by showing that this effect is susceptible to learning effects and that adaptation and learning are modulated by the aspectual make-up of the sentence.

Experiment. We contrast the processing of ambiguous RCs in past imperfective with past progressive. Spanish *imperfective* is ambiguous between simple past and the habitual reading. This ambiguity is not readily available with past progressive, which more easily describe punctual events. As habituais are not compatible with PRs in Spanish, we expect stronger adaptation to RC-reading with *imperfective*, i.e. when habitual reading is available.

Method: 80 native Spanish speakers participated in a forced-choice RC-attachment task. We manipulated VERB TYPE (*perceptual* vs. *stative*) of the matrix sentence and ASPECT (*imperfective* vs. *progressive*) of the embedded clause (Table 1). In line with previous results, we predict higher proportion of HA preference with perceptual (PR-compatible) than with stative (PR-incompatible) verbs. We also predict *selective adaptation* to the high ratio (1:2) of unambiguous RCs in the experiment, observable as a stronger increase in overall proportion of Low Attachment choices for imperfective than progressive aspect with perceptual verbs.

Results: We observed a strong effect of VERB TYPE on attachment (z-value=-8.676, p<.0001), with more HA for perceptual than stative. A 3-way interaction between VERB TYPE, ASPECT and HALF (first vs. second half of the experiment) for perceptual-imperfective sentences (z-value=-3.804, p=.0001) but not for perceptual-progressives (z-value=-0.982, p=0.32) show that the effect is susceptible to selective learning effects.

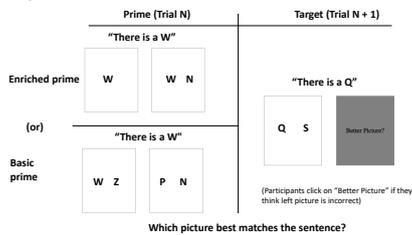
VERB TYPE	ASPECT	Sample Sentence	Table 1: example of sentence stimuli
Perceptual	Imperfective	a. Juan vio al asistente del médico que leía el periódico. John lives with the assistant of the doctor that read the newspaper.	
Perceptual	Progressive	b. Juan vio al asistente del médico que estaba leyendo. John saw the assistant of the doctor that was reading.	
Stative	Imperfective	c. Juan vive con el asistente del médico que leía el periódico. John lives with the assistant of the doctor that read the newspaper.	
Stative	Progressive	d. Juan vive con el asistente del médico que estaba leyendo. John lives with the assistant of the doctor that was reading.	

References [1] Staub et al. 2018; [2] Grillo & Costa 2014; [3] Crain & Steedmann 1985; [4] Altmann & Steedman 1988; [4] Trueswell et al., 1999; [6] Ni et al 1996; [7] Clayards et al. 2008; [8] Kamide 2012; [9] Fine et al. 2010,2013; [10] Fernandes et al 2018, CUNY; [11] Grillo et al. 2015a (Cognition), 2015b (CUNY).

BILINGUAL PRIMING OF PRAGMATIC ENRICHMENT
 Lewis Bott (Cardiff University), Susanne Münz (Cardiff University)

A central question in bilingual language research is whether abstract representations are shared across languages or whether there are distinct representations for each language. Previous research has shown that some syntactic representations are shared (e.g. Hartsuiker, Pickering & Veldtkamp, 2004). Here we show that a different sort of process, pragmatic enrichment, also involves shared representations across languages.

A common sort of pragmatic enrichment involves combining the basic meaning of a sentence with the negation of a stronger alternative. For example, “John closed the door,” can be enriched to *John did not slam the door*, even though the speaker did not say that the door was not slammed. Grice (1975) viewed these sorts of enrichments (*implicatures*) as the result of adherence to a set of abstract conversational maxims combined with domain general reasoning. This suggests a language independent process. However, many others have argued that the alternative used in the derivation must involve a lexical component (e.g. Levinson, 2000) and that the degree of enrichment varies across lexical expression (van Tiel et al., 2016) within a language, both of which suggest a language specific component.

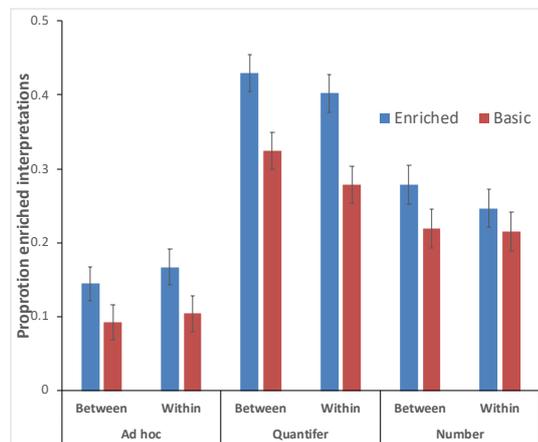


To test between a language-specific and a language-general model of implicature derivation, we used a structural priming task that tested pragmatic enrichments (Bott & Chemla, 2016) on German-English bilinguals. Participants saw pairs of pictures and clicked on the picture that best described the sentence (Fig 1.). Priming occurred if interpretations of the prime sentence influenced interpretations to the target sentence. We used

three types of enrichment (1) quantifiers (2) number (3) *ad hoc* expressions. Figure 1 shows an example. Prime-target pairs could either be within-language (English-English or German-German) or between-language (English-German or German-English). If enrichment procedures are shared across languages, enrichment priming should occur equally across languages as within.

Results with N=204 participants showed large within-language priming effects $t(203) = 6.0, p < .001$, and large between-language priming effects, $t(203) = 7.97, p < .001$, for all expression types, and with no difference in the size of the effect within/between languages, $F(1,203) < 1$. Participants were more likely to derive pragmatic enrichment after they had done so in the prime trial than when they had not, and this effect was independent of whether the prime and target were in the same language.

Our study makes two contributions. (1) We show that bilinguals exhibit shared pragmatic representations just as they do for some syntactic representations (2) Implicatures may be best thought of as involving competition between non-lexical components, i.e. conceptual competition (e.g., \exists vs \forall), rather than the traditional lexical competition (*some vs all*).



Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, 15(6), 409-414. Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, 91, 117-140. Grice, H. P. (1975). Logic and conversation. 1975, 41-58. Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, 33(1), 137-175.

PLANNING UNTIL THE END OF THE SENTENCE?

Agnieszka Konopka (University of Aberdeen) & Nele Ots (University of Tartu)

agnieszka.konopka@abdn.ac.uk

How extensive is early message planning? When generating SVO descriptions of simple events (e.g., “*The man is pushing the car*”), speech onsets and early eye movements are sensitive to the ease of naming the first character (“man”) but not the second character (“car”). This suggests priority encoding of message elements that will be lexicalized first (i.e., left-to-right incremental planning; [1,2,3,4]). Here we test to what the extent speakers devote processing resources to the patient [2] in events that explicitly draw attention to this character, and we assess to what extent patient planning is constrained by properties of the agent.

Experiment: 51 native, eye-tracked English speakers described pictures of events where we manipulated the number and complexity of the patient(s). Target events showed an agent acting on one patient or on one of two similar patients. In one-patient events, the patient was a simple object (e.g., a car) or an object with a distinctive feature (a car with a surfboard on top). Both events elicited *unmodified* descriptions: “*The man is pushing the car*”. In two-patient events, the two patients differed in color or size (e.g., one purple car and one red car) or in a distinctive feature (the target car had a surfboard on top), and elicited descriptions with short prenominal and long postnominal modifiers respectively: “*The man is pushing the purple car*” and “*The man is pushing the car with the surfboard*”. Across events, short modifiers ranged from 1 to 5 syllables ($M=1.6$); long modifiers ranged from 2 to 14 syllables in length ($M=4.8$).

If speakers begin to encode the patient early during message planning (primarily 0-400 ms after trial onset; [2]), they should be more likely to fixate the patient(s) when describing events with more complex than less complex patients(s). Thus, gaze shifts in one-patient events should show whether the conceptual complexity of the patient influences planning when the linguistic output is held constant. Gaze shifts in two-patient events should show the extent of early planning of the patient – specifically, whether the message plan includes only information that the patient must be modified in some way (irrespective of modifier type) or includes more detailed modifier information (as indexed by modifier length).

Results (1): Speech onsets (Mixed Models). The expected modifiers (none vs. short vs. long) were produced in 87% of all SVO sentences. Speech onsets were predicted independently by Agent Codability (shorter onsets in sentences with easier-to-name than harder-to-name agents), Patient Number, Modifier Type and Length (longer onsets in descriptions of two-patient than one-patient events, and with longer than shorter modifiers).

Results (2): Time-course of planning (Growth Curve Analyses). Early eye movements (0-400 ms) were also predicted independently by Agent Codability (more fixations away from the agent and towards the patient(s) when the agent was hard to name) and by Modifier Type. In one-patient events, speakers were more likely to fixate the patient when this character included a salient (but never-mentioned) feature than when it did not, indicating an influence of conceptual but not linguistic complexity of the patient. In two-patient events, speakers were more likely to fixate the target patient when it was described with a postnominal than a prenominal modifier and with a longer than a shorter modifier, showing early sensitivity to patient complexity. Importantly, speakers began fixating agents rapidly after 400 ms (indicating linguistic encoding of the agent) and these fixations were not delayed by Modifier Type or Length, confirming that speakers had begun conceptual but not linguistic encoding of the patient by 400 ms. Effects of Modifier Type and Length were observed again when linguistic encoding of the agent was completed (i.e., before speech onset).

Conclusions: The time-course results are consistent with a planning strategy where, early on (0-400 ms), speakers prepare conceptual and linguistic information for the sentence-initial character (the agent) but primarily conceptual information for the sentence-final character (the patient). The extent to which speakers plan the patient can be modulated but not completely determined by properties of the agent, suggesting that early message planning for SVO sentences can encompass conceptual information about all message elements.

[1] Klaus et al., 2016, *QJEP*. [2] Konopka & Meyer, 2014, *Cog Psych*. [4] Meyer et al., 1998, *Cognition*; [4] Opperman et al., 2008, *JML*.

**THE ROLE OF ALTERNATIVE CONSTRUCTIONS FOR QUANTIFIER SCOPE AMBIGUITIES:
A COMPARATIVE STUDY**

Barbara Hemforth (CNRS, Paris Diderot) & Lars Konieczny (Univ. Freiburg)
barbara.hemforth@linguist.univ-paris-diderot.fr

We will present a series of questionnaire experiments, showing that French, English, and German differ systematically with respect to the preferences for sentences that are ambiguous between linear scope and inverse scope interpretations of all-quantifiers and negation (1).

- (1) a. Alle meine Freunde sind nicht in den letzten Marvel-Film gegangen.
b. All my friends didn't go to see the last Marvel movie.
c. Tous mes amis ne sont pas allés voir le dernier film de Marvel.

We will argue that these differences are the result of differences in the grammar of the three languages under investigation here. Close alternatives to (1) exist for German and English which are only compatible with a wide scope interpretation of the negation (2), thus corresponding to the inverse scope reading of (1). The parallel alternative in French may occur in very informal spoken French but it is highly marked and not available in standard French.

- (2) a. Nicht alle meine Freunde sind in den letzten Marvel-Film gegangen.
b. Not all my friends went to the last Marvel movie.
c. ?? Pas tous mes amis sont allés au dernier film de Marvel.

We will moreover argue that the particular role of the preverbal position in German main clauses as a default topic position increases the linear scope preference compared to English.

Experiment 1: We presented participants (50 French, 34 fem, age 19-75, mean: 40.7; 59 English, 21 fem, age: 19-59, mean: 36; 57 German, 44 fem, age: 18 to 60, mean: 31.2) with 10 experimental sentences like (1a-c), simultaneously with two possible interpretations: i. linear scope interpretation: *None of my friends went to see the last Marvel movie.* ii. inverse scope interpretation: *Some of my friends went to see the last Marvel movie.* Participants had to judge how well each of the the interpretations fit the target sentence on a 5-point scale. All experiments were run on Ixet Farm, target sentences were mixed with 20 filler sentences.

French speakers showed a strong preference for inverse scope while German speakers strongly preferred linear scope. English speakers showed a slight preference for linear scope but less pronounced than German speakers (Figure 1, all ps < .01 in maximal linear mixed models).

Experiment 2: German and English differ from French with respect to a variety of parameters beyond the existence of an alternative constructions (e.g. a marked prosody for inverse scope). To test the *alternative construction hypothesis* directly, we ran a second experiment: The role of alternative constructions has been shown to depend on their accessibility in the language but also in the local environment (Authors, 2013, 2014, 2018). We replaced 10 of the filler sentences from Exp. 1 with sentences like *Not all packages weighed more than five pounds.* (with the following interpretations to judge: Linear scope interpretation: *Some packages weighed 5 pounds or less;* pragmatically enriched interpretation: *Some packages weighed more than 5 pounds.* This experiment was run with 60 English speakers. The presence of the alternative construction significantly increased linear scope preferences for English speakers (Figure 2, p<.01).

Conclusion: Our experiments show that the existence of alternative constructions as well as fine grained differences in the grammars can explain crosslinguistic differences not only in syntax but also in semantic interpretation.

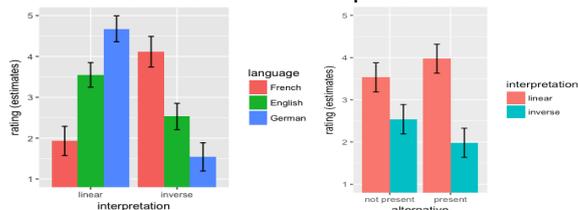


Figure 1

Figure 2

Friday Poster.60

TRAJECTORIES IN BILINGUAL PRODUCTION OF GRAMMATICAL GENDER AGREEMENT: LANGUAGE EXPERIENCE AND CROSS-LANGUAGE INFLUENCE

Hamutal Kreiner (Ruppin Academic Centre), Eva Smolka (University of Konstanz)

& Tamar Degani (University of Haifa)

hamutalk@ruppin.ac.il

Grammatical gender (GG) refers to the gender classification of inanimate nouns. In many languages GG requires agreement of associated sentential elements such as determiners and adjectives. Previous findings indicate that bilingual speakers are particularly challenged in producing GG agreement in their non-native language. This study examined to what extent the developmental trajectory of bilingual production of GG is modulated by factors operating at the long-term scale such as language experience and proficiency, and by factors associated with cross-language influence that operate on real-time.

Experiment 1 examined the acquisition of gender agreement in 130 kindergarten, first and second grade children, half monolingual Hebrew speakers, and half Russian-Hebrew bilingual speakers. Participants were presented with 48 pictures of inanimate objects selected such that their gender was congruent in Russian and in Hebrew for half the objects and incongruent for the other half (neutral gender was excluded). Participants produced a noun phrase in Hebrew that included the noun and a color adjective (e.g. blue_[masculine] flower_[masculine]). Overall, bilinguals showed more agreement errors than monolinguals. Critically, bilinguals produced more errors on incongruent than on congruent gender items, and agreement errors on incongruent nouns persisted even in second grade. By contrast, bilingual performance on congruent nouns improved with age and by second grade it was comparable to that of native speakers. Long-term factors such as L2 accumulated exposure and proficiency further affected agreement error rate.

Experiment 2 used a similar task to examine the effects of cross-language incongruency on the production of GG agreement by adult highly proficient bilinguals. Twenty-one native German speakers and 20 Russian-German highly proficient bilinguals produced a noun phrase in German that included a determiner, a color adjective, and the object name (e.g., *'eine blaue Blume'* - 'a_[masc.] blue_[masc.] flower_[masc.]'). Analysis of agreement errors on correctly named objects revealed no significant difference between bilingual and monolingual speakers, as expected for highly proficient bilinguals. However, bilinguals had more agreement errors on nouns with incongruent compared to congruent gender whereas monolinguals did not show such incongruency effect. Correct responses' latencies showed a similar pattern, and were further affected by long-term factors including switching behavior and L1-frequency.

Taken together these findings suggest that bilinguals may attain high proficiency in their L2, producing GG agreement as fast and as accurate as native speakers, yet their performance remains susceptible to cross-language interference. This seems to reflect two different processes operating on different time-scales: On the long-term scale, the general improvement trajectory indicates that the effect of reduced exposure to the target language of bilinguals as compared to monolinguals diminished with increased language exposure. However, on the moment-to-moment time-scale, dual activation of both languages leads to cross-language interference in case of incongruency even for highly proficient bilinguals. We will discuss the findings with regard to models of language production and models of bilingual language systems assuming dual-language activation.

Friday Poster.61

STATISTICAL LEARNING IN INFANTS, AND ITS RELATIONSHIP WITH LANGUAGE DEVELOPMENT: A STUDY OF NONADJACENT DEPENDENCY LEARNING

Rebecca L. A. Frost (Max Planck Institute for Psycholinguistics), Caroline Rowland (Max Planck Institute for Psycholinguistics), Samantha Durrant (University of Liverpool), Michelle Peter (University of Liverpool), Amy Bidgood (University of Liverpool) & Padraic Monaghan (Lancaster University)

Contact: rebecca.frost@mpi.nl

To acquire language, infants must learn how to spot individual words in speech, and master the constraints that govern the way those words are used. These tasks are highly complex; speech is rapid and continuous, and individual words can be combined in an infinite number of ways. Substantial evidence from language acquisition research indicates that implicit statistical computations could play an important part in these tasks (e.g. Saffran, Newport, & Aslin, 1996, Gomez, 2002).

Infants' capacity to use statistical information to identify words and structure has been suggested to be contingent on additional cues, such as pauses between words. Marchetto and Bonatti (2013; 2014) demonstrated that 12- and 18-month-olds could use nonadjacent transitional probabilities to identify words from continuous speech, but could only discern structural information about these words when speech was segmented. This is in line with prior findings with adults (Pena et al., 2002). However, recent advances in the adult literature indicate that adults can perform both of these tasks at the same time, in the absence of additional cues, when methodological confounds in the original studies are addressed (Frost & Monaghan, 2016).

We adapted the stimuli used by Marchetto and Bonatti (2013; 2014) in line with Frost and Monaghan (2016), and explored whether infants too can segment and generalise statistically-defined non-adjacent dependencies in the absence of additional cues, to shed light on the nature of these processes in infant language acquisition. Evidence of sensitivity to both word boundaries and structure would suggest infants may be able to perform both tasks together, using similar statistical processes.

We familiarised 71 17-month-old infants with a continuous stream of artificial speech, then examined their implicit knowledge of words and structure using an eye-tracked adaptation of the head-turn preference paradigm. At test, infants completed segmentation trials, which measured their preference for words (e.g. **bamuso**, **lifodu**) versus part-words (which straddled word boundaries, e.g. **solifo**), and generalisation trials, which examined their preference for part-words versus rule-words (a trained dependency intervened by a novel syllable, e.g. **baveso**).

Linear mixed effects analysis indicated that 17-month-olds could use nonadjacent dependencies to identify individual words in continuous speech, and could distinguish words from close competitors – building on Marchetto and Bonatti's (2013) demonstration that infants can use nonadjacencies to distinguish between words and previously unheard strings, indicating the powerful nature of these computations in infants. In addition, there was evidence to suggest infants were also sensitive to the structure these words contained. Critically, infants' performance on the statistical segmentation test was found to relate significantly to their vocabulary development (indexed by UK-CDI scores), indicating a link between these abilities in infant language acquisition. Further, the nature of this relationship differed for infants with high versus low CDI scores, who presented with novelty and familiarity preferences, respectively. This relationship will be discussed.

Friday Poster.62

COMPLEX MANDARIN MOTION EVENT DESCRIPTIONS ARE NOT SERIALIZATIONS OF VERBS: EVIDENCE AGAINST THE EQUIPOLLENTLY-FRAMED VIEW

Ziwei Li, Qili Wang & Johannes Gerwien (Heidelberg University)
Li@hulclab.eu

Researchers investigating motion event encoding have been trying to classify the languages of the world according to how frequently their speakers make use of two different lexicalization patterns. In the 'verb-framed pattern', the conceptual components MOTION and PATH are conflated in the verb root; in the 'satellite-framed pattern', MOTION and PATH are expressed in verb-external 'satellites' (Talmy 1985). For Mandarin, a language which provides its users with so-called serial-verb-constructions (SVC), a third category was introduced in order to capture the language-specific lexicalization pattern that goes along with these constructions: In the 'equipollently-framed pattern' (Slobin 2004), MANNER and PATH are expressed by equivalent linguistic elements. Given that each element in a construction like 'zou jin lai' (~ *walk enter come*) can occur alone in a sentence, Slobin argues, all three must be verbs. Furthermore, due to the lack of morphological markings, in his view, one cannot determine which one is the main verb in cases where two, or all three elements appear in one string. In this research, we use an auditory perception/recall test to evaluate Slobin's proposal by asking which components of a SVC native speakers are able to perceive as separate units, or which they automatically cluster together, respectively.

Mandarin native speakers (N=20) were asked to respond to 4 different types of linguistic strings (N=240, 60 of each type) by indicating whether the second element in each string was a word of Mandarin (yes/no button press) and if it was, to say it aloud. In type 1 strings, four units were combined, one encoding manner, one encoding path, one deixis, plus an aspect marker (e.g. *zou jin lai le*). Type 2 strings were also composed of four units, one encoding manner, one path, one encoding a goal, and one encoding deixis (e.g. *zou jin men lai* - *walk enter door come*). These two types are considered typical exemplars of Mandarin SVCs (Chen & Guo 2009). Type 3 strings were again composed of 4 units, one encoding manner, one encoding path, one non-Mandarin syllable, plus one unit encoding deixis (*zou jin */auf/ lai* ~ *walk enter */auf/ come*). Type 4 strings (control), also considered SVCs, were composed of 2 non-motion verbs (e.g. *shuo zan-zou* ~ *speak play*, or *chang tiao* ~ *sing dance*). In half of them, a non-Mandarin syllable was inserted after the first verb. In type 1, 2, and 3, all units were monosyllabic, but in type 4, also bisyllabic (Mandarin) units were used. All stimuli were audio-recorded as a whole by a native speaker of Mandarin. Non-Mandarin syllables were chosen from his L2 (German). Stimuli were presented via headphones in a randomized order.

We hypothesized that if the units that encode path and manner are indeed of the same morphosyntactic category, participants should be able to perceive them as separate units. Therefore, responses in type 1, 2, and 3 stimuli were expected to largely indicate 'yes'-responses and verbal reports were expected to refer to the path element as the second word.

In type 4 strings, participants reliably detected the non-Mandarin syllable (99% accuracy). When there was no non-Mandarin syllable, they correctly reported the second word in 98% of all cases. We take this to validate the method applied. In type 1, 2, and 3 strings, the element encoding path was never the most frequently reported perceived second word (type 1: 26%, type 2: 38%, type 3: 36%). The most frequent response (63%) for type 1 strings was PATH+DEIXIS, indicating that these elements are perceived as clustered information. In type 2 strings, participants reported the noun to be the second word in 38% of all cases, indicating the inability to separate the MANNER and PATH components. A similar conclusion is valid for type 3 strings, where participants responded that the second word was not a word of Mandarin in 63% of all cases.

These findings suggest that Mandarin speakers are not equally well able to perceive the morphemes encoding MANNER (first element) and PATH (second element) as separate units and that the degree to which they are, differs between the two different structures we tested. These findings cast doubt on the claim that the linguistic means that Mandarin speakers frequently use to describe motion events are indeed serializations of (main) verbs, which is in line with previous theoretical accounts (c.f. Lamarre 2007, Bisang 1992). Critically, the notion of equipollently-framing (Slobin 2003) depends on the assumption of verb-serialization.

PROCESSING SINGULAR THEY WITH GENERIC AND SPECIFIC ANTECEDENTS

Lauren Ackerman (Newcastle University)
lauren.ackerman@ncl.ac.uk

Although the use of *they* as a singular pronoun is well established, it is typically used in generic contexts, such as when the gender identity of the referent is unknown (e.g., 1a) [1,2]. Increasingly, singular *they* is used as a genderless or nonbinary pronoun, that is, as a personal pronoun for people who are known to be gender non-conforming (e.g., 1b). Previous processing studies have primarily focused on the generic use, and suggest there could be a cognitive benefit to using a genderless pronoun with referents of unknown gender as compared to known or biased gender [3,4,5]. This study investigates whether the specific use of reflexive singular *they* (i.e., for nonbinary people or referents whose gender identity is intentionally concealed) demonstrates the processing benefit for antecedents of unknown gender, akin to the observed processing benefit for generic antecedents of unknown gender.

- 1)
 - a. Anyone might be able to dress themselves every morning. (unbiased, generic)
 - b. Taylor might be able to dress themselves every morning. (unbiased, specific)
- 2)
 - a. A mechanic might be able to dress themselves every morning. (biased, generic)
 - b. Joseph might be able to dress themselves every morning. (biased, specific)

In an eye-tracking while reading task, participants (N=40) read target sentences (N=24) like those in (1) and (2), which were counterbalanced across conditions and interspersed with fillers from three unrelated experiments (total=98). After completing the task, participants answered standard demographic questions and reported judgments on how familiar they were with certain linguistic and social phenomena (e.g., singular *they*, LGBT communities). This information was used as fixed effects in models of reading times.

Of primary interest was reading behaviour at *themselves*, an established but nonstandard reflexive anaphor. Reading times at *themselves* can indicate initial antecedent retrieval, bonding between the anaphor and antecedent, and feature checking processes as the parser evaluates whether the anaphor and antecedent are compatible [6,7]. In the durations of first fixation on *themselves*, an interaction between antecedent gender bias and specificity is observed ($\chi^2(1)=5.1$, $p=0.025$). The generic/unbiased antecedent condition was read slower than generic/biased, whereas specific/unbiased antecedents were read slower than specific/biased. A main effect of specificity obtained in go-past reading times, with generic *themselves* read faster than specific ($\chi^2(1)=3.9$, $p=0.049$). There was also a marginal effect of antecedent gender ($\chi^2(1)=3.6$, $p=0.057$) and marginal interaction ($\chi^2(1)=3.4$, $p=0.067$), though both appear to be driven by a high mean reading time in the unbiased-specific condition. Both second pass (SP) and total time (TT) durations exhibit a main effect of specificity, with generic antecedents leading to shorter reading times than specific (SP: $\chi^2(1)=4.7$, $p=.03$; TT: $\chi^2(1)=5.9$, $p=.015$).

This study replicates previous work which observes a processing benefit for singular *they* with generic gender-unbiased antecedents (e.g., 1a) as compared to generic gender-biased antecedents (2a) in early reading times through an interaction of antecedent gender and specificity. This pattern does not extend to the specific condition (1b, 2b), in which it appears reading times on the anaphor were longer for the gender-unbiased names than the biased names. Overall, generic antecedents led to shorter reading times at the anaphor *themselves* than did specific antecedents. This is evidence that *themselves* as used for presupposed individuals (i.e., specific, named individuals) is processed differently than *themselves* used for generic individuals, regardless of whether the individual's gender identity can be determined by the antecedent. "Singular they" might therefore be best described as two phenomena: one for generic antecedents and one as a personal or specific pronoun. While most previous work has focused on the generic version, work on personal singular *they* can also inform how the gender identity of referents is conceptually represented and linguistically encoded. This study thus taps into the interface between linguistic, social and domain-general cognition.

References: [1] Baron. 1971. A reanalysis of English grammatical gender. *Lingua*. [2] Whitley. 1978. Person and number in the use of we, you, and they. *American Speech*. [3] Foertsch & Gernsbacher. 1997. In search of gender neutrality: Is singular *they* a cognitively efficient substitute for generic *he*? *Psychological Science*. [4] Doherty & Conklin. 2017. How gender-expectancy affects the processing of "them". *QJEP*. [5] Prasad, et al. 2018. The P600 for singular 'they': How the brain reacts when John decides to treat themselves to sushi. *CUNY poster*. [6] Van Gompel & Liversedge. 2003. The influence of morphological information on cataphoric pronoun assignment. *JEP: LMC*. [7] Sturt. 2003. The time-course of the application of binding constraints in reference resolution. *JML*.

Friday Poster.64

**MORPHOLOGICAL GENERALIZATION BEYOND SURFACE SIMILARITY:
ARGUMENT STRUCTURE AND INFLECTIONAL CLASSES IN HEBREW**

Yael Farhy (University of Potsdam)
farhy@uni-potsdam.de

Morphological generalization engages morphological operations or rules to generate inflected or derived forms of novel words that were not encountered before (such as *ploamphed* as the English past tense of *ploamph*). In previous research, frequency (what is the most common pattern for existing words?) as well as phonological and semantic similarity of novel items with existing word forms have been argued to be crucial determinants of morphological generalization (e.g., Rumelhart & McClelland, 1986; Ramscar, 2002). However, almost all previous studies on the topic involved Indo-European languages, which may have led to a neglect of other sources for morphological generalization.

Against this background, the present study examined morphological generalization in Hebrew, specifically the two most common Hebrew verb inflection classes (*Paal* and *Piel*). Our focus is on how argument structure modulates morphological generalization. In addition to native speakers of Hebrew we also examined proficient late bilinguals, to determine how morphological generalization is affected by speaker characteristics (Kam & Newport, 2009).

In two elicited production experiments every item consisted of two Hebrew sentences presented on a screen with a missing word in the second sentence. The participants (L1 and L2 Hebrew speakers) were asked to fill in the blank with a novel verb from the novel noun presented in the first sentence (e.g., novel noun *desel* [root *D-S-L*] → 7 possible answers, in line with the 7 verb classes: *dasalti* [*Paal*], *disalti* [*Piel*], *hidsalti* [*Hifil*] and so on). Experiment 1 (L1: n=28, L2: n=23) examined the effect of the novel root's phonological similarity with existing Hebrew roots on generalization. Three phonological similarity conditions included novel items derived from a simulation of the Analogical Modeling of Language (AML) framework (Skousen et al., 2002): (1) novel roots similar to *Paal*, (2) similar to *Piel* and (3) not similar to any class. For both groups, *Paal* and *Piel* responses were the most common (over 80% of all responses). In both groups the results showed an overall increase in *Piel* responses in the *Piel* similarity condition (39.4% of all responses in the condition) compared to the no-similarity condition (33.2%; $p=.014$) but not compared to the *Paal* similarity condition (35.5%; $p=.137$). Instead, *Paal* responses remained constant across similarity conditions (*Paal* Similarity: 52.8%, No Similarity: 53.9%; *Piel* Similarity: 48.9%). In Experiment 2 (L1: n=40, L2: n=30), argument structure was manipulated, with two conditions, one with a Direct Object (+DO) after the blank for the novel verb form and one without (-DO). *Piel* responses for the novel verb were expected to be affected by this manipulation, since existing *Piel* verbs (unlike *Paal* ones) are typically transitive (Berman, 1982). Also in Experiment 2 *Paal* and *Piel* responses were the most common (over 80% of all responses). The results showed a strong increase in *Piel* responses in the +DO condition (45.9%) compared to the -DO condition (28.7%), for both the L1 and L2 groups ($p<.001$), even though the argument-structure effect was significantly weaker for the L2 compared to the L1 group ($p=.008$). *Paal* responses remained constant across conditions (+DO: 42.2%, -DO: 43.6%).

The findings show that phonological similarity effects are less ubiquitous for morphological generalization than previously thought. For Hebrew inflectional class generalization they do not play any crucial role. Instead, argument structure is in this case a more relevant prediction of both L1 and L2 speakers' morphological generalizations. We conclude that the sources of morphological generalization are (at least in part) language-specific (Semitic-specific in this case).

**TOPICALITY IS NOT A PREREQUISITE OF TOPIC DROP:
EVIDENCE FROM A RATING STUDY ON GERMAN**

Lisa Schäfer, Robin Lemke & Ingo Reich (Saarland University)
lisa.schaefer@uni-saarland.de

Background. Huang (1984, p. 549) characterizes German as a “zero-topic but non-pro-drop language” allowing for Topic Drop (TD), i.e., the omission of an argument in preverbal position (the “prefield”) of a V2 sentence (see Fries 1988; cf. Haegeman 1997, 2007 for the similar phenomenon of “diary drop” in English and French). The term TD is used by several authors (e.g. Cardinaletti, 1990; Klein, 1993; Zifonun et al., 1997) who assume – explicitly or implicitly – that TD may only target arguments that are aboutness topics (Reinhart 1981). This assumption is questioned by Frick (2017) who presents evidence from a Swiss German text message corpus showing that this kind of omission also targets non-topical elements such as expletives. Since her corpus studies – in line with Auer (1993) – show significantly more cases of TD with 1st person singular (1SG) compared to 3rd person singular (3SG), she adduces factors such as identifiability and givenness which are correlated with grammatical person as trigger for TD. We aim at clarifying if there is a phenomenon justifiably to be called *Topic Drop*. For this purpose we analyzed the impact of topicality and person as well as of their potential interaction on TD in German empirically by conducting an acceptability rating study.

Method. In a 2 x 2 x 2 (TOPICALITY x PERSON x OMISSION) within subjects design, 48 participants (*Clickworker*) read items like (1) presented as text messages and rated the acceptability of the last utterance (in italics) on a 7-point Likert scale (7 = fully acceptable). We employed a preceding question (1a) to set the aboutness topic of the discourse (IV TOPICALITY) and manipulated the target utterance (1c) in two ways: For OMISSION the topic was either realized or omitted and for PERSON the topic pronoun was varied between 1SG and 3SG assuring that the grammatical person might be unambiguously reconstructed from the inflection of the verb in the target sentence.

- (1) a. A: was gibt's neues bei (tom | dir)?
A: *are there any news from (tom | you)?*
- b. B: seit neustem sehen tom und ich uns einmal die woche in der musikschiule.
B: *just recently, tom and i started seeing each other once a week at the musical school.*
- c. B: ((er) gibt mir | (ich) gebe ihm) jetzt klavierunterricht.
B: *((he) gives me | (i) give him) piano lessons.*

Results. Linear mixed effects models with random intercepts for subjects and items and by-subject and by-item random slopes for PERSON and OMISSION and their interaction (lme4, R) show no significant interaction between TOPICALITY and OMISSION. This indicates that TD is not more acceptable when the omitted subject has been established as topic. However, there is a significant interaction between PERSON and OMISSION ($\chi^2 = 10.986, p < 0.001$) suggesting that utterances with TD in 3SG are significantly less acceptable than those in 1SG.

Discussion. Our experimental data suggest that topicality is not a prerequisite of the preverbal omission of arguments, and, thus, question the adequacy of the term *Topic Drop* used by parts of the literature. In addition, our results – significantly higher acceptability ratings for the omission of the 1SG pronoun – are in line with the corpus findings of Auer (1993) and Frick (2017) that TD in 1SG is more frequent than in 3SG. Further research will show whether these differences in acceptability and in distribution in the corpus studies hinge on the disambiguation of the referent by verbal morphology, speech situation or givenness.

Selected references: Auer, P. (1993). *Zur Verbspitzenstellung im gesprochenen Deutsch*. Deutsche Sprache, 23:193–222 • Frick, K. (2017). *Elliptische Strukturen in SMS*. Berlin/Boston • Fries, N. (1988). *Über das Null-Topik im Deutschen*. Sprache und Pragmatik, 3:19–49 • Huang, C.-T. J. (1984). *On the distribution and reference of empty pronouns*. Linguistic inquiry, 15:531–574

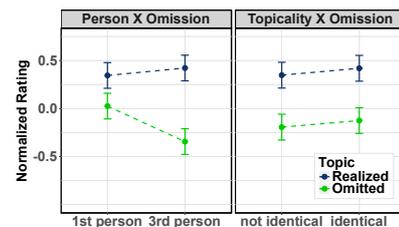


Fig. 1: Normalized ratings and 95% confidence intervals for PERSON x OMISSION and TOPICALITY x OMISSION.

DISCOURSE ACCESSIBILITY OF SEMANTICALLY INTEGRATED REFERENTS

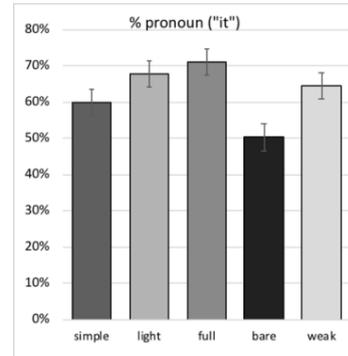
Eva Wittenberg (UC San Diego)
ewittenberg@ucsd.edu

We have many ways to package descriptions of events into linguistic form: “Morgan hugged Sophie” and “Morgan gave Sophie a hug” describe the same scene. However, the choice of construction may have linguistic and discourse-related implications. This project focuses on semantically integrated eventive nouns, like “hug”, in English, and asks whether and how the semantic integration of a noun into a predicate affects its later discourse accessibility at the level of both linguistic and conceptual representation. A strictly form-based view of discourse accessibility (Kamp, 1981) would not predict that degree of semantic integration modulates accessibility. The opposite prediction is made by a processing model that places the generation of event structures at its center (Ziegler et al., under review.) In this view, comprehenders create an event representation while comprehending discourse. If the linguistic event description contains a semantically incorporated noun, the noun itself is not encoded into the event model. The activation of the noun fades in working memory, and pronominal access to it becomes more difficult. Studies 1 and 2 test these predictions.

In **Study 1** (N=60), we used five constructions with varying levels of semantic integration (referents underlined in (1-5)) in a forced-choice production task (Scholten & Aguilar Guevara, 2010). Participants had to pick either a non-personal pronoun (“it”) or a full definite DP (“the note”) for (1); “the kick” in (2), etc.) in order to fill in the blank:

- (1) The teenager gave a note to his rival. ___ was rather mean. (full verb, referent: object)
- (2) The teenager kicked his rival. ___ was rather mean. (simple verb, referent: event)
- (3) The teenager gave a kick to his rival. ___ was rather mean. (light verb, referent: event)
- (4) Martin plays piano every day. ___ is unfortunately not very well tuned. (bare noun)
- (5) Claire had to go to the supermarket. ___ was just around the corner. (weak definite)

We replicated results by Scholten & Aguilar Guevara (2010), finding that pronominal reference to weak definites (5) and bare nouns (4) is dispreferred as compared to fully compositional structures ((1); $z > 2.06$, $p < 0.04$; see Figure). Simple verbs (2) patterned like bare nouns (4; $z < 2.01$, $p > 0.05$). The rate of pronominal reference for light verbs (3), which ‘unpack’ the event into an NP, was significantly higher compared to simple verbs ($z = -2.918$, $p < 0.01$), but it was only numerically less than for full verbs ($z = -0.98$, $p > 0.05$.) However, like Scholten & Aguilar Guevara (2010), we recognize the coarseness of forced-choice production data.



Study 2, a self-paced reading task (N=209), revealed more fine-grained distinctions. People read the pairs of sentences given by (1-3) but with blanks filled in with either demonstrative or non-personal pronouns (“that” or “it”). For “it”, people read fastest after full verbs (“give a note”), followed by reading “it” after light verb constructions (“give a kick”), and slowest reading “it” after simple verbs (“to kick”; strongest effect at the modifier (“rather”); e.g., main effect of Integration $\chi^2 > 12.4$, $p < .003$ using mixed-effects models and model comparisons; Levy, 2014). Thus, the degree of semantic integration modulates discourse accessibility in comprehension, too. We found a different effect for “that”: People read fastest after a light verb construction (3), compared to the other two conditions ($\chi^2 = 7.02$, $p < .04$). This fits well with observations that “that” has a general preference to refer to events (Çokal et al., 2008.)

Discussion. Both studies provide evidence that semantically integrated nouns are harder to access in discourse. These results need to be mapped onto a theory of the mental structure and processing of the discourse representation (Kaiser, 2016). We argue that these data are evidence for a model of language processing that places event construal at its center: comprehending a structure containing a semantically incorporated noun, the discourse structure that is encoded in memory is primarily that of the activity or event, with a lower resting activation of the noun (Ziegler et al., under review).

INTERFERENCE IN THE PROCESSING OF GRAMMATICAL SENTENCES: THE CASE OF MULTIPLE NEGATIONS

Iria de Dios-Flores (Universidade de Santiago de Compostela & Basque Center on Cognition, Brain and Language)
 iria.dedios@usc.es

This paper investigates the processing of grammatical sentences containing multiple negation markers. In Standard English, when two negations appear in the same clause they engage in a double negation dependency (DN), canceling each other out (e.g. (1c)). Strong disruption is expected in these cases, since DN is a dispreferred operation used only in restricted pragmatic contexts [1]. Conversely, cases like (1b) should be unproblematic since the two negative markers appear in different clauses and thus two independent propositions are negated only once. Yet, the results from this study provide evidence that the processing of the negative adverb *never* in (1b) is disrupted by the linearly preceding but structurally irrelevant negative quantifier *no*. The motivation for this investigation is to inform the landscape of hypotheses about a related phenomenon: the illusory licensing of negative polarity items (NPIs). Previous research has shown that unlicensed NPIs, like *ever* in (2), can be perceived as acceptable when occurring after a structurally inaccessible negation. A recent study [2] demonstrated that increasing the distance between the NPI and the inaccessible licenser cancels the illusion (a pattern that was not predicted by previous accounts [3,4]). Since it is generally assumed that NPIs are licensed by the compositional-semantic properties of entire propositions, in [2] they propose that illusions could arise as a consequence of unstable semantic representations available upon encountering *ever* in (2). That is, intrusion happens when the context containing the ineligible licenser has not been fully encoded by compositional semantic operations.

This hypothesis makes a prediction about the structures investigated here: the same unstable encodings that improve the perception of (2) could degrade the perception of (1b) inasmuch as it could be temporarily perceived as a DN dependency. We test this using three tasks: self-paced reading (fig. 1), speeded acceptability judgement (fig. 2), and untimed acceptability rating (fig. 3). The materials, adapted from [2] for maximal similarity, varied the presence/location of the negative determiner *no* with respect to *never*, resulting in the contrasts in (1a-c). The results are consistent across measures in that the processing of *never* in grammatical sentences like (1b) is disrupted when linearly preceded by *no*. Crucially, responses for DN conditions (i.e.(1c)) indicate a more degraded perception and slower recovery from disruption. It must be noted that (1b) is also rated quite low in offline measures (figure 3), which is interpreted as an instantiation of how processing complexity can result in the perceived unacceptability of grammatical sentences. Furthermore, these results are incompatible with the possibility that NPI illusions arise due to the misinterpretation of *ever* in (2) as *never* (i.e. (1b)), since continuations with *never* are shown to generate disruptions.

- (1) **a.** The authors [that **the** critics recommended] have **never** received a Pulitzer. (single negation)
b. The authors [that **no** critics recommended] have **never** received a Pulitzer. (≠ clause negation)
c. No authors [that **the** critics recommended] have **never** received a Pulitzer. (double negation)
 (2) *The authors [that **no** critics recommended] have **ever** received a Pulitzer. (NPI illusion)

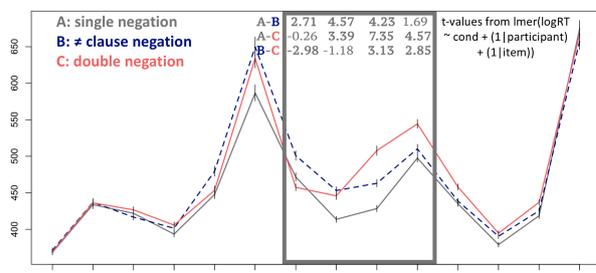


Fig.1: mean rt for non-cumulative window SPR task. N=36

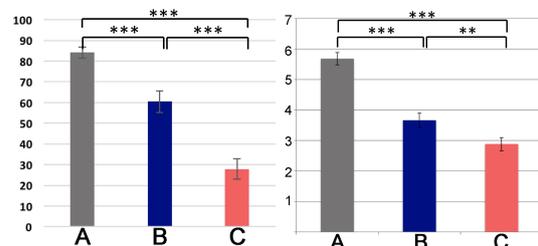


Fig.2: mean % of YES responses in the speeded acceptability task. N=28.

Fig.3: mean rating from a 1-7 untimed acceptability rating. N=24.

References: [1] Larrivé. 2016. The markedness of double negation. [2] Parker & Phillips. 2016. Negative polarity illusions and the format of hierarchical encodings in memory. [3] Vasishth, Brüssow, Lewis & Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. [4] Xiang, Dillon & Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence.

CAN ENTROPY EXPLAIN SUCCESSOR SURPRISAL EFFECTS IN READING?

Marten van Schijndel (Johns Hopkins University) & Tal Linzen (Johns Hopkins University)
vanksy@jhu.edu

Reading times (RTs) are influenced by the surprisal (predictability) of upcoming material that has not yet been fixated (*successor effects*; Kliegl et al., 2006). Surprisingly, successor effects have been found even in paradigms in which the upcoming word is not visible, not even parafoveally (Angele et al., 2015; van Schijndel and Schuler, 2017). Angele et al. hypothesized that successor surprisal predicts RTs because it approximates reader uncertainty about upcoming observations (i.e. entropy), which might underlyingly affect RTs.

To test this hypothesis, we derived surprisal and entropy estimates from a long short-term memory (LSTM) language model trained on 90 million words of English Wikipedia. Successor surprisal is the negative log probability of the word which actually occurred after the current word, and entropy is the expected value of these successor surprisals. We evaluate these predictors against self-paced RTs from the Natural Stories Corpus (Futrell et al., 2017).

The Pearson correlation between entropy and successor surprisal was $r = 0.45$: a considerable correlation but far from 1. It is still possible that the shared component of the two variables explains the effect of successor surprisal on RTs, however. We tested whether this is the case by entering the RTs into a linear mixed-effects model with entropy and successor surprisal as predictors, along with the surprisal, sentence position and length of the current word.¹ Successor surprisal and entropy both predicted RTs (entropy: $\hat{\beta} = 4.87, p < 0.001$; successor surprisal: $\hat{\beta} = 3.47, p < 0.001$); this suggests that the effect of successor surprisal cannot be reduced purely to entropy.

So far we have assumed that readers' uncertainty is based on their estimates of the probability of the entire vocabulary. Inspired by bounded rationality (Simon, 1982), we next consider the possibility that readers' uncertainty only takes into account the K most likely next words. Can entropy explain the effect of successor surprisal when computed only over those words?²

Table 1 shows the Pearson correlation between entropy and successor surprisal as a function of K . The correlation was *weaker* as entropy was computed over smaller K . Likewise, entropy was a weaker predictor of RTs as K decreased (Table 2), suggesting that humans are sensitive to uncertainty over a large set of possible continuations. Across values of K , the regression coefficient for successor surprisal was inversely related to the coefficient for entropy: successor surprisal is a better predictor when entropy is calculated over a smaller number of items. This supports an intermediate position, where some but not all of the success of successor surprisal in accounting for RTs is due to its correlation with entropy.

In summary, we have shown that entropy and successor surprisal are both robust predictors of RTs, regardless of whether uncertainty is calculated over the full vocabulary or only the most likely upcoming words. This suggests that entropy alone is unlikely to be the full explanation for successor surprisal effects.

	r		$\hat{\beta}_H$	$\hat{\sigma}_H$	$\hat{\beta}_s$	$\hat{\sigma}_s$
Best-5	0.212	Best-5	3.1940	0.6894	3.9566	0.5325
Best-50	0.335	Best-50	3.4326	0.7030	3.8539	0.5372
Best-500	0.397	Best-500	4.1081	0.6917	3.6624	0.5381
Best-5K	0.434	Best-5K	4.6732	0.6975	3.5206	0.5390
Total (50K)	0.454	Total (50K)	4.8664	0.7003	3.4727	0.5399

Table 1: Correlation between successor surprisal and entropy when entropy is computed over the best K continuations.

Table 2: Fixed effect coefficients for entropy (H) and successor surprisal (s) on self-paced RTs over the baseline.

¹The model formula was: $RT \sim \text{word_length} + \text{sentence_position} + \text{surprisal} + \text{entropy} + \text{succ_surprisal} + (1|\text{word}) + (1 + \text{word_length} + \text{sentence_position} + \text{surprisal} + \text{entropy} + \text{succ_surprisal}|\text{subject})$

²Our language model had a vocabulary of 50000 words, so entropy previously used $K = 50000$.

PUTTING PREFIXES IN FRONT: MORPHOLOGICAL PRIMING IN L1 AND L2 GERMAN

Laura Anna Ciaccio (University of Potsdam)
 ciaccio@uni-potsdam.de

While a large number of masked priming studies have investigated the processing of suffixed words, both in native (L1) and non-native (L2) speakers, much less is known on prefixed words. Current models of morphological processing make different predictions about how prefixed words are processed. Models that posit stem identification as the key mechanism underlying morphological processing predict a processing disadvantage for prefixed compared to suffixed words because, in prefixed words, the stem is not the first element of the word (see the results by Bergman, Hudson, & Eling, 1988; Colé, Beauvillain, & Segui, 1989; Kim, Wang, & Taft, 2015). In contrast, affix-stripping models (Rastle, Davis, & New, 2004; Taft, 1979; Taft & Forster, 1975), according to which affixes are stripped off pre-lexically, predict that prefixed and suffixed words are processed in the same way. In the present study, we directly compare priming from prefixed and suffixed words on the same targets, in L1 and L2 speakers of German. Morphological priming was tested with Germanic and Latinate items, for a total of 24 targets (see examples). We additionally compared morphological priming to orthographic and semantic priming. In the mean RTs, we found facilitation on target recognition following both prefixed and suffixed primes, with both Germanic and Latinate items and in both groups. We fitted a linear mixed effect model to log-transformed RTs including the fixed factors Prime Type, Set, Group, and their interactions. The model showed main effects of Prime Type for both prefixed and suffixed primes (both $t_s > 7.76$), with no difference between them ($t = 0.53$), no two-way interactions between Prime Type and Set or Prime Type and Group, and no three-way interactions (all $t_s < 1.21$), suggesting that morphological priming effects were comparable for prefixed and suffixed primes, across the two sets and the two groups. Additional models comparing morphological priming to orthographic and semantic priming revealed that morphological priming effects were larger than effects of form or meaning overlap (all $t_s > 2.20$). Our results indicate robust morphological priming effects of the same magnitude for both prefixed and suffixed words, which are consistent with the predictions of affix-stripping accounts of masked morphological priming. Furthermore, in line with previous research on L2 processing of suffixed derived words, our results demonstrate native-like derivational priming effects in L2 speakers, not only for suffixed words (e.g. Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Jacob, Heyer, & Veríssimo, 2017), but also for prefixed words.

Examples of morphologically related prime-target pairs:

Germanic *unsauber* 'dirty' and *Sauberkeit* 'cleanness' - *sauber* 'clean'
 Latinate *inaktiv* 'inactive' and *Aktivität* 'activity' - *aktiv* 'active'

Set:	L1 group (N=48)		L2 group (N=48)	
	Germanic	Latinate	Germanic	Latinate
Unrelated Prime				
RT (SD)	625 (133)	631 (138)	754 (211)	716 (209)
Accuracy	93.1%	87.9%	89.1%	85.9%
Prefixed Prime				
RT	610 (152)	614 (147)	716 (210)	699 (233)
Effect	15	17	38	17
Accuracy	97.9%	90.5%	91.3%	88.4%
Suffixed Prime				
RT	600 (132)	604 (142)	732 (232)	689 (214)
Effect	25	27	22	27
Accuracy	95.8%	92.9%	90.6%	89.9%

Friday Poster.70

GENDER-BIASES IN LANGUAGE PROCESSING: EXPLICIT BELIEFS ABOUT EVENT OUTCOMES VS. IMPLICIT LINGUISTIC EXPECTATIONS

Titus von der Malsburg (University of Potsdam, MIT), Veronica Boyce (MIT), Till Poppels (UCSD), Roger Levy (MIT)

As we speak, write, listen, and read, information sources ranging from long-term stereotypes to recent events and the specific discourse context are rapidly recruited to form expectations about the unfolding linguistic expression. A fundamental open question is how these sources of information are reconciled when they are in conflict. For instance, a sentence like *'The surgeon pricked herself with a needle'* elicits measurable surprise at *'herself'*, even though female surgeons are increasingly common [1,2,3,4,5]. We report results from three large-scale online studies that investigated how gender-stereotypical knowledge and beliefs about event outcomes jointly shape linguistic performance.

Study 1 (N=24,863) was conducted during the 2016 U.S. presidential campaign. It tested the relative contributions of gender-stereotypical knowledge and unfolding explicit beliefs about the outcome of the election (female or male winner) to the production and comprehension of statements about the future president (data collection from June 2016 to January 2017). In a production task, US citizens completed text fragments that were designed to elicit pronominal references to the future president (*'After the inauguration in January 2017, the secret service will protect the US president and . . . his/her/their family'*). Language comprehension was tested with self-paced reading and 132 mini discourses which contained masculine/feminine/neutral pronouns referring to the future president. While the production of male and neutral pronouns (singular *'they'*) was modulated by relative beliefs about the electoral outcome (measured in a third task), feminine pronominal references were consistently disfavored in both production and comprehension. This was true even at times when a female president was perceived as the most likely election outcome. This pattern held across the complete demographic and political spectrum. Thus, even in the face of recent and salient countervailing information, implicit biases can dominate language production and comprehension above and beyond biases already factored into our beliefs about specific real-world event outcomes.

How stable are these stereotypes? Study 2 (N=2,609, UK citizens) had an overall similar design but was conducted in early 2017 during the run-up to the UK general election which decided the next prime minister. In this study, we found that gender-neutral pronouns (singular *'they'*) were strongly favored in production and comprehension. Masculine and feminine reference were rarely produced and elicited much longer reading times than neutral references, presumably suggesting that salient counter-examples to a stereotype (PMs Thatcher, May) may go a long way toward weakening the influence of gender-stereotypes on linguistic behavior.

Do the gender-biases observed in Study 1 generalize across a larger set of roles nouns that exhibit gender biases of various strengths and polarities? Study 3 (N=717, US citizens) measured gender-biases associated with 80 common role nouns (e.g., *'barber'*, *'diplomat'*, *'baker'*, *'ice skater'*, *'babysitter'*) and tested production and comprehension preferences when these nouns were embedded in texts (*'Before the manicurist took out the trash, . . . he/she/they . . .'*). While production preferences were strongly modulated by stereotypical knowledge, we also found a general dispreference for feminine pronouns in production (30% less feminine than masculine pronouns for equi-biased nouns) that held even for role nouns with strong female-bias.

In sum, our results demonstrate that gender-stereotypes as well as general linguistic gender biases can strongly shape linguistic behavior even when they are in conflict with current beliefs about the world. More generally, this work demonstrates the potential of psycholinguistic methods combined with large-scale online experimentation for investigations of how changing world knowledge is integrated over time into specialized processing domains such as language, and for revealing and deepening our understanding of implicit biases.

[1] Sturt, JML (2003), [2] Kennison & Trofe, J of Psych Res (2003), [3] Hale, NAACL (2001), [4] Levy, Cognition (2008), [5] Duffy & Keir, Mem & Cog (2004)

CHILD-LIKE ADULTS: TESTING DISTRIBUTIVITY USING A DUAL TASK

Anna de Koster, Jennifer Spender and Petra Hendriks (University of Groningen)
a.m.b.de.koster@rug.nl

Distributive contexts (Fig 1) seem to require distributive markers like ‘each’, e.g. (1). Adults generally reject distributive readings when no distributive marker is present, e.g. (2) preferring collective interpretations (Fig 2)^{[1][2]}. Adults also tend to reject the use of distributive markers as in (1), when collective meanings (Fig 2) are intended.



Fig. 1 Distributive



Fig. 2 Collective

- (1) Each boy is building a snowman.
- (2) The boys are building a snowman.

Young children initially accept both readings with both sentences. Adult preferences develop late^{[2][3][4]}: children reject ‘each’ with collective readings (*Each-Col*) at age 7;0 but only reject ‘the’ with distributive readings (*The-Dist*) at age 11;0^[4]. Dotlačil^[3] argues that adult collective preferences for distributively unmarked sentences arise via implicatures. Non-distributive DPs like ‘the’ get interpreted collectively, because if the speaker intended distributivity, they would have used ‘each’. Implicatures require reasoning about alternatives, for which working memory (wm) capacity is necessary. De Koster et al.^[4] found a significant correlation between wm capacity and children’s *The-Dist* rejection rates, supporting this explanation. Given both the theoretical predictions^[3], and experimental results^[4] we predict loading adults’ wm in a dual task will limit implicature calculations, and increase acceptance rates for *The-Dist*.

Method 40 Dutch adults (MA:22) judged the truth values of 64 sentences like (1) and (2) (TVJT) (using the Dutch *elke* as a translation for *each*) combined with distributive (Fig 1) and collective (Fig 2) pictures while keeping 3 digits (low memory load) or 6 digits (high memory load) in memory. 3- or 6 digits were presented in different blocks, with block order as a

between subjects factor. Mixed effect models were used for analysis.

Results TVJT Responses Strikingly, the *The-Dist* condition acceptance rate for both 3- and 6-digits, is much higher than the control condition without a wm-load (16 participants, MA:24), showing a large effect of loading working memory. In fact, we found no significant difference between 3- and 6-digits. Even a wm-load of 3 digits already resulted in an acceptance rate of *The-Dist*, similar to children^[4]. Further, the rate of acceptance of *Each-Col* was unexpectedly high as well, also seemingly affected by wm loading. Conditions *Each-Dist* and *The-Col* were at ceiling (both 99% acceptance).

Table 1. Acceptance rates for a low and high wm-load, plus a control group

	3 Digits	6 Digits	Control
The-Dist	80%	78%	41%
Each-Col	57%	57%	32%

Results Reaction Times (RT) Mean RTs for *The-Dist* and *Each-Col* were both significantly longer than other conditions, with no significant difference between 3- and 6-digits. Longer RTs for *The-Dist* are consistent with time needed for implicature calculation^[5]. The Longer RT for *Each-Col*, was unexpected, but likely related to Dutch ‘*elke*’ being less distributive than ‘*each*’.

Table 2. Average Reaction Times in ms

	3020	The-Col	2639
The-Dist			
Each-Col	3109	Each-Dist	2721

Our Conclusions Our results show that loading wm increases adults’ acceptance rates for *The-Dist*, supporting Dotlačil’s claim that *The-Dist* is rejected via an implicature. Our results also suggest children’s development towards adult-like responses may depend on wm capacity.

[1] Syrett and Musolino (2013) Collectivity, distributivity, and the interpretation of plural numerical expressions.[2] de Koster et al. (2017) Children’s understanding of distributivity and adjectives of comparison. [3] Dotlačil (2010) Anaphora and distributivity. A study of same, different, reciprocals and others.[4] de Koster et al. (2018) Are Children’s Overly Distributive Interpretations and Spreading Errors Related? [5] Bott and Noveck (2004) Some utterances are underinformative: The onset and time course of scalar inferences.

AGREEMENT ATTRACTION IN GERMAN SOV STRUCTURES: AN ERP STUDY

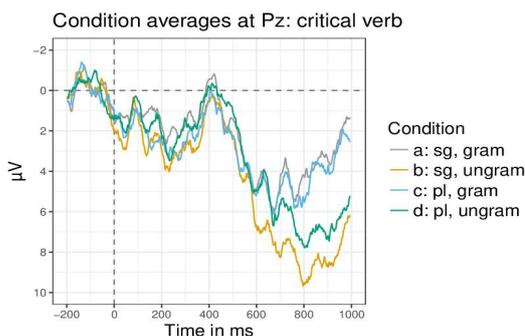
Robin Schäfer, Sol Lago & Titus von der Malsburg (University of Potsdam),
rschaefer@uni-potsdam.de

The efficient processing of subject-verb dependencies is a central component of language comprehension. Nonetheless it is surprisingly error-prone in some specific syntactic configurations. Consider the following example: **The chemist with the test tubes are conducting an experiment.* It has been found that the number mismatch at the verb leads to reduced processing difficulty when an interfering noun, the so-called attractor (*test tubes*), bears the same number marking as the ungrammatical verb [1,2,3]. In comprehension, these effects have been attributed to errors during memory retrieval. However, there are two types of retrieval mechanism that can potentially explain this effect: i) Either subject retrieval is *always* triggered at the verb in which case a feature mismatch could lead to the erroneous retrieval of the attractor noun [1]. ii) Alternatively, misretrieval may *only* happen during an error correction process triggered by an ungrammatical, number-mismatching verb [1,2]. To disambiguate between these two accounts, we conducted an ERP study in German that addressed the role of attraction in modulating the P600 component, which has been previously linked to error correction mechanisms [e.g., 4]. We hypothesized that only under the error-driven account of attraction, the P600 would be modulated by the occurrence of an attractor in the case of a number-mismatching verb.

Methods: Items (N=120) consisted of embedded SOV clauses. We manipulated the grammatical number of the object attractor and of the verb (sg/pl). Experimental items plus 140 filler sentences were read by 33 participants in RSVP mode (SOA = 450ms). Linear mixed models with maximal random effects structures were used to analyze the average potential at electrode Pz in the time window from 600 to 1000ms post-onset of the verb.

	subject	attractor	adv	adv	verb
a.	Pia erzählt, dass der Mann _{NOM/SG}	die Frau _{ACC/SG}	gestern	heimlich	beobachtete _{SG} , ...
b.	*Pia erzählt, dass der Mann _{NOM/SG}	die Frau _{ACC/SG}	gestern	heimlich	beobachteten _{PL} , ...
c.	Pia erzählt, dass der Mann _{NOM/SG}	die Frauen _{ACC/PL}	gestern	heimlich	beobachtete _{SG} , ...
d.	*Pia erzählt, dass der Mann _{NOM/SG}	die Frauen _{ACC/PL}	gestern	heimlich	beobachteten _{PL} , ...

'Pia says that, yesterday, the man secretly watched the woman/women ...'



Results: Ungrammatical sentences elicited a highly reliable P600 effect ($b=3.00$, $t=8.82$). A reliable interaction of grammaticality and attractor number indicated that the P600 effect was reduced in the presence of a plural marked attractor ($b=-2.06$, $t=-2.73$). These results lend support to the reanalysis account of agreement attraction proposed in [1,2]. Under the alternative retrieval account which does not assume error correction, we did not expect a modulation of the P600 in response

to ungrammatical sentences. Our results extend the findings by [1,2], by demonstrating attraction effects consistent with error-correction in a different syntactic configuration. Our results also extend [3,5] by showing ERP effects of agreement attraction in another language and [3] by showing them in a different syntactic configuration (SOV vs PP-mod construction).

References: [1] Wagers, Lau & Phillips (2009). *J Mem Lang*. [2] Lago, Shalom, Sigman, Lau & Phillips (2015). *J Mem Lang*. [3] Tanner, Nicol & Brehm (2014). *J Mem Lang*. [4] Metzner, von der Malsburg, Vasishth & Rösler (2016). *Cog Sci*. [5] Kaan (2002). *J Psycholinguist Res*.

CAN GAPPING BE EMBEDDED? A CROSSLINGUISTIC PERSPECTIVE

Gabriela Bîlbîie (Univ. of Bucharest), Pegah Faghiri (Univ. of Cologne), Israel de la Fuente (Univ. of Lille & STL) & Anne Abeillé (Univ. Paris 7 & LLF)
 gabriela.bilbie@gmail.com

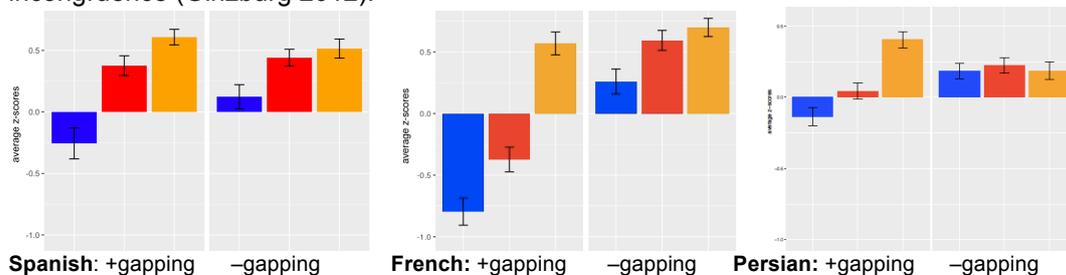
It is usually assumed (Hankamer 1979, Neijt 1979, Johnson 2009) that gapping cannot be embedded (1a). However, Farudi (2013) claims that Persian is an exception (1b) and Garcia Marchena (2015) has Spanish examples from a spoken corpus.

- (1) a. *Alfonso stole the emeralds, and I think [that Mugsy the pearls]. (Hankamer 1979)
 b. Mahsā in ketāb-ro dust dār-e va Minu mi-dun-e [ke māmān-esh un ketāb-ro].
 Mahsa this book-OBJ like have-3SG and Minu IMPFV-know-3SG that mother-3SG that book-OBJ
 'Mahsa likes this book and Minu knows that her mother (likes) that book.' (Farudi 2013)

We show, based on three acceptability judgment tasks for Spanish, French and Persian, that there is crosslinguistic variation with respect to embedded gapping, and that a more general semantic constraint seems to be at work: non-factive verbs embed more easily than factive ones (Kiparsky & Kiparsky 1970, Karttunen 1971), independently of ellipsis and language type. We ran 3 experiments (24 experimental items and 24 control items each), using a 2x3 design (gapping, embedding-nonfactive, embedding-factive), with similar materials in each language. The rating scale was 1-10 in Spanish and French, and 0-10 in Persian. 50 participants completed the Spanish experiment, 48 for French, and 116 for Persian.

- (2) a. [+gapping, +embed, +factive]
 (S) En el bar, Pablo pidió una cerveza y **me molesta que** Juan (pidiera) un whisky.
 (F) Au bar, Paul commande une bière et **ça m'ennuie que** Jean (commande) un whisky.
 'At the bar, P. orders a beer and it bothers me that J. (orders) a whisky.'
 b. [+gapping, +embed, -factive]
 (S) En el bar, Pablo pidió una cerveza y **sospecho que** Juan (pidió) un whisky.
 (F) Au bar, Paul commande une bière et **il semble que** Jean (commande) un whisky.
 'At the bar, P. orders a beer and it seems that J. (orders) a whisky.'
 c. [+gapping, -embed]
 (S) En el bar, Pablo pidió una cerveza y Juan (pidió) un whisky.
 (F) Au bar, Paul commande une bière et Jean (commande) un whisky.
 'At the bar, P. orders a beer and J. (orders) a whisky.'

We only report significant results ($p < 0.01$) using linear mixed-effects models. In Spanish, embedded gapping is as acceptable as embedded non-gapping under non-factive verbs. In all three languages, there is an interaction between gapping and embedding, but embedded gapping (mean z-scores 0.06 in Spanish, -0.58 in French, -0.06 in Persian, for both factives and non-factives combined) is more acceptable than the ungrammatical controls (mean z-scores -0.93 in Spanish, -1.27 in French, -1.70 in Persian). Moreover, factivity is significant in all languages: embedded clauses under a factive verb are less acceptable than under a non-factive verb. Interestingly, this effect is not correlated with ellipsis in French (no significant difference between gapping and non-gapping). We conclude that the No Embedding Constraint on gapping cannot be maintained. Moreover, in all these languages, embedded clauses are sensitive to the semantic class of the embedding verb. The difficulty with coordinating a simple clause and a complex clause may result from a more general parallelism constraint on coordination (Frazier & Clifton 2000) and the further penalty on factive verbs may come from their non-assertive nature (Hooper 1974) and/or from the QUD-incongruence (Ginzburg 2012).



Friday Poster.74

EVIDENCE FROM EYE-TRACKING SHOWS QUALITATIVELY SIMILAR PROCESSING OF NOVEL ITEMS BY L1 AND L2 SPEAKERS

Anna Tsiola (University of Illinois at Urbana Champaign) & Kiel Christianson (University of Illinois at Urbana Champaign)

tsiola2@illinois.edu

This study uses eye-tracking to examine L1 and L2 novel item processing in terms of lexical access and semantic integration. L2 processing is generally slower (e.g., Moreno & Kutas, 2005) and shows increased difficulty with semantic integration (e.g., Moreno et al., 2008). Chaffin et al. (2001) showed that L1 speakers can establish the meaning of novel items from context, and Pellicer-Sánchez (2016) found a slower rate of learning new items for L2 than for L1 speakers, potentially due to difficulties with inferring meaning from contextual cues. It is still unclear whether slower L2 processing is the result of problematic lexical access (e.g. Moreno & Kutas, 2005) or semantic integration (Hahne & Friederici, 2001).

This study asks whether L1 and L2 speakers differ in (1) the way they attempt to access and integrate novel items, (2) their attention to contextual cues to derive their meaning, and (3) how many novel items they can recall. 42 native English speakers and 41 L1-Chinese, L2-English speakers participated in an eye-tracking experiment inducing incidental learning conditions, followed by a receptive vocabulary post-test. Each stimulus contained 2 sentences presented together in separate lines; Sentence 1 contained the target word (novel/familiar) and an informative/uninformative context (resulting in a 2x2 repeated design). Sentence 2 included a repetition of the target word and its hypernym. Below is a stimulus example from the novel word-informative context condition:

He picked up **the barhep**_{word1} from the floor **to play some music**_{context} and frowned.

He had realized that **the barhep**_{repetition} is a difficult **instrument**_{hypernym} to learn by yourself.

Mixed effects models were used to analyze early (first fixation duration, gaze duration) and late (total time, regression path duration) eye-tracking measures on the 4 interest areas (bolded above), to examine lexical access and semantic integration processes respectively. These measures allow us to locate group differences in early and late cognitive processes and tease apart difficulties due to problematic lexical access or semantic integration.

Both groups had similar reading patterns in measures indexing lexical identification, lexical access, and semantic integration, with inflated reading times for novel words and for contexts that were informative and thematically linked to the preceding target word, as readers were making the word+ context meaning connection. The groups differed in semantic integration at the sentence level. The L2 group showed increased reading times at informative contexts with novel words, whereas the L1 group showed this cost only with familiar words. Considering that there was a documented integration cost with familiar words, the L1 group's pattern suggests that they did not actively attempt semantic integration from contextual cues but relied on the hypernym, which was fixated longer when the word had been novel and the context uninformative. The L2 group were actively making meaning but were less efficient in allocating their attention to the relevant areas in the text. Finally, L1 and L2 speakers retained a similar (very small) number of items in the post-test, (L1 $M=0.09$, range: 0-7, L2 $M=0.07$, range: 0-5, $t(1646.4) = 1.5$, $p = 0.13$).

These findings suggest (1) that L2 speakers' processing of novel items patterns mostly in the same way as L1 speakers', and (2) the group differences in the "context" region indicate that L1 and L2 speakers may develop different strategies or approach the task as incidental vs intentional learning respectively. Preliminary data from a follow-up think-aloud experiment support group differences in depth of processing (Craik 2002). In sum, the process of deriving the meaning of a new word from context, adding it to the mental lexicon, and retrieving it later, does not differ substantially between L1 and L2 speakers, suggesting alternate sources for the quantitative differences commonly observed in the literature.

Friday Poster.75

THE EFFECT OF EVENT DEPICTIONS ON SECOND LANGUAGE LEARNING

Huong Thi Thu Nguyen, Katja Münster, Carsten Schliewe, Pia Knoeferle
Humboldt-Universität zu Berlin
nguyetxh@hu-berlin.de

This research is inspired by visual context effects on real-time language processing in adults (Knoeferle et al., 2005; Tanenhaus et al., 1995) and children (Münster, 2016; Trueswell et al., 1999). In addition, it is also motivated by the important roles of visual context in studies on language learning as well as mechanisms such as cross-situational language learning (Koehe et al., 2015; MacDonald et al., 2017; Yu et al., 2011) and learning in multimedia contexts (Mayer, 2005).

We investigated whether German (the first language – L1) adults experience facilitation in learning Vietnamese (the second language – L2) phrases from (a) event photograph presence (present vs. absent) and (b) verb mapping (similar vs. different between L1 & L2), yielding four learning conditions (present-similar; absent-similar; present-different; absent-different). German native speakers (N=32/experiment, L1=German, no L2 < age 6; ages 18-31 in exp1 & 2 and 32-45 in exp3, no prior knowledge of Vietnamese) participated in three experiments. **Experiment 1: Training:** For each training trial, participants listened to two nouns and two complete verb-noun phrases: The nouns occurred together with a photograph of its referent (e.g., *sách/ book* & *bàn/ table*), displayed on a computer screen. In the event-present conditions only, each verb-noun phrase was further accompanied by an event photograph (e.g., *đọc-sách/ reading book* & *lau-bàn/ cleaning table*) during training. **Testing:** For each testing trial, participants performed a binary forced-choice photograph selection task to test L2 learning. They inspected two object photographs again before listening to a verb only (e.g., *đọc/ read* or *lau/ clean*) related to one of them; then they had to complete the verb fragment by selecting one of the two object photographs. The procedure of experiment 1 included two similar parts (Each part: learning and immediate post-trial testing). **Experiment 2:** We re-used the learning method from experiment 1 and tested participants with the same requirements as in the first experiment; however, we expanded the language materials to avoid repeating each verb twice in the whole experiment. Moreover, in testing, participants listened to a full verb-noun phrase (e.g., *đọc-sách/ reading book*) before inspecting two event photographs (e.g., *đọc-sách/ reading book* or *chọn-sách/ choosing book*) and selected the matching one out of these two photographs. Experiment 2 had 3 parts (Part 1: learning and post-trial testing; Part 2: only re-learning all trials; Part 3: only testing the generalization of items by presenting new photographs). **Experiment 3** assessed whether the results from experiment 2 replicate with a slightly older participant group (35-45 years of age) to assess age-related variation in learning.

We measured participants' response latencies and accuracy in selecting one of two photographs following learning. In experiment 1, no reliable effects of the manipulated variables emerged. With a changed testing context, however, in experiments 2 & 3, participants were faster and more accurate when event photographs were present versus absent. Moreover, these effects changed across experiment part / learning (less benefit when testing was delayed and required generalization to new photographs, as was the case in Part 3 of experiments 2 & 3 in comparison with the immediate post-trial testing on the training photographs in Part 1). While main effects of event presence and experiment part on accuracy as well as reaction times replicated, other effects (of language mapping) did not replicate (in experiment 3) which might be caused by age-related differences. In conclusion, event photographs can substantially enhance Vietnamese verb-noun phrase learning but only in a suitable learning and testing context (comparing exp1 with exp2). Furthermore, changes in the results of learning second language phrases (comparing exp2 with exp3) suggest that age-related variation may play a role in the learning outcome.

**THE ROLE OF EXPECTATIONS IN REFERENTIAL AMBIGUITY PROCESSING:
EVIDENCE FROM RUSSIAN**

Veronika Prokopenya (St.Petersburg State University) & Ekaterina Saenko (St.Petersburg State University)
v.prokopenya@spbu.ru

Reference resolution, i.e. the process of relating a linguistic expression to an object of reality, is one of the key aspects of linguistic processing. In a discourse, the same object or person may be mentioned numerous times with the use of different expressions, including pronouns – thus, listener or reader has to link a particular expression to the one used before. Referential ambiguity arises in those cases when a pronoun may be interpreted as relating to several mentioned referents. Previous studies have shown that in particular contexts the initial interpretation of such a pronoun may not be fully disambiguated, i.e. the shallow processing may take place.

The aim of our eye-tracking research was to find out how referentially ambiguous sentences are processed during reading and whether shallow processing is indeed possible until the disambiguation occurs, or the pronoun is immediately linked to one of the referents, and what is the role of readers expectations. Two groups of experimental stimuli were constructed: (1) equal anticipation of both referents: (2) first referent is anticipated more than the second one (these predetermined expectations were confirmed in the pretest):

(1) *Tanya spokojno ehala po ulitse, kogda Nadya neozhidanno vyiletela iz-za ugla na rollovak. K schastyu, ona zatormozila pochti momentalno^a, i Nadya/Tanya proehala spokojno^b, dazhe ne zametiv ugrozyi stolknoveniya. – Tanya was calmly moving down the street when Nadya suddenly came off the corner on rollerblades. Fortunately, she managed to stop quickly^a, and Nadya/Tanya moved on quietly^b without even recognizing the possibility of a crash.*

(2) *Ljusya silno obidela Lera na novogodney vecherinke v shkole. Ochevidno, ona pyitalas pomiritsya pervoy^c, no Lera/Ljusya otklyuchila telefon^d, tak kak vremeni dlya vvyasneniya otnosheniy ne bylo. – Lucy insulted Lera during New Year party at school. Obviously, she tried to mend fences first^c, but Lera/Lucy turned the phone off^d as there was no time for sorting out their relationships.*

There were 64 experimental items in total (32 per group), each presented in 4 conditions: 2 ambiguous and 2 unambiguous (referents of different gender) – where the pronoun referred either to the first or to the second character. 36 participants had to read experimental items at a comfortable pace and answer comprehension questions that randomly followed 1/3 of items. The 2x2x2 design was elaborated, with ambiguity (ambiguous vs unambiguous), referent (1st vs 2nd), and stimuli group (1 vs 2) as factors. The results differed between stimuli groups. For the group 1 no significant difference in reading times of a region containing a pronoun (a) was found between ambiguous and unambiguous contexts ($p > .05$), while the reading times of the disambiguating region (b), and the time spent on regressions to the previous text (go-past time) were significantly longer in ambiguous contexts ($p < .001$). For the stimuli group 2 significant increase in processing time (first-pass time) of a pronoun region (c) was found only in unambiguous cases when the pronoun referred to the less anticipated second referent ($p = .003$). Also, disambiguating region (d) reading times were significantly longer in ambiguous contexts compared to unambiguous ones ($p < .001$), moreover the pairwise comparison showed that the reading times in ambiguous contexts were significantly longer in those cases when the pronoun referred to less anticipated referent.

These results allow us to make the following conclusions. During reading of sentences with formally ambiguous pronouns but with predetermined expectations referential relations are established immediately, in favor of anticipated referent, in the same way as in the unambiguous contexts. Apparently, in the sentences with equally anticipated referents full reference resolution occurs at an early stage as well, i.e. the pronoun is immediately linked to one of the referents (no evidence for shallow processing was found), and this leads to the re-processing of a sentence if the primary interpretation turns out to be incorrect.

STRUCTURAL PRIMING OF THE GERMAN PASSIVE IN LANGUAGE PRODUCTION

Yvonne Portele (Goethe University Frankfurt)

The tendency of speakers to reuse a previously processed structure has been found for many linguistic structures in different languages. The relevance of these *structural priming* experiments has been reinforced by Branigan & Pickering (2017), who claim that they might provide us with evidence about linguistic representations. In order to do this, the effect of structural priming has to be disentangled from general influences on sentence production, some of them pooled under the notion of *inherent accessibility* (e.g. *animacy*).

German priming studies have mainly focused on dative alternations. If, however, structural priming is the robust mechanisms it has been claimed to be, we expect to find structural priming of the German passive. Curiously, there seem to be no published results showing this effect, the only indirect study being Loebell & Bock (2003) with non-significant effects.

I investigated this issue using a classical picture description paradigm (+ cover recognition task) using photographs of persons showing them in real-life situations:

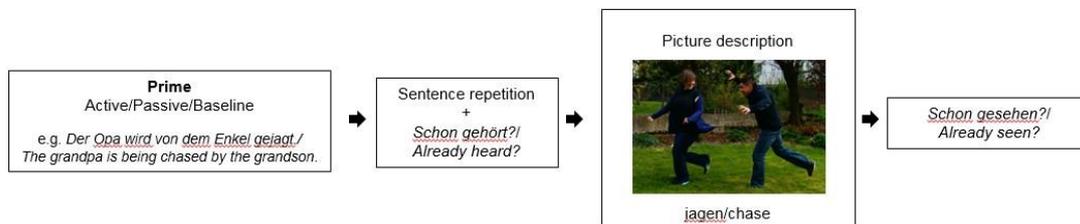


Figure 1: Exemplary experimental trial.

Two variables found to influence structural priming and/or sentence production were included to stimulate the production of non-canonical structures; verb repetition between prime and target (i.e. *lexical boost*) throughout the experiment and (in)animacy of the agent (balanced). Picture descriptions (*responses*; Fig.2) show a main effect of prime. More precise analyses also show a main effect of animacy; changing voice (including to anticausative (AC) structures; Tab.1) is a valid structural option for speakers of German to promote more accessible referents into the subject position.

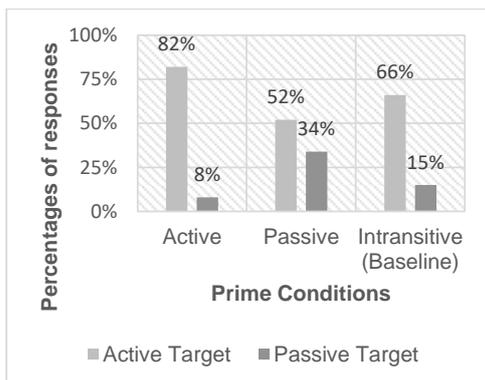


Figure 2: Percentages of active & passive targets. 'Other' responses excluded.

Table 1: Percentages (n) of target responses in the different prime and animacy conditions.

	Prime Agent	Target		
		Active	Passive	AC
Act	an	98.3 (118)	0.8 (1)	0.8 (1)
	inan	67.0 (79)	15.3 (18)	17.8 (21)
Pas	an	79.8 (95)	20.2 (24)	0
	inan	25.2 (30)	47.9 (57)	26.9 (32)
Base	an	96.7 (116)	3.3 (4)	0
	inan	35.3 (42)	26.1 (31)	38.7 (46)

Note: Act=Active, Pas=Passive, Base=Baseline, an=animate, inan=inanimate, AC=anticausative

After this first demonstration of passive priming in German, it is now a future task to successively eliminate general influences on sentence production to extract the purely *structural priming* effect.

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, 40, 1-61.

Loebell, H. & Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, 791-824.

Friday Poster.78

BETA-BAND ACTIVITY TO THE RESCUE OF NON-NATIVE PROCESSING

Laurent Dekydtspotter (Indiana University, Bloomington), Kate Miller (IUPUI), Mike Iverson (IUB), Yanyu Xiong (IUB), Kyle Swanson (UB), and Charlene Gilbert (IUB)

ldekydts@indiana.edu

We report on ERP activity as displaced expressions are determined to belong to both a matrix and an embedded clause at bridge *dit que* 'said that' as in (1a-d) by native (NSs;n=24) and advanced L1-English non-native (NNSs;n=22) speakers of French. Processing the bridge requires a process memory step in which predictions induced by *dit* on a short timescale are transferred to a longer timescale for integration with upcoming content. (1a-d) manipulates predictions in displacement: N-complements (*à propos de lui*; 1b,d) are selected by head noun *décision* and accompany it in the recursion. NP-modifiers (*le concernant*; 1a,c), not selected by the noun, do not (Chomsky, 2005; Lebeaux, 1988). When antecedents are available, pronouns can be bound in recursive movement (1b) or coreferential in discourse (1a). Without matching antecedents (1c,d), processes of anaphora are thwarted.

- (1.) a. Quelle décision le concernant est-ce que Paul a dit que Lydie avait rejetée ?
b. Quelle décision à propos de lui est-ce que Paul a dit que Lydie avait rejetée ?
c. Quelle décision le concernant est-ce que Lydie a dit que Paul avait rejetée ?
d. Quelle décision à propos de lui est-ce que Lydie a dit que Paul avait rejetée ?
'Which decision regarding/about him did Paul/Lydie say that Lydie/Paul had rejected?'

Discrete rhythms of neuronal oscillation meet not only external speech cues but also internal mental activity: gamma, prediction; theta, unification; and beta, management of cognitive focus (Lewis et al., 2015). Characteristics of ERPs in the time and time-frequency domains might reveal NS-NNS timing differences (Boxell & Felser, 2017) and fragility of representations (Dekydtspotter & Miller, 2013). Respondents completed a RSVP task (300ms/word, 250ms/ISI) with 25 quadruples crossing Structure (N-complement/NP-modifier) and Antecedent Gender (Match/Mismatch) (1a-d). EEG was recorded via a 64-electrode EGI system. Data were preprocessed with a .05-100.5Hz bandpass filter and cleaned of artefacts via epoch/channel rejections and Independent Component Analysis. 87% of NS and 86% of NNS trials were retained. First, average amplitudes, with 50ms baseline into critical words *dit* 'said' and *que* 'that' (Phillips et al., 2005), were compared over four regions (Fiebach et al., 2002). Mixed-effect models from 250-550ms after critical words' onsets revealed modifier-complement differences at *dit* in NSs ($p=.02$) but at *que* in NNSs ($p=.027$). Effects were non-focal in NNSs. Time-frequency analysis with 750ms baselines at 4-40Hz revealed main-effect modifier-complement differences during *dit* 'said' at 34-40Hz (gamma rhythm) for 104ms with greater differences in mismatch than match ($p = 0.007$) and during *que* 'that' at 6Hz for 265ms with increased modifier-complement power-differences in match than mismatch ($p = 0.004$). Post-hoc analyses of average power over these clusters revealed greater power for modifier than complement in mismatch in gamma ($p = .0005$) and theta rhythms ($p = .014$). A main group effect at 18-20Hz (beta rhythm) occurred for 150ms 250-400ms into verb *dit* 'said' in the left anterior region and at the midline, with greater modifier-complement power differences for NNSs than NSs ($p = .008$). Main-effects of group also arose for 250ms during *que* at 5-6Hz (theta rhythm) with greater modifier-complement differences for NNSs than NSs over the 300ms window ($p = 0.024$). At 17-18Hz, smaller modifier-complement differences in NNSs than NSs for 85ms during *que* 'that' reversed the preceding difference on *dit* 'said' ($p = 0.028$). Delayed ERPs in NNSs confirm timing differences. Modifier-complement differences in match vs. mismatch at *dit* and *que* respectively suggest information shared across timescales. Group differences in beta-band activity suggests its role in managing the cognitive focus in NNSs.

FALSE POSITIVES IN GROWTH CURVE ANALYSES OF VISUAL WORLD PARADIGM DATA

Yujing Huang (University of California, Davis) & Jesse Snedeker (Harvard University)
 yujinghuang@fas.harvard.edu

Visual World Paradigm (VWP) data are often analyzed by reducing the pattern of eye-gaze fixation in a time window to a single number (e.g., proportion of looking time on Target). But this removes some of the fine-grained temporal information that makes the VWP attractive for studying language processing. Growth curve analysis (GCA) is a statistical method that makes use of this information by treating time as a predictor of the changing fixation proportion across the time window (Mirman, 2008). GCA has been widely used with VWP data. Previously, we found that one implementation of the GCA model (based on Koring et al., 2012) generated a very high rate of false positives, raising concerns about the use of GCA's. The present study explores whether other implementations have similar problems.

We collected two additional data sets from outside our lab and performed Monte Carlo simulations on their original GCA analyses to calculate the false positive rate of each model. Table 1 lists the *Percent* of false positives in 1000 iterations and the result of a *Fisher.test* comparing the false positive rate to the expected rate of 5%. The models in these studies varied in their treatment of the dependent variable (raw proportion or e-logit transformation), the size of time bin used (20, 40, 200 ms) and the number of time polynomials used. All three implementations resulted in unacceptably high rates of false positives. Critically, the main findings of both Pozzan et al. (2017) and Mirman & Graziano (2012) still obtain when more robust analyses are used and thus we see no reason to reconsider the theoretical conclusions of either paper.

Term	Pozzan, Gleitman & Trueswell (2017)		Mirman & Graziano (2012)		Huang & Snedeker (unpublished)	
	Percent	Fisher.test	Percent	Fisher.test	Percent	Fisher.test
Time*condition	33.5%	<0.001	5%	1	68%	<0.001
Time^2*condition	41.2%	<0.001	7.3%	0.04	53%	<0.001
Time^3*condition	23.4%	<0.001	59.8%	<0.001	44%	<0.001
Time^4*condition	NA	NA	47.6%	<0.001	33%	<0.001

Table 1

We suspect that the inflated false positive rate is attributable to the high degree of autocorrelation in eye-tracking data: where you look at Time X is predicted by where you were looking at Time X-1, even at lags up to 1 second. When we conducted a parallel a Monte Carlo simulation on a word learning data set (Mirman et al. 2008), which did not have the same autocorrelation structure, we found no indication of anti-conservativity (Type I error was approximately 5% for each temporal polynomial). In our presentation, we will present additional findings on the autocorrelation structure of these data sets, how this changes with temporal window size and different approaches to modeling data with this structure (including Cho et al., 2018).

References

Mirman, D., & Graziano, K. M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *Journal of experimental psychology: General*, 141(4), 601.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of memory and language*, 59(4), 475-494.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, 108(1), 271-280.

Pozzan, L., Gleitman, L. R., & Trueswell, J. C. (2016). Semantic ambiguity and syntactic bootstrapping: The case of conjoined-subject intransitive sentences. *Language Learning and Development*, 12(1), 14-41.

Friday Poster.80

INVESTIGATING THE INFLUENCE OF SIMILARITY STRUCTURE OF LANGUAGE NETWORKS ON VISUAL WORD RECOGNITION: INSIGHTS FROM MEGASTUDIES

Cynthia S. Q. Siew (University of Warwick; National University of Singapore)

Contact: cynsiewsq@gmail.com

The mental lexicon, the part of long-term memory where lexical representations are stored, can be represented as a language network consisting of words that are connected if they are phonologically, orthographically or phonographically (both phonologically and orthographically) similar to each other. The goal of this project was to investigate whether local and global structural properties of words in the language network influenced visual word recognition.

Orthographic and phonological language networks were constructed for various languages (English, Dutch, French, Malay). In these networks, nodes represented lexical representations. In orthographic networks, connections were placed between words whose spellings differed by an edit distance of 1 letter (cf., Coltheart et al., 1977). In phonological networks, connections were placed between words whose phonological transcriptions differed by an edit distance of 1 phoneme (Luce & Pisoni, 1998). Network analyses of these language networks revealed that the overall similarity structure of all languages shared strikingly similar properties (i.e., a small-world structure with short average path lengths and large clustering coefficients)—properties suggesting that language networks are easy to navigate despite their size (Watts & Strogatz, 1998).

To determine if local and global structural characteristics of words affect lexical retrieval, a number of network measures (i.e., degree, local clustering coefficient, closeness centrality) were computed for individual words. Degree refers to the number of connections incident on a given node (i.e., orthographic or phonological neighborhood size). Clustering coefficient refers to the extent to which a word's neighbors are also connected to (or are neighbors of) each other. Both degree and clustering coefficient are measures of a word's local similarity structure. On the other hand, closeness centrality quantifies a word's global similarity structure and measures how "close" a word is to other words in the network by computing the mean of the shortest path between a given word and every other word in the network.

Using behavioral data from a variety of language databases (English Lexicon Project, British Lexicon Project, Dutch Lexicon Project, French Lexicon Project, Malay Lexicon Project), stepwise regression models were employed to select the best model (while minimizing model complexity) that could account for reaction times and accuracies in speeded naming and visual lexical decision. Predictors included a number of lexical variables previously shown to affect recognition (e.g., length, word frequency) as well as network similarity measures derived from the language networks as described above. Across languages and both tasks, all network similarity measures consistently emerged as variables in the best performing model and were shown to inhibit recognition and naming performance. Specifically, being orthographically or phonologically similar to other words at local and global levels of the language network inhibited lexical retrieval.

Although the variance accounted by network similarity measures is small compared to word frequency, their membership in the best performing models suggest that lexical processes involved in visual word recognition are universally sensitive to the local and global similarity structure of language. The present findings demonstrate that the network structure of the mental lexicon influences visual word recognition performance and could have important implications for models of visual word recognition.

References

- Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI*. Lawrence Erlbaum Associates.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing, 19*(1), 1–36.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

TALKER-SPECIFIC ADAPTATION OF INFERENCES BASED ON SCALAR ADJECTIVES

Cameron Morgan, Bethany Gardner, Rebecca Lawrence & Chigusa Kurumada (University of Rochester)
ckuruma2@ur.rochester.edu

Individual differences abound in language use. Theories in socially and contextually situated language processing predict that listeners must be sensitive to variability across talkers and make use of the information in comprehension (Münster & Knoeferle, 2017). In pragmatic processing, perceived reliability of the talker’s language use is considered to modulate real-time inference generation: Contrastive interpretation of a scalar adjective (e.g., “the *tall* glass” as the taller of two glasses) can be canceled when the talker violates expectations of cooperative language use (Grodner & Sedivy, 2011). We examine the mechanism supporting such talker specific inferences.

E(xperiment) 1: [Design] 48 participants were randomly assigned to the Reliable-Talker or the Unreliable-Talker condition. The Unreliable Talker: 1) was introduced as a patient of linguistic and social impairment; 2) mislabeled objects; and 3) frequently over-modified. In contrast, the Reliable Talker modified appropriately (Table 1). Participants completed 16 target and 32 filler trials. Target sentences had the scalar adjectives *large* or *small* (e.g., “Click on the *large* dog”) presented with a visual scene with either one set of contrasting items (e.g., a large and a small dog) or two sets of contrasting items. We measured anticipatory eye movements after the adjective onset as a proxy of contrastive inference based on the adjective (Sedivy et al., 1999). [Results] Anticipatory eye movements based on the scalar adjectives dissipated in the Unreliable-Talker condition ($p < .001$), replicating the talker-based adaptation of pragmatic inferences (Grodner & Sedivy, 2011).

E2: [Design] 24 participants were tested on a modified version of the Unreliable-Talker condition. In this experiment, participants did not receive explicit top-down instructions about the talker. The bottom-up information about the reliability of language use from E1 remained identical. [Results] Compared with the Reliable-Talker condition of E1, anticipatory eye movements based on the scalar adjectives dissipated in the (bottom-up only) Unreliable-Talker condition ($p < .001$), just as they did in the Unreliable-Talker condition of E1. This critically extends the previous work, demonstrating that listeners can extract the unreliability of the talker’s adjective use exclusively from the bottom-up information and modify their real-time language comprehension accordingly.

E3: [Design] 48 participants were tested in a two-talker version of E1 wherein each participant was exposed to two talkers (male and female) who were either a) both reliable or b) reliable and unreliable. This was to test the key hypothesis: If the adaptation of contrastive inferences is truly talker-specific, participants’ eye movements to the reliable talker should be the same across these two conditions. [Results] As hypothesized, the main effect of the between-subject conditions was not significant, supporting the idea that participants maintained their contrastive inferences when the unreliable input is produced by a distinct talker. The result, however, remained rather inconclusive due to limited statistical power. We are currently conducting **E4**, testing the same hypothesis as E3 with more trials, controls, and participants.

Thus, listeners are exquisitely sensitive to the talker’s reliability in pragmatic use, modulating real-time generations of contrastive inferences even in the absence of explicit, top-down instructions. We argue that the talker-based adaptation of pragmatic interpretations facilitates language comprehension by preferentially allocating processing costs (i.e., considering—and moving eyes to—a referent) in cases where the adjective information reliably signals an intended referent.

Table 1: Reliability manipulations used in E1-E4.

	Instruction	Filler Item manipulations
Reliable	Reliable	<ul style="list-style-type: none"> • Informative adjective use • Correct references
Unreliable (E1, E3, E4)	Unreliable “The talker has a linguistic and social impairment”	<ul style="list-style-type: none"> • Redundant adjective use (e.g., “the small red tomato” with only one red tomato) • Erroneous references (e.g., calling a toothbrush a “hairbrush”)
Unreliable (E2)	Reliable	Same as Unreliable (E1)

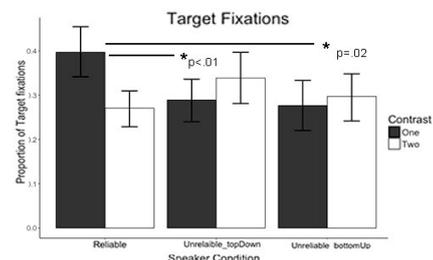


Figure 1: The average target-fixation proportion (proportion of fixations to the target object with respect to all possible object fixations) within the predetermined window of analysis (i.e., a 500-ms window starting 200 ms after the adjective onset) in **E1&2**.

(NOT) FORGETTING VERBS IN HINDI DOUBLY CENTER-EMBEDDED STRUCTURES

Samar Husain (Indian Institute of Technology Delhi) and Sakshi Bhatia (University of Massachusetts Amherst)

samar@hss.iitd.ac.in, sakshibhatia@umass.edu

Working-memory constraints are known to induce 'forgetting' effects in center embedded constructions in English (Gibson and Thomas, 1999). Such forgetting effects are argued to underlie the illusion of grammaticality observed in sentences with a missing verb phrase: **The patient who the nurse who the clinic had hired met Jack*. However, this forgetting effect has not been observed in head-final languages such as German and Dutch (Vasishth et al., 2010; Frank et al., 2016; but see, Häußler and Bader, 2015). The asymmetry in processing of German and Dutch vis-à-vis English has been attributed to the parser's adaptability to certain language characteristics such as head directionality. Recent work on the processing of the head-final language Hindi has questioned the infallibility of the prediction processes in such languages (e.g., Apurva and Husain, 2018; Bhatia and Husain, 2018). In light of these recent results, both the 'forgetting hypothesis' and the 'adaptation hypothesis' need to be tested further cross-linguistically.

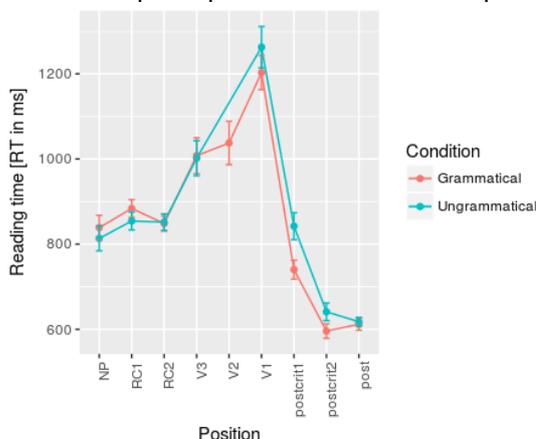
In this study, we utilize doubly center-embedded structures in Hindi of the type previously employed by Vasishth et al. (2010). In condition (a) all the verbs (i.e., the inner most verb V3 followed by V2 and V1) were present. In condition (b) V2 was missing, thus making it ungrammatical. We conducted a centered self-paced reading experiment (N=45) using 24 items (48 fillers) followed by comprehension questions on 66% of trials.

(a) NP1_i [Relpro_i NP2.object_j [Relpro_j NP3 NP4.object V3] V2] V1, postcritical.region ...

(b) NP1_i [Relpro_i NP2.object_j [Relpro_j NP3 NP4.object V3] \emptyset] V1, postcritical.region ...

The forgetting hypothesis proposed by Gibson and Thomas (1999) predicts a slower reading time at V1 (and possibly at the postcritical region due to spillover) in (a) vs (b), owing to the fact that V2 has presumably been forgotten at NP4. However, if Vasishth and colleagues are correct then we should see an opposite pattern – reading times at V1 in (b) should be slower than (a) because of not encountering the required number of verbal heads.

Log RTs were analyzed using linear mixed-effects models. The results show a significant difference ($t=2.4$) between the two conditions at the postcritical region such that reading time in condition (b) was slower (Mean=842.1, SE=31.6) than condition (a) (Mean=740.1, SE=22.3). This suggests that Hindi native speakers are able to distinguish between the two conditions, and are not susceptible to forgetting of the second verb. This reading time data is compatible with two underlying states: (i) the parser is making correct structural integrations, or (ii) the parser is using a surface cue (e.g., counting the number of Relpros since these clearly mark clause boundaries) to track the upcoming heads. The comprehension accuracy data for this experiment - only 33% of the total participants exceed 70% in their comprehension accuracy for the items - makes (ii) seem more likely. In contrast, 100% of the participants exceed 70% comprehension accuracy in the filler sentences.



The results of this study pattern with other head-final languages in that Hindi native speakers are not susceptible to verb forgetting effects in center-embedded structures. This result is consistent with the predictions of the language adaptability hypothesis (Vasishth et al., 2010). However, the low comprehension accuracy suggests a shallow parsing strategy where the required structural integrations may not be taking place in spite of the successful tracking of the number of verbal heads.

Friday Poster.83

WORKING MEMORY CONSTRAINTS OVERRIDE PREDICTION IN PROCESSING HINDI CENTER EMBEDDED CONSTRUCTIONS

Apurva (Indian Institute of Technology Delhi) & Samar Husain (Indian Institute of Technology Delhi)
apurva@hss.iitd.ac.in, samar@iitd.ac.in

The evidence for working-memory (WM) constraints such as similarity-based interference due to identical case-markers on preverbal nouns (Lewis & Nakayama, 2001; Vasishth & Drenhaus, 2011) or memory load (e.g., Levy and Keller, 2013) in head-final languages is weak. Such languages are assumed to have robust predictions and provide evidence for expectation-based accounts (Hale, 2006; Levy, 2008). In particular, increased preverbal nouns and their case-markers are known to facilitate processing (e.g., Levy and Keller, 2013). In this study we provide new evidence in favour of WM constraints during the processing of Hindi center-embedded constructions and show that the data does not support expectation-based accounts. We first replicate an SPR (self paced reading) experiment by Vasishth (2003) and then conduct a Cloze completion study to reinterpret the SPR result using expectation based account. All analyses were done using linear mixed-effects models.

The SPR experiment (N=40) had a 2x2 design: Case-marker similarity (Identical.Case vs Non-identical.Case) and level of clause embedding (Single.Embedding vs Double.Embedding) was crossed. Below (a)=Single.Embedding, Non-identical.Case and (b)=Double.Embedding, Non-identical.Case conditions. (c) and (d) show the Identical.Case conditions that have Accusative (ACC) case-marker on the final nouns (N3-inanimate for Single and N4-inanimate for Double.Embedding conditions). 24 sets of items were used.

- (a) N1-animate=ERG N2-animate=ACC [N3-inanimate=Nominative **Non-finite-Verb1**] Verb
- (b) [N3-animate=ACC [N4-inanimate=Nominative **Non-finite-Verb1**] Non-finite-Verb2] Verb
- (c) N1-animate=ERG N2-animate=ACC [N3-inanimate=ACC **Non-finite-Verb1**] Verb
- (d) [N3-animate=ACC [N4-inanimate=ACC **Non-finite-Verb1**] Non-finite-Verb2] Verb

Log RTs were used for analysis. Results show a main effect of Case-marker ($t=-5.6$): RT in Non-identical-Case < Identical-Case at the innermost non-finite verb (NFV1). Nested contrasts show this pattern holds for Single.Embedding ($t=4.07$) as well as Double.Embedding conditions ($t=3.6$). These results successfully replicate the finding in Vasishth (2003). The slowdown at the innermost verb was interpreted by Vasishth as retrieval difficulty at NFV1 due to similarity-based interference (as a result of similar ACC case-markers on the nouns). However, this slowdown could also be explained by the expectation-based account if the critical NFV in the Identical.Case conditions is less expected. To ascertain this we conducted a Cloze completion study.

The Cloze study (N=25) used the incomplete versions of (a-d). Items were presented until the last noun and were run using Linger. Participants' responses were coded for the target verb type (class and category) and overall grammaticality of the completion (grammatical vs ungrammatical). The completion data showed that for the Single.Embedding conditions, Prob(NFV1|N3,N2,N1) were .35 (for Non-identical.Case) and .67 (for Identical.Case). For the Double.Embedding conditions, Prob(NFV1|N4,N3,N2,N1) were .30 (for Non-identical.Case) and .65 (for Identical.Case). The difference between Identical and Non-identical.Case was significant ($z=-4.7$). The completion data shows that for the Single.Embedding/Double.Embedding conditions the innermost NFV is highly expected in the Identical.Case vs Non-Identical.Case conditions. The expectation accounts would therefore predict faster RT at NFV1 in the Identical.Case vs Non-Identical.Case conditions. This prediction does not hold for the SPR data discussed earlier.

Grammaticality analysis of the completion data shows a significant effect of embedding ($z=10.2$) as well as case-marker similarity ($z=3.5$). Grammatical completions reduce in Double.Embedding vs Single.Embedding conditions and in Identical.Case vs. Non-Identical.Case conditions.

Our work reinforce the role of WM constraints in sentence comprehension. Similarity-based interference (due to identical case-markers) leads to processing slowdown at the verb in spite of high expectation. And memory load (measured as no. of preverbal nouns and similarity of case-markers) leads to reduced grammatical predictions in a Cloze task. These results go against the prediction-based theories and provide new evidence for the influence of WM constraints in a head-final language (cf. Vasishth & Drenhaus, 2011; Levy & Keller, 2013).

References: Hale (2006) NAACL; Levy (2008) Cognition; Levy and Keller (2013) JML; Lewis & Nakayama (2001) CSLI; Vasishth (2003) Routledge; Vasishth & Drenhaus (2011) DnD

**ORTHOGRAPHY IN SECOND LANGUAGE WORD LEARNING AND PRONUNCIATION:
FRIEND OR FOE?**

Pauline Welby (Aix-Marseille Université), Audrey Bürki (Universität Potsdam) & Elsa Spinelli (Université Grenoble-Alpes)

buerki@uni-potsdam.de

Second language (L2) learning is a key issue in contemporary societies. Its relevance is reflected in the rapid growth of language learning apps (e.g., *Babbel*, *Duolingo*). Identifying the determinants of successful learning is fundamental not only for the development of such technologies but for all applied and theoretical research in L2 learning. This study contributes to this endeavor by investigating the influence of orthography on novel word learning.

Learning words requires both acquiring new labels and associating them with concepts, and knowing how to pronounce these labels. The role of orthographic information on L2 expressive vocabulary learning has not been documented. Some studies reported that different grapheme-to-phoneme correspondences (GTPCs) between L1 and L2 can lead to non-target like pronunciations (Bassetti & Atkinson, 2015), other found that orthographic information leads to *more* target-like productions (Steele et al. 2005) and others reported mixed results (Erdener & Burnham, 2015). Given the omnipresence of written material in language learning and teaching, it is crucial to determine the consequences of orthographic exposure. This study investigates the influence of orthography in novel word learning and pronunciation.

Twenty-six native speakers of French participated in the experiment. They learned to associate pictures of 20 novel objects with their corresponding labels, i.e., new English words, produced by a native speaker of Canadian English. The novel words were all monosyllabic and contained the vowel <i> or <o>, which have different GTPCs in English and in French (North American English: <i> ~ /ɪ/ (*disk* [dɪsk]), <o> ~ /ɑ/ (*bog* [bɑg]); French: <i> ~ /i/ (*disque* [disk] 'disk'), <o> ~ /ɔ/ in closed syllables *bogue* [bɔg] 'husk'). Ten novel words were presented in spoken and written form (Audio+Ortho condition), ten were presented in the spoken form only (Audio condition). The following day, the participants named the pictures aloud.

If seeing the orthographic form helps building phonological representations and their associations with concepts, more correct responses and shorter naming latencies are expected in the Audio+Ortho condition than in the Audio condition. If seeing the orthographic form influences the pronunciation, vowels are expected to be more /i/-like or /ɔ/-like (compatible with French GTPCs) in the Audio+Ortho condition than in the Audio condition (with lower first formant (F1) and higher second formant (F2) for <i> tokens, lower F1 and F2 for <o> tokens). The analyses confirmed these predictions. Correct response were more frequent in the Audio+Ortho condition (51% vs. 39%, $p = 0.04$) and naming latencies for correct responses were 172 ms shorter in the Audio+Ortho condition ($p = 0.023$, see Fig. 1). F1 was lower in the Audio-Ortho condition for both vowels ($p = 0.022$) and F2 was higher in the Audio+Ortho condition for <i> vowels and lower for <o> vowels ($p = 0.042$, see Fig. 1).

These data demonstrate for the first time a beneficial role of orthography on L2 word retrieval in a production task. They further show that orthography can be beneficial and detrimental in the same language learning task, a finding that any recommendations on including or limiting exposure to orthographic material in second language teaching (Rafat 2016) should take into account. These data further show that orthography contributes to shaping L2 speakers' linguistic knowledge.

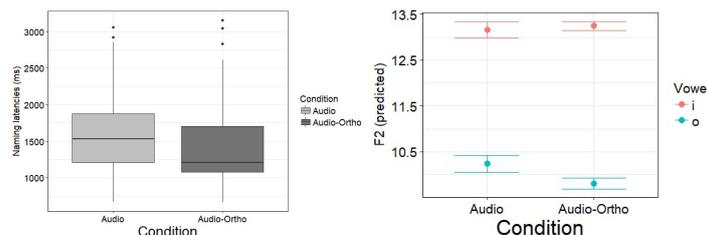


Figure 1. Response latencies (left-panel) and F2 values as predicted by the statistical model (right panel)

PROCESSING INFERRED RESULT STATES IN DISCOURSE
 Sarah Hye-yeon Lee & Elsi Kaiser (University of Southern California)
sarahhl@usc.edu

Language comprehension involves rapidly combining (a) bottom-up information from specific words/phrases with (b) top-down information from discourse context and real-world knowledge. We investigate how comprehenders combine information from lexical semantics and discourse context to compute and utilize inference-based interpretations. As a novel test domain, we especially focus on verb classes that do/do not linguistically encode results.

Background Lexical semanticists (e.g. Rappaport Hovav & Levin 1998) have identified two verb classes based on what the verb lexicalizes: **manner verbs** (e.g. *hit, strike*) lexicalize the manner of the action, and **result verbs** (e.g. *break, shatter*) lexicalize the result state of an action. In this study, we take advantage of the fact that although manner verbs do not lexicalize result states, they can be used to *infer* result states (e.g. Rappaport Hovav & Levin 1998, Talmy 1991): (1) Mary struck the vase. → The vase is broken.

Prediction If contextually-triggered (top-down) result inferences guide expectations about the presence of a result state, they will facilitate processing of upcoming result-denoting phrases as effectively as bottom-up sources of meaning (lexicalized result in result verbs).

Exp 1 (N=40) used self-paced reading. We manipulated (i) context type (**result supporting** vs. **neutral**) and (ii) verb type (**mannerV** vs. **resultV**) for a 2x2 design ((2), 34 targets, 48 fillers). The result-supporting context specifically focuses on the “fate”/result state of the object, but the neutral context does not. Nonce nouns were used to avoid noun semantics influencing the plausibility of result attainment. All targets contain a result phrase (e.g. “damaged”). How do context type and verb type influence how quickly the result phrase is read?

(2) Trevor called and asked Mary **what happened to**_{res.supp.Context}/**about**_{neutralContext} the merick.

She replied that she **hit**_{mannerV}/**broke**_{resultV} it in the morning on Monday.

She said that it is **damaged**_{result-phrase} and that she feels very sorry about this.

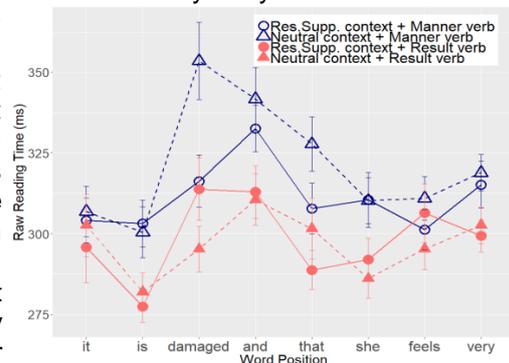
Results For RTs at the **result phrase** (e.g. “damaged”), there is a main effect of verb type (lmer, $t=2.71$), no main effect of context type ($t=0.8$), and crucially, a significant interaction ($t=2.05$): in the result-supporting context, the mannerV condition was read as quickly as the resultV condition ($t=0.19$). In the neutral context, however, the mannerV condition was slower than the resultV condition ($t=3.14$).

Exp 2 (N=40) tests whether neutral context RTs at the result phrase in Exp.1 were driven by a categorical linguistic verb class distinction or gradient knowledge associated with how likely it is that each verb-result pair (e.g. *hit-damaged*) will obtain in the real world. After Exp.1, the same participants rated questions such as (3) on a 7-point Likert scale for all the verb-result pairs they saw in Exp.1.

(3) Mary hit the merick. — How likely is it that the merick is damaged?

Results Does likelihood-of-result (Exp.2: how likely a verb is to yield a result state) predict RT of the result phrase (Exp.1) in the neutral context? Our results suggest the answer is no. Surprisingly, there is no meaningful relationship between the likelihood-of-result ratings and SPR RTs at the result phrase given neutral context ($t=1.24$).

Discussion Our results demonstrate that (1) in a neutral context, lexicalized verb meaning (result/manner) (i.e. linguistic verb class) plays an important role in guiding people’s expectations about the presence/absence of a result state. We found no evidence, however, that real-world knowledge plays a central role in this. (2) Crucially, when the context drives comprehenders to compute an inferred-result interpretation, RTs suggest that inference-based top-down information is used as effectively as bottom-up lexicalized information. This is shown by the fact that in the result-supporting context, the result phrases in the mannerV condition were read as fast as in the resultV condition.



THE ROLE OF PREDICTION ERROR IN LINGUISTIC GENERALIZATION AND ITEM BASED LEARNING

Masa Vujovic, Michael Ramscar and Elizabeth Wonnacott
(masa.vujovic.15@ucl.ac.uk)

Discriminative learning models frame language learning as a process by which *prediction error* is used to discriminate uninformative cues and to reinforce informative cues. This theoretical approach makes predictions about learnability since some learning contexts facilitate prediction error to a greater extent than others. Ramscar et al. (2010) demonstrated that learning of appropriately general word meanings is facilitated when learners (both adults and 2-year-olds) view referents *before* hearing their labels, compared to when they hear labels and *then* view referents. This is predicted under *discriminative learning*, since only in the former case is there opportunity for prediction error, which enables discrimination of the appropriate set of semantic features for a given label. Less work has addressed the predictions of this approach for linguistic learning at morphological/syntactic level; however, Ramscar (2013) proposes discriminative learning as an explanation of the cross-linguistic preference for suffixing (build-er, finish-ed) over prefixing (un-happy, dis-advantage) in marking linguistic generalizations (Greenberg, 1963). Specifically, as a consequence of linear order in discriminative-learning, suffixing benefits learning of abstract common dimensions from the preceding lexical items (which enables generalization to novel items), whilst prefixing benefits item-based learning (Arnon & Ramscar, 2012; Ramscar 2013).

Method: The current work explores this in a computational simulation and an experiment with humans in which we compared generalization and vocabulary learning in artificial languages. The languages described alien characters, with two noun categories marked by phonological and semantic cues and each accompanied by one of the two affixes (category1_affix: *ge*, category2_affix: *ma*). Nouns could precede (“suffixing” language) or follow (“prefixing”) the affix. Computational modelling using a discriminative implementation of the delta rule (Widrow & Hoff 1960) suggested that that participants exposed to a “prefixing” language would show better vocabulary learning, whereas those exposed to a “suffixing” language would be better at generalizing the correct affix to new category members. Four groups of participants (N=42 per condition) were trained on four versions of the artificial languages; we manipulated whether the language was prefixing or suffixing, vocabulary size and type (category) frequency.

Results: We found evidence of generalization in both prefixing and suffixing conditions, but, in line with the prediction, only participants in the suffix condition showed appropriate generalization over low type frequency items. While we did not see the predicted benefit of prefixing for vocabulary learning, there was overall better vocabulary learning for low type frequency items in both conditions. Interestingly, these were the items for which we did not see strong generalization in the prefix condition. The results demonstrate the crucial role of *prediction error* in linguistic generalization, and the importance of linear order in human language learning. They also speak against any simplistic account in which item-based learning necessarily proceeds generalization.

References

- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292-305.
- Greenberg, J. H. (1963). *Universals of Language*. London: MIT Press.
- Ramscar, M., et al. (2010). The effects of feature-label order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909-957.
- Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, 46(4), 377-396.
- Widrow, B., & Hoff, M. E., (1960). Adaptive switching circuits, in *IRE WESCON Convention Record*, Part 4, New York: IRE, 96–104.

Poster Presentations - Abstracts Saturday

Poster Abstracts Saturday, Sept 08, 2018

Saturday Poster.1

THE ROLE OF VOWEL DURATION FOR PERCEIVED VOWEL QUALITY OF CZECH VOWELS: DATA FROM NATIVE AND NON-NATIVE LISTENERS

Nikola Paillereau (Institute of Phonetics, Charles University), Radek Skarnitzl (Institute of Phonetics, Charles University)

nikola.paillereau@mac.com, radek.skarnitzl@ff.cuni.cz

The Czech vowel system has five short and five long monophthongs, /ɪ i:, ε ε:, a a:, o o:, u u:/, and three diphthongs, /ou̯ au̯ ɛu̯/. Traditionally, the short–long pairs of vowels are described as having the same quality (very similar F1 and F2 values), and differing in quantity (long vowels being approximately twice as long as their corresponding short counterparts. (Hála, 1941). Most recently, Skarnitzl and Volín (2012) showed important F1 and F2 difference and different duration ratios in pairs of high vowels, especially in [ɪ/i:] but also in [u/u:]. Because the mean differences for F1 and F2 in both pairs exceeds 5%, the *just noticeable difference* in vowel formants (Kewley-Port and Zheng 1999), we aim to investigate whether the spectral difference is perceived by a) native Czech listeners (experiment 1), and if yes, to what extent the spectral difference can trade for durational difference in discriminating vowels of each pair, and also by b) native French listeners (experiment 2), whose native vowel system is based on four degrees of aperture without contrastive length.

Experiment 1 consists of three series of identification tests with goodness ratings comprising Czech isolated vowels (5 long and 5 short × 4 repetitions). Listeners are 27 Bohemian Czech monolingual listeners (M age = 11.2). In the 1st identification test, /i: ε: a: o: u:/ are shortened to one half of the original duration. In the 2nd test, /i: ε: a: o: u:/ are shortened to 100 ms each (corresponding to about one third of the initial duration). In the 3rd test, short /ɪ ε a o u/ are lengthened to twice the original duration. The original vowels were also identified and the result serves as reference.

Results are expressed in terms of a *Fit Index* (FI) value (Guion et al. 2000), which is obtained by multiplying the identification score by the goodness rating, with a maximum of 5. Whereas results of test 1 indicate that *the half-long* [i:]-like stimuli were categorized as *long* /i:/ (in 79% with FI at 3.7) and other stimuli were identified as phonologically *short*, those of test 2 show that all *the 100-ms* stimuli were most of the time categorized as *short* vowels (including [i:]-like stimuli which were identified as short /ɪ/ in 80%, with FI at 3.9). Results of test 3 show that all *the lengthened* stimuli were mostly categorized as phonologically *long* vowels. Nevertheless, the [i:]-like stimuli were identified as phonologically *long* /i:/ in only 55% with FI at 2.5 and as short /ɪ/ in 35% with FI at 1.5.

Experiment 2 is an identification test, administered to ten non-southern French listeners (M age = 39) and comprises two sets of stimuli: ten French isolated vowels [i, e, ε, a, u, o, ɔ, y, ø, œ] × 3 repetitions, and ten Czech isolated vowels × 3 repetitions. All vowels were intermingled and listeners were not told that vowels were from different languages.

Czech vowels are classified as “fair” and “poor” examples of French vowels according to Guion’s et al. (2000) classification based on FI value for Czech and French (reference) vowels. Results show that most Czech vowels are “fair” exemplars of French vowels. The best exemplar is the Czech short [ɪ], assimilated to French /e/ with an FI at 1.1 s.d. from the reference. Fairly well assimilated is also Czech short [u] to French /o/ (FI at 1.38 s.d. from the reference). The long vowels [i:], [u:] can be considered as “poor” exemplars of the French vowels /i/ and /u/, with an FI at respectively 3.33 and 4.72 s.d. from the reference.

Conclusion: Czech listeners perceive the spectral difference that exists between [ɪ i:] but duration still plays a role in perceiving the /ɪ/i: contrast. Indeed, when shortened to about one third of its original duration, the shortened [i:] is categorized as short /ɪ/. It therefore seems that all Czech vowels must maintain certain duration to be interpreted as phonologically long by Czech listeners. Even though Czech listeners are not much sensitive to a qualitative contrast between [u u:], French listeners assimilate these Czech vowels to two different native categories /o u/. They also assimilate Czech [ɪ i:] to French /e i/. Our results thus suggest that spectral differences in short–long pairs of high vowels are perceptually salient, and that the salience depends on listeners’ linguistic background.

Saturday Poster.2

TESTING IMPLICIT LEARNING WITH CASE MARKING VARIATION

Heeju Hwang (University of Hong Kong), Jeong-Ah Shin (Dongguk University), YooLae Kim (Dongguk University) & Bailiang Li (University of Hong Kong)
heejuhwang@gmail.com

INTRODUCTION Structural priming is known to be insensitive to the identity of closed class elements such as prepositions. For example, Bock (1989) found that primes containing a different preposition than targets (e.g., A cheerleader saved a seat for her friend) were equally successful in priming target constructions (e.g., The girl handed the paintbrush to the man) as were primes containing the same preposition (e.g., A cheerleader saved a seat to her friend) (see also Ferreira, 2003; Pickering & Branigan, 1998; Scheepers, 2003). Implicit learning models (Change et al., 2006; Reitter et al., 2011) suggest that closed class elements do not have long-lasting and cumulative effects on structural priming. This claim, however, has not been tested with case-markers, which are a key closed class element in many languages such as Korean. Here we test whether morphological variation in case marking leads to long-term structural priming in Korean. Similar to English ditransitive constructions, Korean allows two alternate ditransitive constructions such as 'Mary-NOM John-DAT book-ACC gave' (DAT-ACC) and 'Mary-NOM John-ACC book-ACC gave' (ACC-ACC). The two structures are commonly suggested to have the same underlying structure but result from morphological variation in case marking (e.g., Oh, 2006; Sohn, 2001; Yook, 2013, but see Jung & Miyagawa, 2004). We investigate whether morphological variation in ditransitive constructions can lead to long-term structural priming in comparison to transitive constructions (actives/passives) that have different underlying structures.

EXPERIMENT We recruited 22 native Korean speakers. We examined long-term structural priming using a sentence completion methodology (e.g., Pickering & Branigan, 1998; Kaschak et al., 2006). The experiment consisted of three phases: pretest phase to measure baseline production of ditransitive and transitive constructions (12 items each), priming phase where participants were exposed to only one dative and transitive construction (24 items each, e.g., Mary-NOM __-ACC __-ACC gave/ Mary-NOM __-by was pushed), and posttest phase to measure the long-term structural priming effect (12 items each). Participants were asked to describe a picture depicting a ditransitive or transitive event using given sentence stems during the priming phase but without stems during the pretest and posttest.

RESULTS The analysis of transitive constructions revealed a significant interaction effect between priming and test sessions ($p < .05$). That is, Korean speakers exposed to passive constructions during the priming phase were significantly more likely to produce passive constructions during the posttest (16% more passives) than those who were exposed to active constructions (2% more passives). This suggests that experience with a particular transitive construction can lead to the long-term adaptation within the sentence production system in Korean. The exposure to a dative construction, however, did not result in any long-term adaptation. Whether participants produced a DAT-ACC or ACC-ACC construction during the priming phase, they produced a DAT-ACC structure 100% of the time during the posttest phase. This suggests that ditransitive constructions that involve morphological variation with the same underlying structure do not cause long-term structural priming as suggested by implicit learning accounts of structural priming. Note, however, that Shin and Christianson (2009) showed that ditransitive constructions resulted in short-term structural priming between Korean and English. This may suggest that morphological variation may cause short-term structural priming without causing long-term adaptation. More broadly, this indicates that different learning mechanisms may be involved with short-term and long-term effects of syntactic experience and that syntactic experience leading to short-term priming does not necessarily lead to long-term adaptation.

Saturday Poster.3

ITEM-BOUND VS CATEGORY-BASED GENERALIZATIONS. AN ENTROPY MODEL

Silvia Rădulescu (Utrecht University), Efi Giannoupoulou (Utrecht University), Sergey Avrutin (Utrecht University), & Frank Wijnen (Utrecht University)
S.Radulescu@uu.nl

What triggers the inductive leap from memorizing items and statistical regularities to inferring abstract rules? We propose an innovative information-theoretic model for both learning statistical regularities and generalizing to new input. Our entropy model predicts that *rule induction is an encoding mechanism triggered by the interaction between input complexity (entropy) and the limited encoding power of the human brain (channel capacity)*.

While traditional cognitive psychology claimed that rule learning relies on encoding of linguistic items as abstract categories (Marcus et al, 1999), as opposed to learning statistical regularities between specific items (Safran et al., 1996), recent views converge on the hypothesis that it is one mechanism – *statistical learning* – that underlies both item-bound learning and abstract rule learning (Aslin & Newport, 2012; 2014; Frost & Monaghan, 2016). However, it is still not clear how a single mechanism outputs two qualitatively different forms of encoding – item-bound and category-based generalization, and what factors trigger the inductive leap from one to the other.

In our model, less *input complexity (entropy)* facilitates finding regularities between specific items, i.e. item-bound generalization, while a higher complexity exceeding *channel capacity* drives category-based generalization. Rule learning is a phased mechanism that starts out by memorizing specific items and finding regularities between them (*item-bound generalizations*) and it gradually moves to an abstract *category-based* encoding, as a function of increasing input entropy.

In two artificial grammar experiments, we exposed adults to a 3-syllable XXY artificial grammar to probe the *effect of input complexity* on rule induction. We designed six experimental conditions with different degrees of input complexity and we used entropy to measure the complexity.

Participants gave grammaticality judgements on four types of test items: correct trained XXY strings, correct new XXY, ungrammatical X1X2Y (three different trained syllables), and ungrammatical new X1X2Y strings. Results showed that when input complexity increases, the tendency to infer abstract rules increases gradually (Fig.1). Also, in the lower entropy conditions participants correctly accepted trained XXY strings, and correctly rejected strings of three different trained syllables (X1X2Y_old), but they did not accept new XXY strings as confidently as participants in higher entropy conditions.

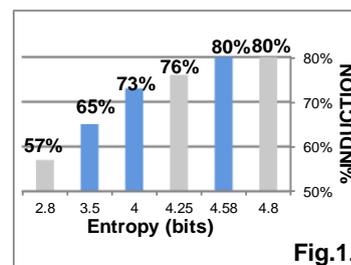


Fig.1.

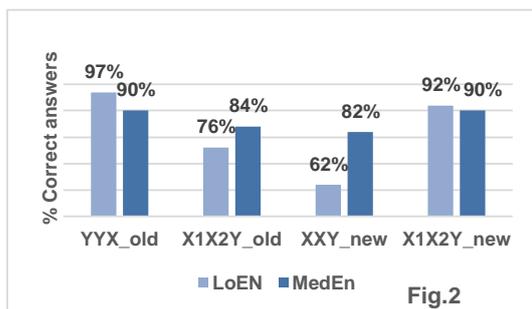


Fig.2

Given that low entropy allows for easy memorization of the specific items and combinations of items, correct acceptance of trained XXY might be supported by memory of the exact items and strings, not necessarily by item-bound generalization. In order to further test the hypothesis that low entropy input facilitates item-bound generalization, we ran another experiment. One group of adults was exposed to the same lowest entropy condition (2.8 bits), and another group to a medium entropy condition (4.25 bits). But in the test, instead of the trained XXY strings, we tested YYX strings with trained syllables (YYX_old). As expected, results showed that participants accepted YYX with trained syllables in the low and medium entropy conditions, based on the rule of *same-same-different*, but in the low entropy condition they accepted new XXY less than in medium entropy. These results support our model that low input entropy facilitates item-bound generalization.

Saturday Poster.4

ATTENTION TOWARD SHAPE AND COLOR IS AFFECTED BY LINGUISTIC STRUCTURES

Maximillian Paulus (University College London), Veronica Mazza, Giulia Calignano & Francesco Vespignani (Università degli Studi di Trento)
francesco.vespignani@unitn.it

Within the study of the relationship between language and visual information, the sentence-picture verification task has been applied to investigate the interpretation of negative sentences using behavioral and electrophysiological measures. Specifically, Lüdtkke et al. (2008) showed that the amplitude of the N400 ERP component following negative sentences does not reflect truth values but responds to a simulation of the negated state of affairs, in line with a two-stage simulation hypothesis (TSSH). In this study, we adapted the Lüdtkke et al. (2008) paradigm in order to measure different ERP components, N2pc and CDA, which reflect visuospatial attention deployment. We implemented different affirmative and negative descriptions of a target as in (1) by also manipulating the syntactic position of shape and color, assuming that the adjective inside the NP is more salient in defining the target than the predicative adjective and thus more efficient in driving visuospatial attention. A further task manipulation consisted in using three shapes (triangular, rectangular and circular) and only two antinomial colors (black and white) in the hypothesis that the antinomial pair could show an early effect of negation in attention deployment toward the actual state of affairs (not white could more easily drive attention toward black objects).

1.a The white figures are (not) triangular / 1.b The triangular figures are (not) white

After word-by-word visual presentation of Italian translations of sentence as in (1) and a 1500ms ISI, a display was briefly presented for 200ms with three objects of the same shape and color on each side of the screen. In half of the trials, one group of objects matched the sentence meaning, while in the other no group corresponded to the description. Subjects (N=16) were instructed to judge if the sentence was true with respect to the display. In each trial the two groups had different colors (white and black) and different shapes, one of which was mentioned in the previous sentence. The EEG was continuously recorded. N2pc and CDA were determined as the difference between PO7 and PO8 sites (contralateral minus ipsilateral to the side of the screen in which the mentioned shape appeared).

Behavioral RT results show an interaction of truth-value by polarity with true sentences responded faster to than false ones only for affirmative sentences. The opposite was found for negative sentences, in line with the TSSH. N2pc and CDA results show an overall preference in directing visual attention toward the side of the display containing the named shapes. The timing and amplitude of the N2pc however suggests that when shape and color are on different sides of the display (false affirmative and true negative), attention may be initially directed toward the named color after sentences in which the color is the modifier of the subject (1.a). With respect to the TSSH, a crucial difference in the timing of the N2pc emerges for false negative sentences in which named shape and color are on the same side: when the color is the predicate as in (1.b), the peak of the N2pc toward the negated state of affairs target is delayed, suggesting that not white implies a partial allocation of attention toward the side containing black objects. The results show that even though negated sentences are interpreted towards their negated state of affairs, this does not prevent the negation from exerting an online effect on the way the visual display is analyzed.

Lüdtkke et al. (2008). *J COGNITIVE NEUROSCI*, 20(8), 1355-1370.

Saturday Poster.5

Independent vs. shared syllable-representations in late Spanish-German bilinguals

Kiara F. Abad Bruzzo, Annett B. Jorschick, & Joana Cholin

Faculty of Linguistics and Literary Studies, Bielefeld University, German

Speakers' ability to speak fluently might rely on stored high-frequency (HF) motor representations for syllables. Stored syllables facilitate planning relative to a slower segment-by-segment assembly used to construct novel or low-frequency (LF) motor representations for syllables from scratch. Evidence for stored syllables comes from studies showing that HF-syllables are produced faster than LF-syllables (Cholin, Dell & Levelt, 2011; Laganaro & Alario, 2006). Most previous work focused on monolingual speakers. A study by Alario, Goslin, Michel, & Laganaro (2010) investigated syllable-frequency effects in early and late high-proficiency Spanish-French bilinguals. Their findings indicate that early bilinguals access independent language-specific syllabic representations when speaking their respective languages, while late bilinguals seem to rely on language shared representations. To further investigate the (dis-)entanglement of syllabic representations in bilingual speech, we tested syllable-frequency effects in late Spanish-German bilinguals of varying proficiency in German.

Using rigorously constructed materials controlling for segmental and metrical factors, participants produced 28 German high- and low-frequency CVC-syllables in a symbol-associating production task (Cholin et al., 2011). Language dominance was evaluated through self-assessment which proved all participants to be Spanish dominant.

The results revealed a significant interaction between Syllable Frequency and Language Dominance: Less proficient speakers showed an inverse syllable-frequency effect (HF-syllables yielding slower RTs than LF-syllables) while speakers with a higher proficiency level showed no RT-difference (with a trend towards a facilitatory syllable-frequency effect).

Post-hoc analyses took Spanish-to-German syllable correspondences into account by separating them into the categories of non-existent, allophonic, or existing in Spanish. We found a three-way interaction between proficiency, syllable frequency and correspondence category. This suggests that Spanish syllable frequencies may have impacted production times in German. For most syllables that exist in both languages (10 out of 28), German HF-syllables tend to be LF in Spanish and vice versa. However, it is not clear if the aforementioned three-way interaction reflects the syllable frequency effect in Spanish.

Against the background of the Alario et al. (2010) data, our results suggest that late bilinguals with a lower proficiency level in German may rely on Spanish syllable representations to construct German syllables anew before they gradually acquire language-specific representations. Moreover, the first representations acquired might be those that do not exist in their native language.

References

- Alario, F.-X., Goslin, J., Michel, V., & Laganaro, M. (2010). The functional origin of foreign accent: Evidence from the syllable frequency effect in bilingual speakers. *Psychological Science, 21*, 15-20.
- Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 109-122.
- Laganaro, M. & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language, 55*, 178-196.

Saturday Poster.6

CALQUES FROM ENGLISH ARE PROCESSED LIKE WELL-FORMED COLLOCATIONS BY NATIVE SPEAKERS OF POLISH: EVIDENCE FROM N400

Marta Marecka (Jagiellonian University), Agnieszka Otwinowska (University of Warsaw),
Joanna Durlik (Jagiellonian University), Jakub Szewczyk (Jagiellonian University), Marcin
Opacki (University of Warsaw), Zofia Wodniecka (Jagiellonian University)
marta.marecka@gmail.com

Multiword expressions (MWE) are combinations of words characterized by a degree of connectedness and recognised as conventional by native speakers (Siyanova-Chanturia, 2013). In native speakers, they are processed faster than expressions with a lower degree of connectedness (Vespigiani et al. 2010), and evoke lower N400 than novel word combinations in ERP studies (Molinaro & Carrerias 2010). One type of MWE are collocations, i.e. “fixed, identifiable, non-idiomatic phrases and constructions” (Benson, Benson, & Ilson, 1997). As a result of contact between a low-prestige language (L1) and a high-prestige language (L2) (Trudgill, 1972), collocations such as *makes sense* can be borrowed from L2 to L1. In such borrowings, called calques, L2 words are replaced by semantically equivalent L1 ones (Haugen, 1950). In the recent years, calques from English appeared in Polish due to a wide-spread use of English in the media and business. We thus ask whether English calques are processed similarly to Polish collocations by speakers of Polish as L1. We first created three lists of expressions in Polish, each containing 180 items: 1) well-formed Polish verb + noun collocations (e.g. *ma sens* – *has sense), 2) collocations calqued from English, where the verb was replaced by a Polish translation equivalent of the English verb (e.g. **robi sens* - makes sense), 3) absurd verb + noun expression, where the verb did not collocate with the noun (e.g. **zjada sens* – *eats sense). All expressions were tested for their mutual information scores against the National Corpus of Polish Language (NKJP). Further, we created 180 carrier sentences, in which the expressions (well-formed, calqued and absurd) were embedded. The same sentence was used across all three conditions (total of 540 sentences – 180x3). Next, we asked 31 native speakers of Polish to judge the acceptability of the expressions on a 5-point Likert scale. The sentences were divided into three lists, such that each participant saw 60 well-formed collocations, 60 calques and 60 absurd expressions, each in a different carrier sentence. The results of the by-subject and by-item ANOVAs revealed significant differences between the conditions, with the calques being assessed as significantly less acceptable than the well-formed collocations, and the absurd expressions significantly less acceptable than the other types of expressions. We then chose 120 carrier sentences, for which the differences between the expressions were most pronounced and embedded them in an ERP experiment. We chose another group of 30 participants, all native speakers of Polish, their self-rated knowledge of English ranging from intermediate to advanced; mean LexTALE (Lemhöfer & Broersma, 2012) scores: 70.86, *SD* = 11.48, range: 46.25 – 93.75. Participants read sentences on the screen and answered YES/NO comprehension questions for randomly chosen 20% of the sentences. The sentences were divided into three lists so that each participant saw 40 sentences with well-formed collocations, 40 calques and 40 absurd sentences. We measured the N400 amplitude within 340-500 ms time window on centro-parietal electrodes in response to nouns, since in previous research N400 has been detected in response to novel, ill-formed and metaphorical expressions (Molinaro & Carrerias, 2010; Vespigiani et al., 2010). The results revealed a significant difference in the N400 amplitude between the well-formed collocations and absurd expressions, but not between the well-formed and calqued collocations. This result did not depend on the participants’ knowledge of English. The results indicate no difference in online processing of correct collocations and calques, even when in the previous off-line experiment those calques were considered less acceptable than the well-formed collocations. This suggests a dissociation between performance measuring explicit judgments and early stages of processing measured by the ERPs. In conclusion, the ERP data seem to have captured the process of change to the Polish language; the influence of English is so pervasive that collocations calqued from this language are likely to become acceptable and widely used by Poles.

Saturday Poster.7

AN EAR FOR LANGUAGE: BASIC AUDITORY SKILLS ARE LINKED TO MORE EFFICIENT NOVEL WORD LEARNING

Marta Marecka (Jagiellonian University), Tim Fosker (Queen's University Belfast), Jakub Szewczyk (Jagiellonian University), Patrycja Kałamała (Jagiellonian University), Zofia Wodniecka (Jagiellonian University)
marta.t.marecka@gmail.com

Amplitude rise times (ART) is the time in which a sound goes from silence to its peak amplitude. Previous studies show a relationship between sensitivity to this basic auditory cue and reading ability (Goswami, 2002). Research suggested that detecting ART helps segment speech and develop phonological representations, necessary for reading (Goswami et al., 2011). A recent paper suggested that the same mechanism might also be used to segment speech in a foreign language, facilitating foreign vocabulary learning (Marecka et al., 2018). Learning a novel word involves learning the word form - i.e. creating its phonological representation in long-term memory - and linking this representation to a concept. We hypothesised that higher sensitivity to ART leads to more efficient segmentation of foreign speech, resulting in better encoding of foreign word forms. This might in turn increase the efficiency of the whole word learning processes, including the efficiency of linking the new word form representation to the concept. To test these hypotheses, we asked 40 adult participants to perform a 3I-2AFC task testing sensitivity to ART and a paired associates word learning test. All participants were native speakers of Polish. None of them was an early bilingual, although all of them spoke at least one foreign language. In the auditory task, participants were presented with triads of tones, one of which had a different ART. Participants had to identify the odd-one-out. The task difficulty was adapted on the basis of responses and the discrimination threshold for each participant was established. In the word learning test, participants were first auditorily exposed to 12 bisyllabic nonwords paired with line drawings depicting common objects. Each nonword-picture pair was presented 36 times (432 trials altogether). All nonwords were pronounceable in the native language of the participants and were matched for the number of consonants and vowels, but varied in terms of phonotactic probability. Participants were then tested on the knowledge of the associations with a Match-Mismatch Word Identification Task. In the task, they heard each of the 12 nonwords, followed either a congruent or incongruent picture and they had to assess if the picture matched the nonword. For each nonword, there was 33 congruent and 33 incongruent trials (792 trials altogether). Behavioural responses as well as ERPs in response to the pictures were collected. In terms of ERP, we were looking for the N300 component, which typically appears in response to pictures in incongruent settings or following a semantically mismatching stimuli (Barret & Rugg, 1990). As such it was taken as an index of semantic processing of pictorial stimuli, similar to N400 for verbal stimuli. The results of mixed-effect linear regression analysis indicated that greater sensitivity to ART is related to greater accuracy rates and faster RTs on the Match-Mismatch Task. However, we found no effect of amplitude rise time detection on the N300 component measured on the pictures. These results suggest that sensitivity to amplitude rise time is connected to more efficient auditory learning of phonological word forms, but not necessarily to a better and more efficient semantic processing of the novel words in the brain.

References:

- Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16), 10911–10916.
- Goswami, U., Wang, H. L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, 23(2), 325–337.
- Marecka, M., Szewczyk, J., Jelec, A., Janiszewska, D., Rataj, K., & Dziubalska-Kolaczyk, K. (2018). Different phonological mechanisms facilitate vocabulary learning at early and late stages of language acquisition: Evidence from Polish 9-year-olds learning English. *Applied Psycholinguistics*, 39(01), 1–35.

Saturday Poster.8

**PARALLEL, CASCADED, INTERACTIVE PROCESSING OF WORDS DURING
SENTENCE READING: THE SENTENCE-SUPERIORITY N400 EFFECT**

Yun Wen, Joshua Snell, Jonathan Grainger

Laboratoire de Psychologie Cognitive, Aix-Marseille Université and Centre National de la
Recherche Scientifique, Marseille, France

yun.wen@univ-amu.fr

Reading is a fundamental cognitive skill unique to humans. It is known that humans can read 200 to 400 words per minute (Rayner, Schotter, Masson, Potter, & Treiman, 2016), although it is unclear yet whether this skilled reading involves a one-word-at-a-time incremental processing (Reichle, Liversedge, Pollatsek, & Rayner, 2009) or parallel, cascaded, and interactive processing (Rumelhart, 1977). Recent behavioural evidence in favour of parallel processing comes from one relatively unstudied phenomenon, the sentence superiority effect. Prior research has shown that single words are easier to identify in a briefly presented syntactically correct word sequence (e.g., the word “man” in the following sentence: the man can run) compared with a scrambled version of the same set of words (can man run the). This is the sentence-level equivalent of the well-known word superiority effect, and just like the word superiority effect reflects the parallel, cascaded transmission of information from letters to words, the sentence superiority effect is thought to reflect the parallel, cascaded transmission of information from words to sentence-level representations. Alternatively, this phenomenon can also be explained by incremental, one-word-at-a-time processing by appealing to extra-linguistic factors such as sophisticated guessing. That is, having identified the first word in the sequence “the man can run”, plus the first letter of the second word, if asked to report the second word in the sequence then participants could use this information to guess the identity of a 3-letter noun beginning with “M”. This kind of sophisticated guessing processes should operate relatively slowly compared with automatic word identification processes. Therefore, in order to rule-out the role of such factors, here we conducted the first electrophysiological investigation of the time-course of the sentence superiority effect. In this electrophysiological experiment, sequences of four words were presented briefly (200 ms) followed by a backward mask and a delayed (500 ms) cue. Participants were asked to report the words at post-cued locations by typing in their responses. Our results revealed a robust and widespread sentence-superiority effect on the N400 component of the event-related potential (ERP) that onsets around 270 ms post-stimulus onset, and therefore well before any sophisticated guessing could have occurred. Therefore, we reject an extra-linguistic explanation of the sentence superiority effect. To our knowledge, this study is the first to provide unequivocal evidence of parallel word processing during sentence reading by using a novel combination of electrophysiological recordings and the rapid parallel visual presentation of word sequences.

References

- Rayner, K., Schotter, E. R., Masson, M. I. E. J., Potter, M. C., & Treiman, R. (2016). So much to read, so little time. *Psychological Science in the Public Interest*, 17(1), 4-34.
- Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, 13(3), 115–119.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (pp. 573–603). Hillsdale, NJ: Erlbaum.

CASE AND (MIS)INTERPRETATION IN NUMBER ATTRACTION: EVIDENCE FROM EASTERN ARMENIAN

Serine Avetisyan, Sol Lago & Shравan Vasishth (University of Potsdam)
 serine.avetisyan@uni-potsdam.de

The production and comprehension of subject-verb agreement have been argued to involve memory retrieval: at the verb, cues are triggered to recall the subject in order to assign or license the verb in number (Badecker & Kuminiak, 2007; Wagers et al., 2009). But subject retrieval can fail when a noun (a so-called “attractor”) partially matches the number retrieval cues of the verb, leading to *attraction errors*: speakers erroneously produce ungrammatical verbs like in (1a), and readers process them faster than in the absence of an attractor (1b):

(1a) *The key to the **cabinets** are rusty. (1b) *The key to the **cabinet** are rusty.

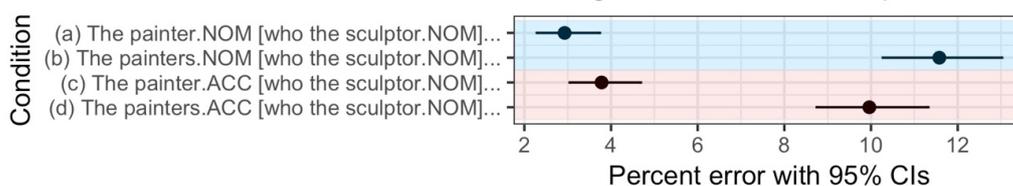
If case is used as a retrieval cue, overt case on nouns distinguishing subjects from non-subjects should reduce misretrievals and attraction errors. However, findings are mixed, with overtly case-marked nouns causing either reduced attraction (Hartsuiker et al., 2001; 2003) or increased attraction (Franck et al., 2006). Here we re-examined this issue in Eastern Armenian (EA), a language in which case is a reliable cue to grammatical role, because animate subjects and non-subjects are unambiguously marked with nominative and non-nominative cases. We first established the existence of number attraction in object-extracted relative clauses (ORCs), using two self-paced reading (SPR) studies. Then we adapted these items to re-examine whether case ameliorates attraction in a forced-choice task.

SPR experiments 1 (n = 48) and 2 (n = 46) manipulated the number of the RC head (SG/PL attractor) and whether the RC subject and verb (dis)agreed in number (grammatical/ungrammatical verb). Both the RC subject and head bore the same case (e.g. a,b). Both experiments replicated the classic finding of attraction errors in ungrammatical conditions, with no effects in grammatical conditions (Jäger et al., 2017). One novelty about the SPRs was that Experiment 2 asked participants who the subject of the RC verb was (e.g., ‘Who ignored?’) in order to establish their final interpretation of the subject-verb thematic dependency. Interestingly, participant’s typed responses indicated that the subject was often misinterpreted, and this effect was strongest in the attraction conditions: comprehenders either mistook the attractor for the subject (24%) or thought that the subject was plural (28%).

Experiment 3 (n = 176) used a web-based forced-choice task (Staub, 2009) to investigate whether case information on the NPs modulated attraction rates. Participants read ORC fragments in rapid serial visual presentation (SOA: 500ms) and had to continue them by choosing between the singular and plural forms of the critical RC verb (e.g. ignore_{SG} vs. ignore_{PL}). Half the conditions contained a nominative-marked attractor (e.g., a,b) and half contained an accusative-marked attractor (e.g., c,d). A Bayesian logistic mixed effects model showed that plural attractors led to more erroneous plural verb responses than singular attractors (Est(imate): 4.6%, 95% credible interval (CrI): [3.1, 6.3]%), but that the attraction rate was lower when the two nouns carried different case: When the two nouns carried same case, the attraction rate was 5.1% CrI: [3.2, 7.4]%, and when the two nouns had different case, the attraction rate was reduced to 3.8% CrI: [2.4, 5.5]%).

Overall, our results provide evidence of attraction errors in Eastern Armenian in constructions where the attractor does not linearly intervene between the subject and the verb. We also show that in ungrammatical sentences, the subject of the RC verb is often misinterpreted, being either confused with the plural attractor (Staub, 2009) or misrepresented as a plural noun (Patson & Husband, 2016). Finally, differential case marking on NPs denoting their grammatical roles reduces attraction effects only slightly.

Wrong-verb choices in Experiment 3



UNDERSTANDING CHANGES IN GARDEN-PATHS AS EXPECTATION ADAPTATION

Wednesday Bushong (wbushong@ur.rochester.edu), Zachary Burchill (zburchill@ur.rochester.edu), and T. Florian Jaeger (fjaeger@ur.rochester.edu)

Sentence processing seems to draw on implicit expectations about syntactic structures. Some theories hold that expectations are continuously *adapted* towards the syntactic statistics of the input^{1,2}, minimizing average surprisal.³ A few studies have found evidence qualitatively compatible with adaptation to new syntactic distributions.²⁻⁵ Principled *models* and their quantitative test against human data, however, have been lacking. We test a Bayesian belief-updating model against data from two garden-path reading experiments^{3,6} (N=77, 415 subjects; 71, 142 items, respectively). The experiments reported conflicting results. We find that both datasets are, in fact, captured by simple belief-updating. The syntactic priors inferred from the reading data are similar across experiments, *and* approximate syntactic statistics of language corpora.

Data. Both studies^{3,6} investigated the main verb (MV)/relative clause (RC) ambiguity, using the same block design (Fig 1), but different items and numbers of items. Subjects were randomly assigned to either the *RC-First* or *Filler-First* group. The RC-First group read only RCs in Block 1. The Filler-First group read only fillers. In Block 2, both groups read RCs and fillers. In Block 3, both groups read MVs. Half of the MV/RCs in each block contained the ambiguity (Latin-Squared).

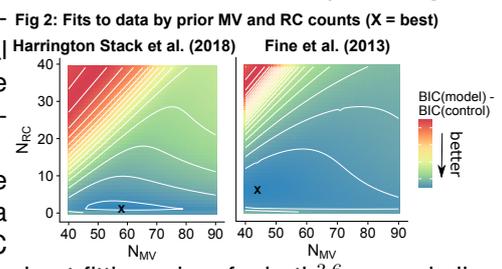
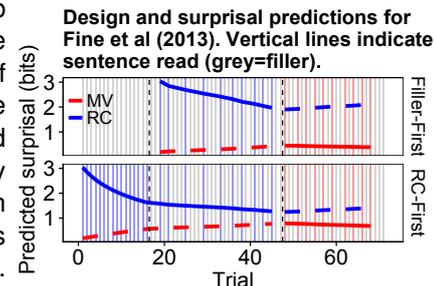
Block-based between-group (factorial) analyses found adaptation in [3] but not [6]. But these analyses do not take into account that [6] doubled the number of MV/RCs per block, changing the predicted expectation adaptation. We ask whether belief-updating explains both results.

Model prediction. The theory of expectation adaptation predicts that listeners incrementally adapt their expectations based on the frequency of MVs and RCs in the input.³ We operationalize this as beta-binomial belief-updating.⁴ This model has two DFs (inferred from the RT data): the prior MV and RC counts (N_{MV} , N_{RC}). The counts encode the prior probabilities of MVs and RCs (e.g., $P(RC) = N_{RC}/N_{RC} + N_{MV}$). The sum of the parameters captures how relevant listeners consider prior experience in the current situation. We then incrementally update expectations (and thus surprisal⁷) each time subjects read an RC/MV (Fig 1).

Analysis. We corrected RTs for word length and log trial order to remove the effects of adaptation to self-paced reading. We fit linear mixed models to both datasets, predicting RTs in the disambiguation region from surprisal, ambiguity, and their interactions. We compare the surprisal model to a control model predicting RTs from the design variables—group x block (structure) x ambiguity.

Results. For both^{3,6}, the surprisal model fits the data significantly better than the control, across a wide range of prior parameterizations (surprisal BIC < control BIC; blue and green regions in Fig 2). The best-fitting priors for both^{3,6} were similar, as expected if subjects on average hold similar prior experience, and thus beliefs ($N_{MV} = 44, 58$; $N_{RC} = 6.1, 1.1$ for [3,6], respectively; $\rightarrow P(RC) = 0.12, 0.01$). The inferred priors make sense: from natural language use, we would expect $\hat{P}(RC) = .011$.⁸

Conclusion. Bayesian belief-updating captures changes in RTs and garden-path effects, even for data reported not to show adaptation.⁶ The fact that the priors—inferred from comprehension data alone—match corpus data supports experience-based theories.⁹ Comprehenders seem to adapt their syntactic expectations to the statistics of recent input.



¹Chang et al 06-*PsyRev*; ²Wells et al 09-*CogPsy*; ³Fine et al 13-*PlosOne*; ⁴Myslin&Levy 16-*Cognition*; ⁵Ryskin et al 16-*JEP*; ⁶Harrington Stack et al 18-*Mem&Cog*; ⁷Hale 01-*ACL*; ⁸Roland et al 07-*JML*; ⁹MacDonald 13-*Frontiers*

Saturday Poster.11

**CODE-SWITCHING PATTERNS EN UN MODELO COMPUTACIONAL: SIMULATING
CODE-SWITCHING IN A BILINGUAL SENTENCE-PRODUCTION MODEL.**

Chara Tsoukala, Stefan Frank, Mirjam Broersma, & Antal van den Bosch (Radboud University)
c.tsoukala@let.ru.nl

People who speak several languages are able to switch from one to the other, a process called code-switching, between or even within sentences. The underlying mechanisms, however, are still not well understood; using computational modeling one can simulate code-switching behavior in multilinguals with the goal to explain the process. Hence, we have employed the bilingual Dual-path, a connectionist model of bilingual sentence production (Tsoukala et al., 2017; based on Chang's (2002) Dual-path model of sentence production). The Dual-path model is trained on message-sentence pairs and it learns to produce a sentence, word by word, given its semantic representation. For instance, the simple message "AGENT=DEF, WAITER; ACTION=EAT;" is expressed in English as "The waiter is eat -ing" and in Spanish as "El camarero está com -iendo".

Using the bilingual Dual-path model we have simulated sentence production in early Spanish-English bilinguals and late speakers of English who have Spanish as a native language. The difference between the models is that the early bilingual one is exposed simultaneously to Spanish and English, whereas the latter learns to produce Spanish sentences before getting exposed to English. We then manipulated language control to allow the model to produce sentences in either language or to code-switch. Interestingly, the model was able to produce code-switches even when it was not exposed to code-switched input.

The model predicts how code-switching patterns differ between early and late bilinguals. The early model code-switches much more frequently: 21.7% of sentences contained a code-switch as opposed to 2.3% in the case of the late model. Furthermore, most code-switches in the early models were complex mid-sentence switches (alternational code-switches; 10.2% as opposed to 0.6%), whereas the late models mainly inserted nouns from their L1 Spanish when producing English. Experimental works have not focused yet on a comparison between code-switched production of early and late bilinguals, but the results are in line with Poplack's (1980) findings; in the Puerto-Rican community in the US, balanced bilinguals produced complex code-switches, whereas Spanish-dominant bilinguals inserted tags and nouns. Using this cognitive model we can proceed in further examining the code-switching process.

References:

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26, 609– 651.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics*, 18(7-8), 581-618.

Tsoukala, C., Frank, S. L., & Broersma, M. (2017). "He's pregnant": simulating the confusing case of gender pronoun errors in L2. In G. Gunzelmann, A. Howes, T. Tenbrink, & E.J. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 3392-3397).

Saturday Poster.12

IS PUPILLOMETRY SENSITIVE TO ISLAND VIOLATION STRENGTH?

Christen N. Madsen II, Ian Phillips, Gita Martohardjono, Richard G. Schwartz
(City University of New York Graduate Center)
cmadsen@gradcenter.cuny.edu

In this study we focus on pupillometry to test its sensitivity to strong and weak island violations in processing (Kluender, 1997). Fernandez et al. (2018) uses pupillometry to investigate wh-gaps in native speakers and L2 learners of English but this methodology has rarely been extended to island violations (but see Fernandez and Engelhardt (2012)).

Five island conditions in Spanish were included as shown below. Sentence types 1-5 were given an increasing “complexity” code reflecting relative strength of violation (e.g., Chomsky, 1986; Sprouse & Hornstein, 2013): comp-trace= 0; wh-islands=1; CNPC = 2; RC and temporal adverbial adjuncts = 3. Stimuli were auditorily presented as natural running speech to 26 Spanish-English late bilinguals and pupil diameter was sampled at 60Hz.

Task-evoked pupillary responses (TEPR) were averaged in 250ms windows and analyzed using linear mixed-effects models in R, with grammaticality, complexity, and their interaction as fixed effects. Interaction of grammaticality and complexity was significant ($p < .05$) in windows 1–2 (0–500ms) and 5–9 (1250–2250ms). To further explore this interaction, TEPR were remodeled separately by grammaticality for these windows. This analysis showed TEPR was significantly modulated by complexity in all windows for ungrammatical items and in window 2 (250–500ms) for grammatical items.

In more complex (stronger) items ungrammaticality elicited an early immediate effect followed by a late and sustained effect. In weak items there was no significant difference between grammatical and ungrammatical. This suggests that pupillometry registers ungrammaticality for strong violations differently than for weak violations and that strong but not weak violations result in higher processing load. We discuss the implications of using this methodology for different types of syntactic complexity.

Sample Stimuli:

Ex. 1: Comp-trace condition

Qué hermana_k confesó Inés^{onset} [(que) ____k había comido la tarta?*
“Which sister did Inés confess (*(that) had eaten the cake?”

Ex. 2: Wh-island condition

Qué enfermera_k confirmó Ignacio^{onset} [(por) qué ____k había llevado la medicina?*
“Which nurse did Ignacio confirm (*(why/that) had taken the medicine?”

Ex. 3: Complex-NP complement condition

Qué insecto_k escuchó Gimena^{onset} [(el reportaje) que ____k invade el campo?*
“Which insect did Gimena hear (*(the news that) invaded the countryside?”

Ex. 4: Extractions out of RC

Qué cine_k ____k mostró el documental || que el crítico odiaba?
“Which cinema showed the documentary that the critic hated?”

**Qué crítico_k mostró el cine || el documental que ____k odiaba?*

***“Which critic did the cinema show the documentary that hated?”*

Ex. 5: Extractions out of temporal adverbial adjuncts

Qué juez_k ____k dio el veredicto || después que el defensor presentó su caso?
“Which judge gave the verdict after the lawyer presented his case?”

**Qué defensor_k || la juez_k dio el veredicto después que ____k presentó su caso?*

***“Which lawyer did the judge give the verdict after presented his case?”*

Saturday Poster.13

This is not only about decomposition: L2 learners process inflected words differently from native speakers

Kira Gor, Anna Chrabaszcz and Svetlana Cook
(kiragor@umd.edu)

We present the results of an auditory lexical decision task (LDT) with English-speaking late second language (L2) learners (N=34), heritage speakers (HS) (N=25), and native speakers (NS) of Russian (N=39). L2 learners and HSs were matched in oral proficiency using the Interagency Language Roundtable (ILR) scale. Both L2 learners and HSs of Russian experienced difficulties (high error rates and increased reaction times) in detecting violations in nonce nouns with incongruent real stems and real inflections (nonce-inflection in Fig 1 and 2), while they were efficient at rejecting nonce nouns with nonce stems (nonce-stem in Fig 1 and 2). A developmental trajectory in nonnative sensitivity to the congruence of noun stems with case inflections in nonce words indicates that while the accuracy rate increases at higher proficiency levels, it significantly lags behind the accuracy rate of NSs even at the Superior level of proficiency (ILR 3 or CEFR C2, see Fig. 2). Based on the reported results and together with our previous findings, we propose a new model of nonnative processing of inflection: The critical difference between NS and L2/HS processing of inflected words is associated not with the initial stage of decomposition—word parsing into stems and inflections—which successfully takes place in all listeners and permits them to reject nonce nouns with nonce stems. It is associated with the later recombination stage, when L2 and HSs fail to detect stem-inflection incongruence, and arguably, to process the morphosyntactic information carried by the inflection. Both L2 and HSs gradually develop sensitivity to the congruence of real inflections with real stems to which they are attached, and engage the morphological information carried by the inflections only at high proficiency.

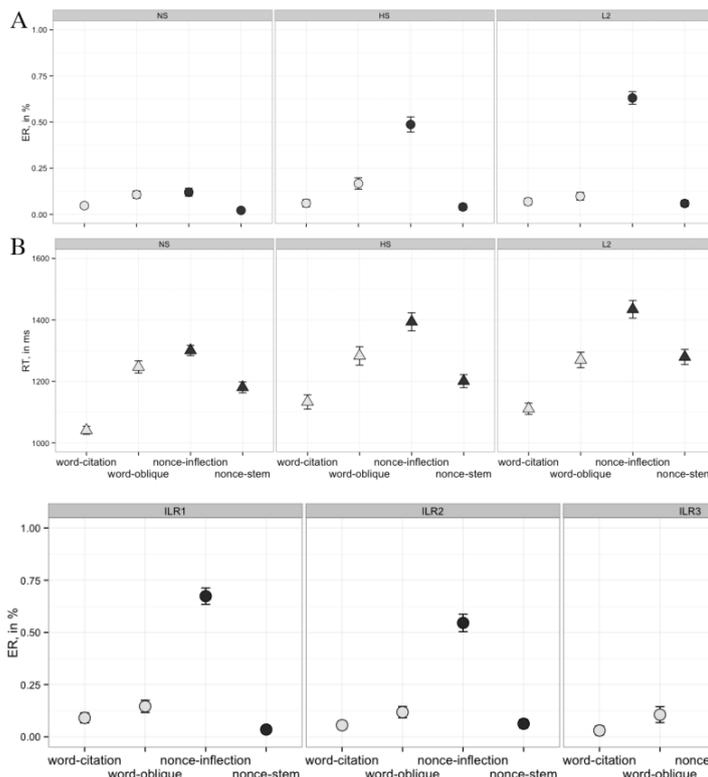


Figure 1. Error rates (Panel A) and reaction times (Panel B) in three participant groups for two types of real and two types of nonce nouns

Figure 2. Error rate in response to real and nonce nouns: A developmental trajectory
Figures correspond to the following oral proficiency levels:
1 – Intermediate; 2 – Advanced; 3 – Superior, and 5 – Native Speaker.

SENSITIVITY TO LANGUAGE STATISTICS IN 1ST AND 2ND LANGUAGE READING

Stefan L. Frank (Radboud University) & Robin L. Thompson (University of Birmingham)
s.frank@let.ru.nl

Proficient language comprehension is highly sensitive to a language’s statistical patterns. For example, the time spent reading a word is linearly related to its surprisal; the negative logarithm of the word’s probability given the sentence so far (Smith & Levy, 2013). We investigated whether the use of internalized language statistics differs between monolinguals and bilinguals and between first and second language reading. We compared word-reading times between two groups of native English speakers (monolinguals ($N = 21$) and bilinguals ($N = 19$), mostly heritage speakers) and proficient non-natives with L1 either Dutch ($N = 20$) or British Sign Language (BSL; $N = 27$). Reading times (RTs) were measured using eye-tracking on 205 English sentences (1931 word tokens) sampled from novels to be representative of the written language (Frank et al., 2013). Word surprisal values were computed by n -gram models ($n = 2, \dots, 5$) that estimate word probability from frequencies of word sequences (up to length n) in an English corpus. Goodness-of-fit of log-transformed first-pass RTs to surprisal was quantified by linear mixed-effects regression, including a predictor for previous word surprisal to capture delayed effects of reading difficulty (i.e., spillover), and predictors for word frequency and length (among others) as covariates.

L2 readers show weaker spillover than native speakers, possibly because reading is less fluent for the non-dominant language (Fig. 1). In all groups, larger n leads to a better fit, indicating veridical knowledge and use of frequencies of longer sequences of words (Fig. 2). Surprisal predicts L1 readers’ RTs more accurately than those of the other groups (Fig. 2). This suggests that the relation between RTs and a language’s statistical properties is weakened by knowledge of any other language, even if it is the L2 or (like BSL) has no orthography.

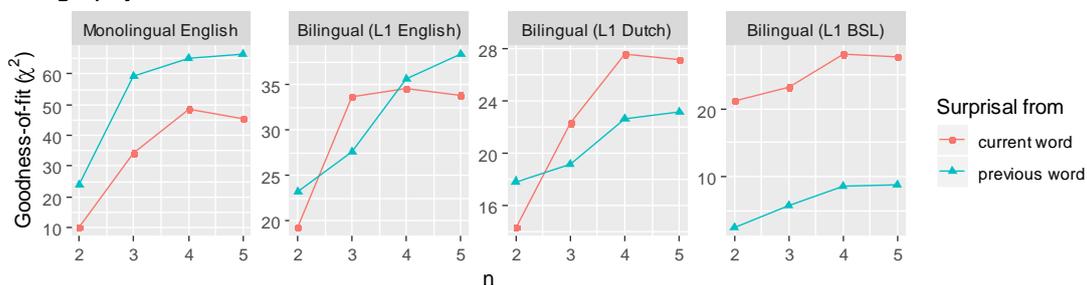


Figure 1: Goodness-of-fit of surprisal to RTs, as a function of the maximum word-string length (n) for surprisal computation. Each point is the outcome of a log-likelihood ratio test comparing a regression model that includes both current and previous word surprisal, to a regression model that includes only one of the two (i.e. effect of one over and above the other). Note that χ^2 cannot be compared between groups because the data sets have different sizes.

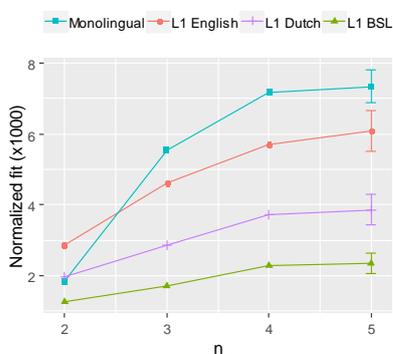


Figure 2: Goodness-of-fit of current and previous word surprisal combined, after correcting for differences in data set size between groups (χ^2 divided by number of data points). Error bars are bootstrapped 95% CIs.

References

Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. (2013). Reading-time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45, 1182–1190.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128, 302–319.

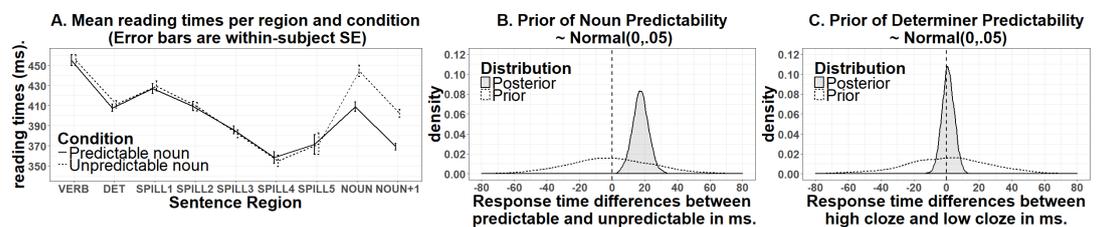
A CRACK IN THE CRYSTAL BALL: EVIDENCE AGAINST PRE-ACTIVATION OF GENDER FEATURES IN SENTENCE COMPREHENSION

Ernesto Guerra (Universidad de Chile), Bruno Nicenboim (University of Potsdam) & Andrea Helo (Universidad de Chile)
ernesto.guerra@ciae.uchile.cl

In the past decade, the idea that prediction is part and parcel of language processing has turned from a controversial hypothesis to a widely accepted assumption. In an influential paper, DeLong et al. (2005) reported a distinct electrophysiological response (i.e., a larger negativity around 400 ms) for indefinite English articles that mismatched phonological features of a high-cloze noun (e.g., 'an' when 'kite' was predicted) compared to articles that matched the noun (e.g., 'a'), effect that correlated with the cloze probability of the article. Similar results have been reported in Dutch and Spanish (van Berkum et al., 2005; Wicha et al., 2004) for event-related potentials (ERPs) and self-paced reading (SPR). These studies exploited the gender markedness of nouns to examine the (behavioral and ERPs) responses to gender-marked articles and adjectives preceding a critical noun. More recently, however, a large-scale multi-lab pre-registered study (Nieuwland et al., 2018) challenged the notion that predictions occur at all levels of linguistic representation.

In a SPR experiment, we examined whether there is reading times (RTs) evidence for predictions occurring at the morpho-syntactic level during sentence comprehension. Previous research has shown that SPR experiments are sensitive to prediction at other levels (see Levy, 2008). Thus, if readers pre-activate gender information, we would expect faster RTs *before* the predicted noun when a determiner matches the gender of the noun. If not, we would expect shorter RTs only upon encountering the predictable noun. We collected data from a large sample of participants (n=120) and items (n=60). Each item had two versions; one with a predictable gender-marked noun and another with an unpredictable noun of the opposite gender (e.g., 'Homer will drink *the*_{FEM/MASC} *refreshing*_{SPILL1}, *though*_{SPILL2} *not*_{SPILL3} *so*_{SPILL4} *cold*_{SPILL5} *beer*_{FEM} *whisky*_{MASC}...'). Using Bayesian linear mixed models with log-transformed RTs, we examined the effect of predictions at the noun, determiner, and spillovers of the determiner. For the noun region, we used condition (predictable vs. unpredictable) as a predictor. For the determiner (and spillover regions), we used the log-probability of the gender (obtained from an independent sample, n=100), and the gender of the determiner as a random variable. We used different priors centered in zero to assess the strength of evidence of predictions in our models (see Tables below).

BF for the effect at the NOUN.					BF for the null effect at the DETERMINER.				
Prior	95% prior interval	Estimate	95% CrI	BF10	Prior	95% prior interval	Estimate	95% CrI	BF01
~ Normal(0, .1)	[-102, 102] ms	17 ms	[8, 26] ms	427	~ Normal(0, .1)	[-102, 102] ms	1 ms	[-6, 9] ms	21
~ Normal(0, .05)	[-51, 51] ms	17 ms	[7, 27] ms	51	~ Normal(0, .05)	[-51, 51] ms	1 ms	[-6, 9] ms	9
					~ Normal(0, .01)	[-10, 10] ms	1 ms	[-6, 7] ms	2



As it is evident from the graph (panel A), the effect of predictability clearly appears, but only upon reading the noun. For the noun region, there is very strong evidence in favor of an effect of cloze (panel B). Critically, for the determiner region, with realistic priors there is substantial evidence *in favor* of a null effect (panel C). Only when assuming a very small effect as a prior, the evidence in favor of a null effect is inconclusive. The same pattern is found in the spillover regions after the determiner. Consequently, our study provides substantial evidence for the absence of a meaningful behavioral effect of gender-based predictions during reading. *If* readers use the contextual gender information, it has virtually no effect in RTs before the upcoming noun.

Saturday Poster.16

**SEMANTIC INTERFERENCE AND MORPHOLOGICAL FACILITATION IN NOUN-NOUN
COMPOUND PRODUCTION: EVIDENCE FROM EVENT-RELATED BRAIN POTENTIALS**

Antje Lorenz (Humboldt-Universität zu Berlin), Stefanie Regel (Humboldt-Universität zu
Berlin), Pienie Zwitserlood (Westfälische Wilhelms-Universität, Münster), & Rasha Abdel

Rahman (Humboldt-Universität zu Berlin)

antje.lorenz@hu-berlin.de

We examined how compounds are lexically stored and processed during speech production. Predictions of the two-stage model, which assumes sequential access to holistic compound lemmas and morpheme-based form representations (Levelt et al., 1999), and its modifications (e.g., ‘multiple-lemma representation account’, Marelli et al., 2012) were tested. Using a picture-word interference paradigm, written distractor words were superimposed onto pictures that were named with compound words. Distractors either overlapped with the compound’s first or second constituent (sun / flower → sunflower), were categorically related to the compound (tulip → sunflower), or to its first constituent (moon → sunflower). In addition to picture-naming latencies, the continuous electroencephalogram (EEG) was extracted. Speech artefacts in the EEG signal were eliminated by using residue iteration decomposition (RIDE; Ouyang et al., 2016), and event-related brain potentials (ERPs) were analysed. Naming latencies confirmed morpho-phonological facilitation for both constituents of compound targets, and semantic interference for compound-related distractors. However, no effects were obtained for distractors that were semantically related to the first constituent of the compound. The EEG data complement the behavioural findings: Distractors from the same semantic category as the compound induced an early posterior positivity, probably reflecting lexical competition in speech production. An additional left-frontal negativity may be related to enhanced cognitive control in the resolution of lexical competition. Distractors from the same category as the compound’s first constituent showed no such effects, supporting single, holistic compound lemmas. Morphological distractors produced a positivity at midline electrodes, probably reflecting facilitation of morpho-phonological encoding. The morphological positivity of first constituents overlapped temporarily with the semantic effect for the whole compound. In contrast, the positivity induced by second-constituent distractors started approximately 50 ms later, and was followed by a frontal negativity. Latency and EEG data corroborate single-lemma, but multiple-morpheme representations for compounds in production.

Saturday Poster.17

LOW-LEVEL VOCAL CUES AFFECT THE ACQUISITION OF HIERARCHICAL STRUCTURE

Antony S. Trotter (Lancaster University), Padraic Monaghan (Lancaster University), & Rebecca Frost (Max Planck Institute for Psycholinguistics)
t.trotter@lancaster.ac.uk

Hierarchical centre embeddings (HCEs) in natural language have been taken as evidence that language is not a finite state system (Chomsky, 1957). Whilst phrase structure may be necessary to produce HCEs, sequential processing may underpin their comprehension (Frank, Bod, & Christiansen, 2012). In this theory, low-level statistical correspondences – e.g. pitch and rhythm variation – provide support for dependency detection. Trotter, Monaghan, and Frost (2017) found that in spontaneously produced HCEs, phrases in the embedded clause are similar in terms of pitch and preceded by a longer pause than elsewhere in the speech, concluding that the pitch similarity and temporal proximity Gestalts are particularly relevant to the comprehension of phrasal clauses. Prior artificial language research has encountered difficulties in demonstrating the acquisition of phrase structure grammar. However, the languages used in these studies rarely feature the range of natural language cues that support the processing of dependencies between words. In this study, we assessed whether pitch similarity and temporal proximity would enhance learning in an artificial language learning study.

64 native English speakers were trained on an artificial grammar containing HCEs. Participants were presented with one ($A_1A_2B_2B_1$) and two ($A_1A_2A_3B_3B_2B_1$) levels of embedding (LoE) sequences, where syllable A_i always co-occurred with B_i in the corresponding position (Lai & Poletiek, 2011). Participants were assigned to one of four cue conditions: baseline, pause, pitch and combined (pause + pitch). For baseline, the only useful cues were the statistics of the grammar. The pause condition employed temporal grouping cues; pauses occur between levels of embedding (e.g. A_1 [pause] A_2B_2 [pause] B_1). For pitch, items within a LoE used similar pitch, with 15Hz difference between levels. Participants received 12 blocks of training and testing, which exposed them to 16 grammatical sequences, then performed a grammatical classification task on 16 novel sequences.

Analyses were conducted using linear-mixed effects modelling. The models included the main effects of block and cue condition, their interaction, and random by-items and -subjects intercepts. Individual models were run for each LoE. For one LoE, there was a significant negative block by pause cue interaction ($\beta = -1.44$, $p = 0.0489$); initially, pause cues boosted performance, which reduced over training. For two LoE, there was a marginally significant effect ($\beta = -0.102$, $p = 0.057$) of combined cues. Why should greater exposure to pause cues diminish performance? One possibility is tension between local (linguistic) and global (temporal) structure. Acoustic cues were present in both training and testing; this may reduce continued attention to the pause cue as reflecting grammatical structure though pauses may provide a head start with the grammar by highlighting particular dependencies. In natural language, diversity of syntactic forms with accompanying prosodic variation may avoid this problem. Nevertheless, we have shown that prosodic cues improve grammar learning of HCEs enabling phrasal groupings to be computed from lower-level auditory processing biases.

References

- Chomsky, N. (1957). *Syntactic Structures*. Oxford, England.
- Frank, S. L. Bod, R. & Christiansen, M. H. (2012). How Hierarchical is language use? *Philosophical Transactions of the Royal Society B*, 279, 4522 – 4531.
- Lai, J. & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures *Cognition*, 118, 265 – 273.
- Trotter, A. S., Monaghan, P., & Frost, R. L. A. (2017). *Auditory-perceptual gestalts assist in the processing of hierarchical structure*. 23rd Annual Architectures and Mechanisms of Language Processing, Lancaster, UK.

AUDITORY-PERCEPTUAL GESTALTS SUPPORT THE PROCESSING OF PHRASE STRUCTURE IN COMPREHENSION

Antony S. Trotter (Lancaster University), Padraic Monaghan (Lancaster University), & Rebecca Frost (Max Planck Institute for Psycholinguistics)
t.trotter@lancaster.ac.uk

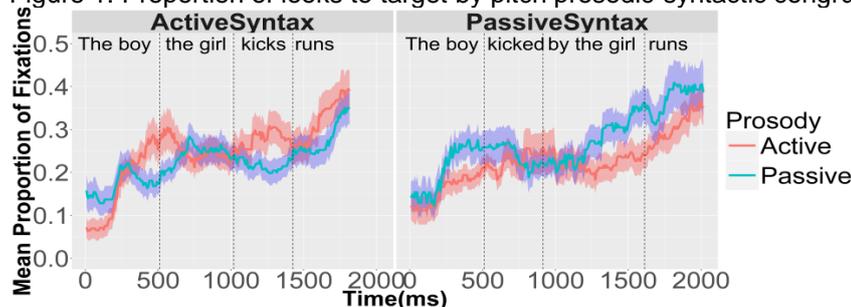
Processing and responding to speech requires rapidly determining the structural dependencies between words to comprehend meaning. While phrase structure may be necessary for producing syntactically complex sentences, it has been argued that sequential processing may be sufficient for comprehension, with low-level statistical correspondences providing critical support for dependency detection (Frank, Bod, & Christiansen, 2012). Two auditory Gestalt grouping mechanisms are particularly relevant to the comprehension of phrasal clauses; temporal proximity (pauses between clauses will render them distinct if they are longer than elsewhere in the speech, supporting their grouping) and pitch similarity (constituents within clauses will be similar in pitch, supporting grouping). Trotter, Monaghan, and Frost (2017) found that for spontaneously produced hierarchical centre-embedded (HCEs) structures, phrases in the embedded clause were similar in pitch, and preceded by a longer pause than elsewhere in the speech (e.g. The boy [pause, pitch reduction] the girl hugged is green). Passives differed; a longer pause and a pitch reduction occurred after the verb phrase of the medial clause (e.g. The boy being hugged [pause, pitch reduction] by the girl is green). HCEs are challenging to process, but here we see evidence that in speech, temporal proximity and pitch similarity provide grouping cues for tracking dependencies. Whilst these cues are present in speech production, are they useful in comprehension?

Using a visual world paradigm, we assessed whether temporal proximity and pitch similarity facilitate the processing of HCE and passive constructions. Prosody was manipulated to be congruent or incongruent with HCE or passive structures. In 128 trials, 32 participants previewed four potential targets for the sentence, each showing agents and patients performing actions, three of which were distractors (agent-verb violation, patient-verb violation, role reversal). Then, participants heard the sentence describing the event, and indicated which scene it described with a key press. Analysis of participant fixations indicates that when pitch-prosodic and syntactic structure are congruent, participants make a higher proportion of looks to target at an earlier timepoint, indicating that prosody affects online syntactic processing (see Figure 1).

References

- Trotter, A. S., Monaghan, P., & Frost, R. L. A. (2017). *Auditory-perceptual gestalts assist in the processing of hierarchical structure*. 23rd Annual Architectures and Mechanisms of Language Processing, Lancaster, UK.
- Frank, S. L. Bod, R. & Christiansen, M. H. (2012). How Hierarchical is language use? *Philosophical Transactions of the Royal Society B*, 279, 4522 – 4531.

Figure 1: Proportion of looks to target by pitch prosodic-syntactic congruency



Saturday Poster.19

Online response to perspective taking in narratives

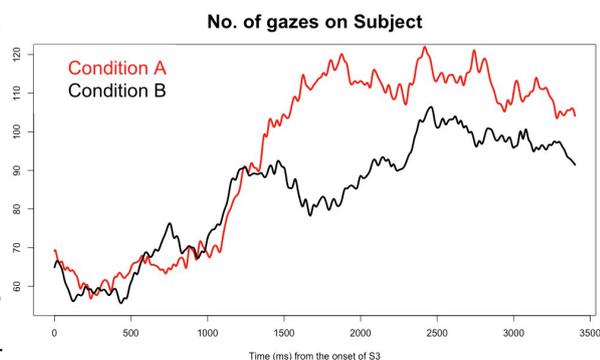
Sara Meuser (University of Cologne), Umesh Patil (University of Cologne) & Stefan Hinterwimmer (University of Cologne)
smeuser2@uni-koeln.de

We test the hypothesis that the most prominent referent is more likely to be the perspectival center of a discourse compared to competing referents. We report an eye-tracking experiment (n=40) in the visual-world paradigm on the processing of perspective taking in narratives. We use free indirect discourse (FID) as the main indicator of a shift in perspective. FID, defined as “mode of narration where we can listen to a protagonist’s thought” (Eckardt, 2014), can only be processed correctly if the reader is able to take the protagonist’s perspective. Characteristics of FID are: interjections, judgmental statements, exclamations, discourse particles, rhetorical questions and a shift in deixis with respect to the discourse referent. For example, in “*Last Friday John wanted to go to a concert. Right before he got to the concert hall he checked his pockets. Oh no, had he really left the tickets at home? Now it was too late!*” the rhetorical question as well as the exclamation can be easily understood as expressing a thought of John because of the interjection *Oh no* and the deictic expression *now*. FID has been a topic of interest for literary scholars as well as linguists, but there is very little empirical research on the processing of FID. Also, lately there has been growing interest on the impact of perspectives in language comprehension, e.g. the interface of Theory of Mind and pragmatics; however, the shift in perspective in narratives has not yet been studied through online experiments.

In our experiment we test processing of perspective taking in FID. Test items (see (1) below) consist of short stories introducing one highly prominent protagonist with a proper name in subject position (R1=*Martin*) and a second minimally prominent referent with an indefinite article in object position (R2=*florist*). In the third sentence we compare an utterance in FID (cond-a) involving at least three indicators — interjection, a deictic expression and a discourse particle, to a sentence of similar content in narrative style (cond-b). Along with R1 and R2 we show pictures of a distractor mentioned in the story and an unmentioned distractor. The hypothesis predicts more gazes on R1 in cond-a compared to cond-b as a result of the change towards R1’s perspective triggered by the FID. As we use several different indicators of FID, we expect the effect to show as the third sentence (S3) unfolds. Figure below shows the number of gazes recorded on R1 from the onset of the critical S3. We modeled the gazes during the first 3400 milliseconds (the mean duration of S3 across items) of S3 using the growth curve analysis (interaction of two experimental conditions with all the terms of a fourth-order orthogonal polynomial as fixed effects). We observed significantly higher number of gazes on R1 in cond-a (significant effect on the intercept term). There was also a significant effect on each of the other three terms of the polynomial indicating a clear effect of perspective taking between cond-a and cond-b. Our findings suggest that more gazes on the most prominent referent in cond-a are due to the ascription of authorship of the utterance in FID. We regard these results as a proof of concept for further research employing the visual-world paradigm for investigating prominence status and perspective taking in language processing.

(1) Martin fragte einen Floristen an einem Stand nach einem Blumenstrauß. Der Geruch der Blumen lockte jedoch einige Wespen an.
[cond-a] Oh, jetzt bloß ganz ruhig stehen bleiben, um die Mistviecher nicht zu reizen.
[cond-b] Um Wespen nicht zu reizen, sollte man ruhig stehen bleiben.

Martin asked a florist at a stall for a bouquet. The smell of the flowers attracted some wasps.
[cond-a] *Oh, better stand still now, not to aggravate those beasts.*
[cond-b] *As not to aggravate wasps one better stand still.*



EVIDENCE FOR SYNTACTIC TRANSFER FROM LANGUAGE AND MUSIC

Mythili Menon and Drew Colcher, (Wichita State University)

Introduction. The effects of language on musical processing has received very little attention. Existing results reveal a discrepancy between differences of perception of pitch versus perception of pitch intervals as shown by studies looking at tonal language speakers versus non-tonal language speakers (Pfordresher & Brown'09, Deutsch et al.'06). These mixed results point to parallels between language processing and musical processing. In this study, we report results from a priming comprehension task from language to music pointing to the existence of a shared cognitive process underlying musical and linguistic syntax.

Our study has two main aims. We investigated (i) whether comprehension of subject-verb agreement errors transferred to dissonance of musical chords played on *guitar* (stimuli that are easily recognized by listeners as being part of the *domain of music*), and (ii) whether comprehension of subject-verb matches transferred to congruence of musical chords.

Experiment. We used a 2 x 2 design with a) agreement match (match/mismatch) and b) ambiguous number (singular/plural) as factors. We provided participants (N= 18, 20 targets, 20 fillers) with subject-verb agreement sentences such as the following:

Linguistic primes (inanimate nouns, counterbalanced for number):

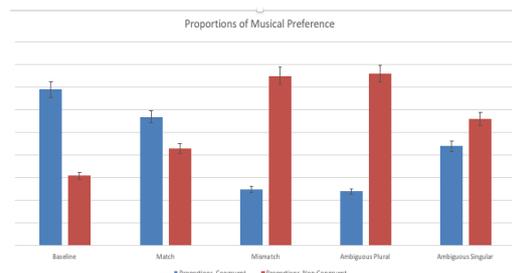
- (1) [The **key**] to [the cabinets] [**was** rusty from many years of disuse] *Match condition*
- (2) [The key] to [the cabinets] [**were** rusty from many years of disuse] *Mismatch condition*
- (3) [The **screen**] of [the **phone**] [**was** cracked from top to bottom] *AmbgSingular condition*
- (4) [The mugs] on [the shelves] [**were** still wet from being washed] *AmbgPlural condition*

Participants read a sentence and answered a comprehension question, following which they heard two musical targets. They were asked to choose which target they preferred. We also included a baseline condition to find which musical targets people preferred overall.

Musical targets: The musical targets were structured using western music theory in the major diatonic scale in two keys - G and C (G major being the base key for guitar) - as well as their relative minor keys (E and A minor, respectively). They consisted of a series of 7 chords (including repeated chords), the last of which acts as a coda that refers back to the body of the piece in a particular way. In the Match condition, all of the chords in the series strictly adhere to scale theory and fit in the given key without the inclusion of any variations. The coda of these Match targets is a repeat of the 2nd or 3rd chord in the series. In the Mismatch condition, the final "coda" chord contains notes that are outside of the 7-note structure of the key in which the rest of the 6 chords are played. In AmbigSingular condition, the "coda" chord contains tonal variations to the scale, that are canonical, recognizable variations in the realm of western music (a so-called "7 chord" in the 5th degree of a minor scale, e.g.); in Ambg Plural, both the coda and one of the body chords of the series (2nd or 3rd chord, depending on the sample) contain such variations. Baseline samples were also recorded, consisting of either 3 or 4 chords in one of the aforementioned keys, with no ambiguity and a different structure than critical conditions.

Results. We find an overlap between syntactic processing and musical processing. In the baseline, we find an overall preference for the congruent musical target. Furthermore, in the subject-verb Match condition, participants continue to prefer the congruent musical condition over the non-congruent target. On the contrary, in the subject-verb Mismatch condition, we find participants prefer the non-congruent musical target significantly. We find no differences between the Ambiguous Singular condition and the Ambiguous Plural condition. In both cases, participants prefer the non-congruent target.

Discussion. Our results show that subject-verb agreement errors activate hierarchical representations that overlap with the syntactic representations of explicitly musical stimuli (guitar notes) on an abstract level. We show that structure is crucial in overlap between harmonic domains. **IN SUM**, The results of our experiment provide striking evidence for the domain-general level of abstraction in the representation of hierarchical structural information.



Saturday Poster.21

Placing Pronouns

Rhythm affects Word Order Preferences in German

Isabelle Franz (MPI for Empirical Aesthetics), Markus Bader (Goethe University Frankfurt) & Gerrit Kentner (MPI for Empirical Aesthetics)
isabelle.franz@ae.mpg.de

We present two questionnaire studies with binary preference judgements, that extend previous work on syntax-prosody interactions and suggest influences of rhythm on word order of object pronouns in speech. 74 participants chose one of two sentences (presented in writing) which varied w.r.t word order and rhythmic well-formedness.

The first study concerns the order of the object pronoun 'ihn' ('him') and the embedded subject (1) in subordinate clauses. Apart from the order, we manipulated stress of the embedded subject (iambic or trochaic). Over and above a strong bias for SO-sentences (see table 1), participants preferred the object pronoun to be fronted ($z = 2.15$, $p < 0.01$), when this promoted a rhythmic sequence – that is, in sentences with a trochaic embedded subject.

- (1a) Péter sagt, dass Márkus/Marcél **ihn** árgert.
 (1b) Péter sagt, dass **ihn** Márkus/Marcél árgert.
Peter says that Markus/Marcel is nettling him.

Stress pattern embedded subject	presentation order	response pronoun first in %	response pronoun first both orders in %
iamb 'Marcél'	OS-first	08.1	12.8
	SO-first	17.6	
trochee 'Márkus'	SO-first	14.9	17.9
	OS-first	21	

Table 1: Responses in percent for the preferred sentences with the pronoun in fronted position

The second study replicates findings by Vogel et al. (2015) who examined the relative order of objects and pronomial adverbs in sentence production. The rhythmic manipulation concerns stress on the trisyllabic object (intital stress vs penultimate stress). Participants showed a strong bias for the sentences with the adverb following the object (see table 2). Also, fronted adverbs were more likely when the object featured initial stress ($z = -4.31$, $p < 0.0001$), promoting a dactylic rhythm.

- (2a) Da könnte der Stéffen Konfétti/Lúftschlangen **draus** básteln.
 (2b) Da könnte der Stéffen **draus** Konfétti/Lúftschlangen básteln.
There Steffen could make confetti/streamers from.

Stress pattern object	presentation order	response adverb first in %	response adverb first both orders in %
antepenult 'Lúftschlangen'	adverbFrontFirst	23.7	23.9
	objectFrontFirst	24.2	
penult 'Konfétti'	adverbFrontFirst	13.2	10.9
	objectFrontFirst	08.5	

Table 2: Responses in percent for the preferred sentences with the pronoun in fronted position

Both studies show clear influences of rhythm on syntactic preferences, suggesting an interaction of phonological and grammatical encoding in speech. We currently conduct a picture naming experiment eliciting the structures in (1) and (2).

References:

Vogel, Ralf; Ruben van de Vijver; Sonja Kotz; Anna Kutscher & Petra Wagner. 2015. *Function words in rhythmic optimization*. In: Vogel, Ralf & Ruben van de Vijver (eds.) *Rhythm in Cognition and Grammar. A Germanic Perspective*. Walter de Gruyter GmbH. Berlin/Munich/Boston. 253-274.

Processing in Parallel: Single-Trial EEG at the Phonology-Morphology Interface

Laurel A. Lawyer (University of Essex)

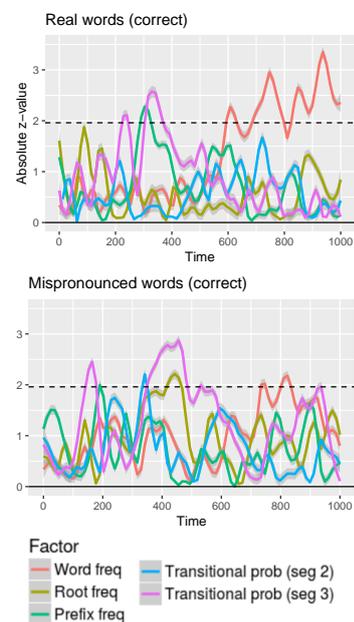
l.lawyer@essex.ac.uk

Introduction. Models of speech perception vary in the degree to which levels of processing (e.g., phonetic, lexical, syntactic) operate autonomously. In fully isolating models (e.g., Fodor, 1983; Norris et al., 2004), phonological information is used to facilitate lexical access but does not feed forward into further levels. However, other models allow activation may spread between levels in either direction during processing (e.g., McClelland & Elman, 1986; Hickok & Poeppel, 2007). Using real and mispronounced morphologically complex words, I use single-trial EEG data to investigate the time course of phonological and morphological processing in a lexical decision task.

Methods. In an ERP experiment, subjects (N=28) were asked to identify correctly pronounced words from a group of 40 complex real words and 80 mispronounced words. Mispronounced items were created by substituting a single segment in the prefix of an existing complex word. A linear mixed-effects model was estimated for every sample point in single-trial EEG data (435 samples per trial, covering -200 to 1400 msec.). The input to the model for each time point contained the amplitude at each electrode site, for each word, for each subject. Lexical predictors were word frequency (CELEX; Baayen et al., 1995), and residualised prefix frequency and root frequency. For mispronounced words, frequency values for the original words were used. Phonological predictors were transitional probabilities for the second segment (TP2), and for the third segment (TP3) in each word, covering the end of the prefix and beginning of the root of each word. Subject, item, and electrode site were included as random effects. Z-values exceeding 1.96 (estimating $p < .05$) for more than three consecutive time points were considered significant.

Results. Subject performance showed good discrimination, with 94% of real words and 87% of mispronounced words correctly labelled. In correct word responses, EEG signal amplitudes were predicted by three factors: TP3 (from 210-250msec and 300-380msec), prefix frequency (280-320msec), and word frequency (600 to 1050msec). In correct mispronounced word responses, both TP2 (320-350msec) and TP3 (140-180msec, and 330-480msec) significantly predicted EEG amplitude, as did prefix frequency (190-200msec), root frequency (420-460msec), and word frequency (730-830msec).

Conclusions. The analysis of factors influencing EEG responses to complex morphological words illustrates both the serial nature of decompositional parsing as well as bi-directional, interactive parsing. For lexical factors, the frequency of prefixes, roots, and whole words effect EEG amplitude in turn. This ordered influence supports feed-forward decompositional models of complex word processing. However, in looking at phonological factors, the significant contributions of TP3 in both early and later time windows suggests a more interactive processing route. The early effect of TP3 occurs in a time window that overlaps with effects found in studies of phonological mismatch (cf. Connolly & Phillips, 1994). However, the secondary influence of TP3 suggests that the processing of phonological sequences is not single-staged, but rather that it occurs in tandem with lexical processing. In this case, knowledge of specific phonological detail remains available and may be used as a means to discriminate nonwords even while lexical access is in progress.



Saturday Poster.23

**INCREMENTAL LEARNING IN WORD PRODUCTION:
TRACING THE FATE OF NON-SELECTED ALTERNATIVE PICTURE NAMES**

Jörg D. Jescheniak (Leipzig University), Franziska Kurtz (Leipzig University), Herbert Schriefers (Radboud University, Nijmegen), & Andreas Mädebach (Leipzig University)
j dj@uni-leipzig.de

In contrast to classical models of word production (e.g., Dell, 1986; Levelt, 1999), recent models view lexical retrieval as an adaptive process. The principle idea is that selecting an (intended) lexical representation (e.g., “bird”) increases the accessibility of this representation for future retrieval, whereas co-activated, but not selected representations (e.g., “cat”) become less accessible for future retrieval due to changes in the connection strength within the conceptual-lexical network (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell & Schwartz, 2010). Evidence for such an incremental learning mechanism mainly comes from the finding of semantic interference in continuous and blocked-cyclic picture naming experiments. Specifically, it was shown that the retrieval of a picture name is slowed down (e.g., “cat”), if category coordinates have been named on the preceding trials (e.g., “bird”, rabbit”, “guinea pig” etc.; see Howard et al., 2006; Damian, 2003). In the present study, we asked whether such context-dependent changes in lexical accessibility can also be demonstrated for true naming alternatives (e.g., “bird” when participants name the picture of a specific bird by saying “duck”) which is to be expected under the incremental learning view. Our starting point was the observation that alternative picture names become phonologically co-activated even when they are eventually not produced (e.g., “bird” when “duck” is produced). This can be inferred from the fact that in experiments using the picture-word interference (PWI) task distractor words which are phonologically related to the alternative name (e.g., “birch”) slow down picture naming more strongly than unrelated distractor words (e.g., “lamp”; e.g., Jescheniak, Schriefers, & Hantsch, 2005). On the incremental learning view this interference effect should be attenuated, when speakers consistently only use one particular name (e.g., “duck”) in a large number of naming episodes. This is because the co-activation of the non-selected alternative name should be reduced due to changes in connection strength.

We conducted a series of PWI experiments (with 32 participants each) in which picture naming latencies were measured. The experimental pictures were photographs of simple objects that were preferentially named at the subordinate level, as established in a norming study (2 exemplars each from 16 basic level categories). Each picture was named 25 times (or 50 times in one of the experiments). At different points in time, we assessed the phonological activation of the alternative name and of the target name itself with auditory distractor words (SOA 0 ms). We implemented two contrasts:

(a) Distractor words phonologically related versus unrelated to the alternative name (e.g., “birch” vs. “lamp” when the target word was “duck”) – interference from related distractors would reflect the phonological co-activation of the alternative name (“bird”) and was expected to decrease in the course of the experiment due to changes in connection strength.

(b) Distractor words phonologically related versus unrelated to the target name (e.g., “dust” vs. “shelf”) – facilitation from related distractors would reflect the phonological activation of the target name and was expected to remain unchanged because on each naming episode the target word needs to be phonologically encoded anew (control condition).

Contrary to our prediction, the interference effect from distractors phonologically related to the alternative name remained stable (as did the facilitation effect from distractors phonologically related to the target name). This was also true when participants were not familiarized with the pictures, when more naming episodes and a consolidation phase were introduced, or more items per participant tested. Overall, these results indicate some limitation of incremental learning in word production.

INCREMENTAL STRATEGIES IN CHILDREN'S LANGUAGE PRODUCTION

Jessica Brough (University of Edinburgh), Holly Branigan (University of Edinburgh), Chiara Gambi (Cardiff University), Hugh Rabagliati (University of Edinburgh)
j.brough@ed.ac.uk

Adults produce language incrementally, and this incrementality can manifest in two different ways. *Lexical incrementality* refers to sentence planning on a concept-by-concept basis, with formulation and articulation beginning with the most accessible entity, which can be determined by low-level attentional cues (e.g. first seen = first mentioned). In *structural incrementality*, sentence planning begins with a general understanding of the structural relationships between entities (e.g., agent, patient, with e.g., animate entities more likely to be considered as agents, and agents more likely to be sentence subject), and that structure guides retrieval of lexical items. Eye-tracking experiments, in which speakers' gaze is analysed as they describe pictures, have provided extensive evidence for incremental production in adults (Kuchinsky et al., 2011) but little is known about online sentence production from a developmental perspective: Do children also produce language incrementally and, if so, do they show both lexical and structural incrementality, or use only one strategy?

40 adults and 48 3- and 4-year-olds (mean age = 47 mo) described events that could be interpreted from two perspectives (e.g., looking for/hiding, see Figure 1, c.f., Gleitman et al., 2007), while their eye movements were tracked. To measure the effect of animacy, scenes showed actions executed by one human and one anthropomorphic shape. To measure effects of perceptual salience, one character was always colourful while the other was in greyscale. We tested how these factors affected whose perspective participants took when describing the scene, i.e., which of the two perspective verbs they chose. This required them to choose who to mention first (as the subject) and second (as the object), e.g. *the woman is looking for the star* vs *the star is hiding from the woman*. We then assessed the relationship between what participants said, and where they gazed.

We found that both children and adults produced speech incrementally: Figure 2 shows that, at the time when both age groups mention the first character, they had already started to shift their gaze to the second character. Strikingly, the size of the eye-voice span was very similar at both ages. Next, we analysed type of incrementality. If speakers were being lexically incremental, then we expected their order of mention to reflect the first character they gazed at. If descriptions demonstrated structural incrementality, word order would reflect the linguistic bias to map the sentence subject onto the most animate entity (the human), i.e. *the woman is looking for the star*.

Adults provided evidence for lexical incrementality. The first character that they gazed at was highly likely to be used as the sentence subject ($p < .05$) and, moreover, on 27% of trials, adults began to speak before even fixating on the second character. These patterns were not present for children. They provided stronger evidence for structural incrementality, demonstrating a bias to use the most animate entity as a starting point for their sentence structures by mentioning the human first. They did this to a greater degree than adults ($p < .01$), whose animacy bias was also present but not significant. Interestingly, salience did not affect either adult or child descriptions.

The current work provides strong evidence that children produce language incrementally and can do so based on structured relations between entities. However, when describing these complex scenes, children did not provide evidence for lexical incrementality as a strategy.

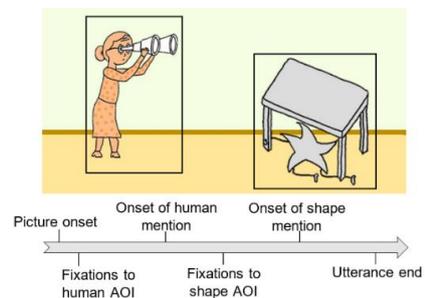


Figure 1. Top: E.g. trial for *Looking For/Hiding*; boxes indicate Areas of Interest. **Bottom:** Expected behaviour indicative of lexical incrementality.

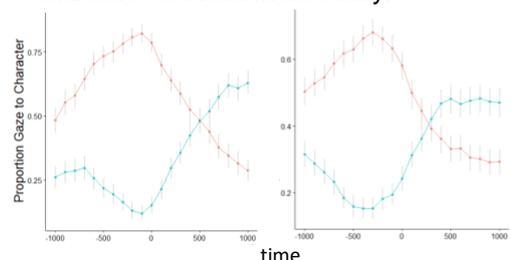


Figure 2. Gaze to agent (red) and patient (blue) as the utterance begins (L = adults, R = children; 0 ms = speech onset)

Saturday Poster.25

**LEXICAL PREDICTION DOES NOT INTERACT WITH MORPHOPHONOLOGICAL
PREDICTION DURING EARLY STAGES OF SENTENCE PROCESSING**

Mikel Santesteban (University of the Basque Country, UPV/EHU), Paolo Lorusso (IUSS Pavia), Anna Hatzidaki (National and Kapodistrian University of Athens), Adam Zawiszewski (UPV/EHU), & Itziar Laka (UPV/EHU)

mikel.santesteban@ehu.eus

During sentence comprehension we make predictions not only about upcoming words but also about their morphophonology based on both semantic and grammatical knowledge (Freuenberger & Roehm, 2016). But what is the timing of these processes and what can this tell us about information prediction prioritization? Previous research (see Martin-Loeches et al., 2006, for a review) has showed mixed evidence supporting either a *syntax-first model* (Friederici, 2002, 2011) that considers each source of information independently at initial stages and integrates it at later ones, or an *interactive model* (Hagoort, 2003) that considers all types of information right from early points of processing. Most studies have manipulated semantic violations, and only a few manipulated cloze-probabilities of nouns (e.g., Gunter et al., 2000; Wicha et al., 2004). We explore for the first time whether verb cloze-probability interacts with object-clitic agreement, and if so, at what stage(s) these processes interact.

ERPs of 23 Spanish native speakers were recorded during a RSVP reading for comprehension task of sentences (n=120) with (a) semantically high- vs. low-cloze predictable (congruent) verbs and (b) gender grammatical vs. ungrammatical object-clitic morphemes; e.g., *El conductor frenó muy bruscamente el tren para intentar pararlo/*pararla vs. aparcarlo/*aparcarla en el andén*, “The driver stopped very abruptly the train_{SG-MASC} to try to stop it_{CL-SG-MASC/*FEM} vs. park it_{CL-SG-MASC/*FEM} at the platform”). If the information sources of semantic and morphophonological prediction are first used in an independent way, we expected additive effects at early stages of processing and an interaction at later ones. Alternatively, if both sources of information are considered simultaneously, we expected interaction effects at both early and late stages (e.g., larger computational cost of morphological agreement for low than high-cloze verbs).

The results favoured the first hypothesis. At early stages, between 350-500 ms (N400), there was a main effect of grammaticality (larger negativity for ungrammatical than grammatical sentences) at left (p=.019) and midline sites (p=.015) and a marginally significant semantic predictability N400 effect at left-parietal (p=.056) and mid-parietal (p=.093) regions (larger negativity for low than high-cloze verbs), but no interaction. At later stages, between 600-800 ms (early P600), we found a grammaticality by predictability interaction in the frontal region, led by a larger positivity for the grammatical-low-cloze condition than for the other three conditions, resulting in a semantic predictability effect only in grammatical sentences (p=.020; larger positivity for low- than high-cloze verbs, *aparcarlo vs. pararlo*). This suggests that semantic integration and interpretation efforts were only made for grammatical sentences.

Taken together, our results suggest that, at early processing stages, lexical/semantic prediction does not interact with morphophonological prediction, so that the two processes operate independently (in contrast to, i.a., Wicha et al., 2004). Morphophonological and semantic processing only interact at later stages where sentence integration, interpretation and repair processes take place, supporting the syntax-first model (Friederici 2002, 2011).

Saturday Poster.26

PREDICTION OVERRIDES SYNTACTIC PRIMING: EVIDENCE FROM HINDI

Samar Husain¹ (IIT Delhi) & Himanshu Yadav² (Jawaharlal Nehru University, New Delhi)

¹samar@hss.iitd.ac.in, ²yadavhimanshu059@gmail.com

Syntactic priming is known to facilitate comprehension of the target sentence if it aligns with the structure of the prime sentence (e.g., Branigan et al., 2005; Thothathiri & Snedeker, 2008). It has also been demonstrated that syntactic prediction leads to facilitation during comprehension (e.g., Tanenhaus et al., 1995; Kamide, 2012). However, little is known about how these two processes (i.e., priming and prediction) interact. Under the assumption that both priming and prediction result in preactivation of upcoming structures, certain theories would not make a distinction between the two (e.g., Pickering & Branigan, 1998). On the other hand, some theories do assume that they are qualitatively different (see Kuperberg & Jaeger, 2016). Given that both priming and prediction are implicated in sentence comprehension, it is important to investigate their interaction. In this work we report a Cloze completion study (N=27) and a self-paced reading study (N=60) to investigate this question. All analyses were done using linear-mixed models (log RTs were used for expt 2).

In the first Cloze completion study (N=27), target items had non-finite clauses such as in (1). There were three conditions which were determined by the case of an NP (LOC, ACC, ERG). A target item was preceded by a prime that had a non-finite clause with an intransitive non-finite verb (NFV) such as in (1). NP=GENITIVE in the prime was the subject of the intransitive NFV (NFV.Intrans). Target items were shown until the Adverbial and had to be completed meaningfully.

(1) Prime: NP=GENITIVE NFV.Intrans Matrix.verb

Target: ... NP=GENITIVE NP=LOCATIVE/ACCUSATIVE/ERGATIVE ADVERBIAL ...

Syntactic priming would predict participants should complete the target with a non-finite clause and an intransitive NFV in all the three conditions. This should of course be followed by some matrix verb. If no priming happens then NP=GEN should be completed as part of a noun phrase signifying possession and no NFV should be used. While the NP=LOC can be structurally integrated with the expected primed NFV.intrans, an NP=ACC or an NP=ERG cannot be integrated. This is because an NFV.intrans in Hindi cannot take an ACC/ERG marked argument. In order to posit an NFV.intrans in the ACC/ERG condition, the NP=ACC/ERG needs to be integrated with the Matrix.verb by making a discontinuous phrase. Such discontinuous constructions are rare in Hindi. Also, NP=ACC could be integrated within the non-finite clause by positing a causative rather than an intransitive NFV.

Completion data shows that structural priming happens only in the case of LOC and ACC case conditions. Lexical priming happens only in the LOC condition. In the case of Ergative no priming happens. Interestingly, while the exact verb class of the prime (NFV.intrans) was not maintained in the ACC condition, the completed verb was nonetheless semantically related, i.e., if the prime had an NFV.intrans to laugh, the target was completed with a causative cause to laugh. This points to semantic priming in the ACC condition. The table below shows % completions in each condition.

Condition	Completion with NFV structure	Completion with NFV.intrans	Completion with NFV semantically related with prime
NP=LOC	Yes (94.8%) *	Yes (63.8%) *	Yes (22.4%) *
NP=ACC	Yes (100%) *	No (6.4%)	Yes (32.2%) *
NP=ERG	No (4.16%)	No (4.16%)	No (0%)

We followed up this with an SPR experiment (N=60). The design of the experiment was same as experiment 1, however this time the three conditions were complete, i.e., ADVERBIAL was followed by an NFV.intrans and a Matrix.verb in all conditions. Reading times were measured at the NFV.intrans region. Results show a significant effect of ACC case (t=-3.10) and ERG case (t=2.87) such that ACC > LOC and ERG > LOC at the NFV.intrans region. No difference was found between ACC and ERG conditions.

Together the two experiments suggest that prediction can override priming processes. This seems to happen when the priming sentence leads to a construction of a low probability structure in the target. At the same time, the effect of priming still persists in the form of semantically related verb. This work suggests that both priming and prediction are actively influencing sentence comprehension, but they are distinct processes. || Reference: Branigan et al. (2005) JEP:LMC; Thothathiri and Snedeker (2008) JML; Tanenhaus et al. (1995) Science; Kamide (2012) Cognition; Pickering & Branigan (1998) JML; Kuperberg and Jaeger (2016) LCN ||

Saturday Poster.27

HOW GOOD IS PREDICTION IN HEAD-FINAL LANGUAGES?

Apurva (Indian Institute of Technology Delhi) & Samar Husain (Indian Institute of Technology Delhi)
apurva@hss.iitd.ac.in, samar@iitd.ac.in

Head-final languages are assumed to be good at making predictions about upcoming elements based on prior elements. 'Anti-locality' effect (e.g., Konieczny, 2000), which is characterized by facilitation at the head with increased head-dependent distance, is a key evidence for this claim. Anti-locality is explained either as increased expectation (Levy, 2008) or as memory reactivations (Vasishth and Lewis, 2006; henceforth V&L) of the predicted phrase. However, a rigorous investigation, which checks if the facilitations due to increased distance is indeed due to robust predictions, is lacking. In this study we investigate if predictions about upcoming verbs are robust or do these falter in the face of increased working-memory. We conduct two Cloze completion studies (N=30) (cf. Levy and Keller, 2013) to probe these questions by revisiting the anti-locality effect found in Hindi by V&L (2006). All data analyses were done with linear-mixed effects model.

The first SPR experiment in V&L (2006) used double center-embedded constructions such as (1). In condition 'a' nothing intervenes between NP4=ACC and NonfiniteVerb1, while in conditions 'b', 'c', and 'd' an Adverbial, a PP and a relative clause (RC) respectively intervene. V&L (2006) found a facilitation at critical NonfiniteVerb1 in conditions 'b', 'c' and 'd' compared to condition 'a'. The facilitation was interpreted as an anti-locality effect. There were 24 sets of items in this study.

(1) NP1=ERG NP2=ACC [NP3=ACC [NP4=ACC \emptyset /Adverb/PP/RC **NonfiniteVerb1**] NonfiniteVerb2] MatrixVerb

Our Cloze completion study (N=30) presented the items from V&L (2006) until the NP4=ACC region for condition 'a' and until the intervening phrases (Adverb/PP/RC) in 'b-d'. Participants' responses were coded for the target verb type (class and category) and overall grammaticality of the completion (grammatical vs ungrammatical). Results show that predictions for the critical NonfiniteVerb1 were as follows: a=72.4%, b=78.3%, c=66%, d=64.9%. The difference between condition a vs b-d was not significant.

The second SPR experiment in V&L (2006) had a 2x2 design crossing embedded 'Relative clause (RC) type' (Subject RC vs Object RC) with 'Distance' between the relative pronoun and the RC verb (Short vs Long). They found a facilitation at the RC verb in the Long conditions compared to the Short (i.e., an anti-locality effect). We conducted a Cloze study (N=30) for these items as well. Similar to the previous Cloze task, incomplete items were presented in all the four conditions to ascertain if the critical RC verb is being predicted. Response coding was same as Expt 1. Results show that predictions for the RC verb were as follows: ORC.Short=90.9%, SRC.Short=95.9%, ORC.Long=92.4%, SRC.Long=94.1%. The two main effects and the interaction for the completion data did not reach significance.

Grammaticality analysis of Expt 1 completion data shows that NonfiniteVerb1 was predicted as part of a single embedded structure which made these completions ungrammatical. The total no. of grammatical completions for all conditions was < 7.5%. Grammaticality analysis of Expt 2 data showed that although the RC verb was frequently predicted, the matrix verb was not, which made the total no. of grammatical completions for all conditions < 42%.

The completion data suggests that while the speed-up (in V&L, 2006) at NonfiniteVerb1 (between 'b' vs 'a' in Expt 1) can be explained by the reactivation-based account, the other speed-ups ('c'/d' vs 'a' in Expt 1; Short vs Long in Expt 2) cannot be explained. This is because a key assumption of the reactivation account is that the memory chunk of the predicted verb that gets reactivated, does not change due to the intervening material. The completion data suggests that this assumption does not hold – differing intervener types (e.g. Adverbial vs PP) lead to different predictions. The expectation-based account might seem to explain the speed-up in 'b' vs 'a' in Expt 1 and ORC.Long vs. ORC.Short in Expt 2. This theory explains anti-locality as pruning of incompatible parses which are deemed to be grammatical, however given the low grammatical completions in both the experiments, it is unclear how the difference in RTs can be explained by this theory.

The two Cloze studies suggest prediction failure in Hindi in the face of increased memory load due to complex structures involving multiple (Expt 1) and single (Expt 2) embeddings. The data does not support some key assumptions regarding processing of head final languages such as (1) prediction is frequently robust, (2) prediction becomes better with more intervening material, and (3) differing intervening materials keep prediction constant. In light of this, the generalizability of anti-locality effect and its explanations by expectation-based and reactivation-based accounts become untenable.

References: Konieczny (2000) JPR; Vasishth and Lewis (2006) Language; Levy (2008) Cognition; Levy and Keller, (2013) JML; Husain et al. (2014) PloS One.

Saturday Poster.28

TRACES OF TRACES

Massimo Burattin (University of Milano-Bicocca) Francesca Foppolo (University of Milano-Bicocca), Carlo Cecchetto (University of Milano-Bicocca, CNRS & Paris VIII)
francesca.foppolo@unimib.it

The nature of the gap in filler-gap dependencies has been a highly discussed topic during the '80s. Back in those days, linguists split between a trace and a traceless approach. Some assumed that a phonologically null category (trace) sits in the gap position (Chomsky 1981 and much following work); others, mostly in the HPSG tradition (Head-driven Phrase Structure Grammar), favored a traceless SLASH mechanism instead (Pollard & Sag, 1987 a.o.). In a later revision, Chomsky (1995) proposed the *copy theory*, according to which a (silent) full lexical copy of the antecedent occupies the gap: when an element undergoes movement, it leaves an empty copy of the same category behind, not simply an empty trace (that resembles a null pronoun), as advocated by trace theory. Some experimental evidence was proposed in favor of trace approaches (Bever & McElree, 1988; Tanenhaus et al., 1985), but no consensus was reached as to interpret experimental findings. Nonetheless, no attempt has been recently made that includes the copy theory as a possible option.

We revive this old debate with novel experimental evidence for traces and their nature. We designed a lexical retrieval task in which, after hearing a given sentence, participants had to decide if a target word was present in the sentence or not. We created 48 sentence pairs (24 object and 24 subject relative clauses - RCs) in Italian in which we manipulated the distance between the filler and the gap, by inserting the same adverbial phrases in different sentential positions. The pair (a)-(b) is an example of an object RC:

(a) Mio nonno ha adottato il cane che [l'anno scorso di fronte a casa di mia zia] hanno abbandonato _ senza pietà.

(b) Mio nonno ha adottato il cane che hanno abbandonato _ [l'anno scorso senza pietà di fronte a casa di mia zia].

My grandpa adopted the dog that <pro> abandoned _ [last year mercilessly in front of my aunt's place].

For each pair, one of these four target words appeared for the retrieval task: (i) the antecedent (i.e. the filler in the filler-gap dependency, *cane*); (ii) a non-reactivated word present in the sentence and occurring close to the antecedent (*nonno*); (iii) a phonological competitor of the antecedent, which is not present in the sentence (*pane*), and (iv) a control word, that is neither present in the sentence, nor phonologically or semantically related to the antecedent (*figlia*). All conditions (sentences, dependency, target words) were balanced for type and frequency and rotated across lists. We tested 50 Italian participants.

Participants were significantly more accurate to make a decision on the antecedent compared to the non-reactivated word (88% vs. 65%: Est=-1.98; SE=0.30; z=-6.58, p<0.001) and they were always faster to make decisions on the antecedent (823 ms) compared to the other targets: vs. competitor (878 ms): t=3.84; vs. non-re-activated (954 ms): t=6.02; vs. control (881 ms): t=3.58. RTs also revealed an effect of RC type: decisions were significantly faster in object than subject RCs (762 ms vs. 936 ms). This advantage for object RCs might depend on the lower syntactic position in which the antecedent is reactivated, which is linearly closer to the decision task in object compared to subject RCs. At the same time, though, we did not find any clear effect of the phonological competitor with respect to the control word, as they behave similarly in accuracy (86% vs. 89%) and RTs (both 890 ms).

The fact that re-activation is sensitive to the structural position of the gap suggests that this effect might be mediated by the presence of a copy/trace of the filler at the gap site, contra a traceless approach. The lack of an effect (neither inhibitory nor facilitatory) on the phonological competitor, instead, does not allow for any conclusive evidence towards a copy or a trace approach.

Saturday Poster.29

IS FRENCH MASCULINE GENDER OVERRATED ?- A CLOSER LOOK AT CCA

Aixiu An, Simon Duverger & Anne Abeillé (LLF, University Paris Diderot)

aixiu0806@gmail.com

Linear order has been shown to play a role for verb agreement in English and German, with postverbal subjects favoring closest conjunct agreement (CCA) (Konieczny & Hemforth 2004, Haskell & MacDonald 2005). It has also been argued to play a role for adjectives gender agreement (Corbett 1991), but this has only been tested experimentally for the Subject-verb gender agreement in South Slavic languages (Gold et al, 2017). According to Corbett (1991) and most French grammar books, French is an exception among Romance languages in using only gender resolution (to masculine) and not CCA when Nouns with different genders are coordinated (*Le garçon et la fille sont beaux/*belles. 'the boy and the girl are beautiful.mp/*fp'*). However, we show that CCA (to feminine) is frequent with French attributive Adjectives and even preferred in certain cases. In a corpus study (131 tokens 'D Adj Nfp et Nmp' and 222 'D Nmp et Nfp Adj' randomly selected in FrWAC (Baroni et al, 2009), we found a significant effect of Adj position: 96.95% of prenominal Adj observe CCA (feminine) while only 42.34% for postnominal Adj.

We ran an acceptability rating experiment with 24 experimental items with 3 factors: Adj position, Adj gender, N animacy, and 12 control items (grammatical and ungrammatical with attraction errors, 5). The closest N was always feminine. We had 40 participants.

(1) pre,fem/masc,+hum: De nouvelles/nouveaux étudiantes et étudiants sont déjà en stage.

(2) post,fem/masc, +hum: Des étudiants et étudiantes nouvelles/nouveaux sont déjà en stage.

'Some new._{fp/mp} student._{fp} and student._{mp} are already in internship'

(3) pre,fem/masc,-hum: De nouvelles maisons et immeubles ont déjà fait l'objet de rénovation.

(4) post,fem/masc, -hum: Des maisons et immeubles nouvelles/nouveaux ont déjà fait l'objet de rénovation.

'Some new._{fp/mp} house._{fp} and building._{mp} have already been renovated'

(5) control, +/-gram: Le fils de la voisine est content/contente d'aller à l'école.

'the son._{ms} of the neighbor._{fs} is happy._{ms/fs} to go to school'

Using a maximalized linear model with rating as the dependent variable, and Position, Gender, Animacy as predictors, we found a significant effect of Gender, an interaction between Animacy and Gender, and between Gender and Position. Feminine Adj were rated higher than masculine ones, contrary to the prescriptive resolution rule, and much higher than attraction errors (all ps < .05). CCA was preferred with prenominal Adj, and dispreferred after human Nouns. We also included the acceptability of prenominal and postnominal Adj (rated by 19 participants) as a control factor, but didn't find significant effects, nor interactions.

We conclude that both syntactic (position) and semantic (animacy) factors play a role in French gender agreement. With prenominal Adj, CCA targets the highest N, whereas it targets the lowest N with postnominal A. With human Nouns, gender is not arbitrary but related to social gender. This confirms that French readers do not process masculine gender as a "default", contrary to the norm (Gygax et al. 2012). This favors a simultaneous, multi-factorial model for sentence comprehension and a single stage model (Pickering et al. 2002, Haskell & McDonald 2005) for production.

ONLY AND CLEFTS: THE INCREMENTAL PROCESSING OF PRESUPPOSITION

Jérémy Zehr (UPENN), Francesca Foppolo (UNIMIB)
 Daniele Scanzi (UNIMIB), Florian Schwarz (UPENN)
 francesca.foppolo@unimib.it

There's an ongoing debate about the backgrounded vs. entailed status of the content associated with presupposition triggers like *only* and *also*. Several studies in the past tested *only* and *also* in a variety of tasks and found contrasting results with respect to the incremental processing of presuppositional triggers (Foppolo & Marelli, 2017; Kim, 2008; Romoli, Khan, Sudo & Snedeker, 2015; Schwarz, 2015). One experimental question that remains open is the status of presupposed and asserted content in different triggers, and the timecourse of incremental processing of presupposed and asserted content.

To address this question, we implemented a Visual-World paradigm in which participants' eye-movements were monitored during the processing of sentences like (a-b) in a scenario like the one in Fig.1, in which two male and two female characters were shown next to their card. The task was to identify the target character following a comment made by a game master in a card game, in which players join a 'Chosen House' as represented by the suit of the card they were dealt. We aimed at testing the incremental processing of presupposed vs. entailed content, which is reversed in *Only* (a) and *clefts* (b) sentences. Both were compared to a control condition (c) with no presuppositional trigger:

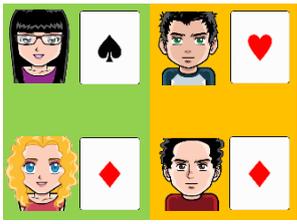
<p>(a) <i>Only Sarah</i> will join the Chosen House of diamonds Presupposition: Sarah has diamonds Assertion: no one else (except Sarah) has diamonds</p> <p>(b) <i>It is Sarah</i> who will join the Chosen House of diamonds Presupposition: exactly one person has diamonds Assertion: Sarah has diamonds</p> <p>(c) <i>Look! Sarah</i> who will join the Chosen House of diamonds Control condition: no presupposition trigger</p>	
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------

FIGURE 1

We designed the critical trials so that participants could use the exclusivity information of the trigger, along with the information about the target character's gender, to identify the target prior to the full disambiguation point constituted by the final noun (*diamonds*). Crucially, of the two characters whose gender was a match (female in this case), only the target character (bottom-left) had a unique suit (*predictable* scenario). In the control trials, both of the target-gender-matching characters had a unique suit, thus preventing early identification of the target (*unpredictable* scenario). If presuppositional information is available early on during processing, convergence onto the target character should happen sooner for clefts than for *only*, for the former but not the latter conveys exclusivity as a presupposition. We tested 46 English and 49 Italian participants on parallel versions of the task. Each participant was either assigned to the cleft or only condition (between-subject variable). Each experiment consisted of 48 experimental items, including 6 test items like (a) and (b), between subjects, 6 control items like (c) and 36 filler items controlling for potential task-specific confounds.

We found a convergence toward the target prior disambiguation in the Only condition compared to the control Look condition, as revealed by a significant effect of scenario (Est = -1.9, SE=.49, t=-3.9) and a significant interaction between trigger and scenario (Est=1.6, SE=.60, t=2.6). This result shows that listeners converged to the target earlier in the predictable than unpredictable scenario, but they did so when they heard *only*, not *look*. Contrary to our predictions, this effect was not recorded for clefts and it was only significant for English, not Italian. This result suggests that the exclusivity inference is derived incrementally to locate the target character, at least in English *only*. This process seems to be delayed in the case of clefts, in which the same content is presupposed. Further research is needed to control for possible intervening factors that might have affected results on clefts in English, as well as the absence of an early effect in Italian.

Saturday Poster.31

DO WE PROJECT SLUICING (AND BY EXTENSION ELLIPSIS) WHEREVER POSSIBLE?

Emilia Molimpakis (University College London), Masaya Yoshida (Northwestern University) & Andrea Santi (University College London)
emilia.molimpakis@ucl.ac.uk

It has been claimed that elliptical utterances are projected whenever possible [1]. Here we present compelling evidence against this claim from two self-paced reading (SPR) studies in Modern Greek and English and a sentence completion study. We argue that the original results were potentially driven by a reflexive gender mismatch effect (GMME) interacting with processing complexity.

In [1], SPR was used with two manipulations: I.) a technically compatible (1c & d) vs non-compatible (1a & b) sluice continuation at the *wh*-phrase due to the preceding P (*with*) and II.) a gender matched (1c) vs mismatched (1d) reflexive at a point compatible with ellipsis projection (4 conditions). The results showed a sign. interaction, with a GMME occurring at the reflexive only in sluice-compatible structures, leading to the claim that we have a default preference for sluicing-, and by extension ellipsis-, projection.

- 1) a) Jane's **grandfather** / b) Justin's **grandmother** told some stories at the family reunion but we couldn't remember with which story about **himself** from the party his brother was so very impressed.
- c) Jane's **grandfather** / d) Justin's **grandmother** told some stories at the family reunion but we couldn't remember which story about **himself** from the party his brother was so very impressed with.

Experiment 1 attempted to replicate these findings in Greek, a morphologically rich language. Using the same design, we tested 24 participants on 40 items (same power as [1]). We found a sign. main effect of gender mismatch ($p < .04$), but no interaction ($p > .8$, *n.s.*), with gender mismatched conditions always showing sign. slow-downs at the reflexive and immediately thereafter. We propose that the GMME in both studies may be driven entirely by an immediate search for the reflexive's antecedent, an independently well-documented phenomenon. In English, the absence of this effect in the non-sluice-compatible condition may stem from the temporary attachment ambiguity at the PP (i.e., *with* can be adverbial) that garden-paths the parser and derails it from detecting the mismatch. In Greek, rich feature marking on the PP makes its attachment unambiguous and allows the parser to notice and process the mismatch.

Experiment 2 simplified the English stimuli by excluding the reflexive (2). Null Hypothesis: if sluicing is not projected where possible, there should be no difference between the 2 conditions at the embedded subject (*Annabelle*), where a sluicing projection is foiled. Our results indeed found no sign. difference between conditions ($p > .86$, *n.s.*), confirming our hypothesis.

- 2) Andrew heard some rumours at the pub, but we did not know for sure ((a) to) which rumours Annabelle responded ((b) to) and I would like to find out somehow.

These data fail to support prediction of a sluice or even a parallel structure. The results of a sentence-completion study with shortened materials from (2) (ending at *for sure* or *for sure to*) further support this null result, showing no preference for sluice, cleft or parallel continuations in either condition. Also, more sluice completions follow the condition with the preposition than without. This indicates that the original assumption underlying the sluice-compatibility manipulation does not fit with preferences of the parser. While evidence for prediction in sentence processing is vast, this work contributes to other recent data in demonstrating its limits. While prediction is indeed useful, the creativity of language puts an upper-limit on this. Future work should consider the role of the coordinator, as "but" may generate significant creative potential.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes*, 28(3), 272-302.

THE IMPACT OF STEREOTYPES AND NOUN ENDINGS ON PROCESSING GENDER IN ENGLISH: COMPARING NATIVE AND NON-NATIVE PERFORMANCE

Julia Müller, Lars Konieczny & Verena Haser (all affiliated with
Albert-Ludwigs-Universität Freiburg)
julia.mueller@pluto.uni-freiburg.de

This study aims at establishing whether the perception of the noun ending *-er* as masculine carries over into English for native speakers of German. Since German is a language with a grammatical gender system (Hellinger & Bußmann, 2003), professions which end in *-er* identify a person as male in German (e.g. *der Gärtner* – “the (male) gardener”) while in English, they are not grammatically marked for gender. Thus, it is possible that Germans might still perceive professions ending in *-er* in English as more male. Alternatively, they might switch to an English native speaker view and judge professions based on their stereotypical associations instead, a phenomenon that has been established in various studies (e.g. Gygax, Gabriel, Sarrasin, Oakhill, & Garnham, 2009). Participants read short texts while their eye movements were recorded. 60 stimuli with three sentences each were constructed. The first sentence introduced an occupational noun which varies in its stereotypical association (with males or females) between conditions. The third and final sentence then either included “men” and “women”, referring back to the occupational noun in the first sentence. For example: “During the last month, the stockbrokers/hairdressers tried to get the business going. Recently, it had gotten a bad reputation. But two of these men/women had a brilliant idea that would turn the ship around.” Professions were taken from Misersky et al. (2014) who had native speakers of seven languages rate them for stereotypicality. Beyond that, professions ending in *-er* in English were contrasted with those that ended in *-or* and those that had endings which rarely occur in German (such as *-ian*). Additionally, 24 items from another reading experiment were used as distractors. Beyond the stereotypical associations, gender, noun endings, and native languages, the participants’ proficiency and acquisition environment of English was established. Furthermore, they filled in a short form of Bem’s Sex Role Inventory (Bem, 1981), two surveys that measure sexist attitudes towards women (the Ambivalent Sexism Inventory, Glick & Fiske, 1996), and men (the Ambivalence Toward Men Inventory, Glick & Fiske, 1999), and the Modern Sexism Scale (Swim, Aikin, Hall, and Hunter, 1995). Results from 64 participants (40 German, 24 English) indicate that *-er* may slightly, but *-or* more considerably, slow down processing for German speakers when used with *women*. Moreover, the increased exposure to incongruent stereotype-person combinations seemed to facilitate the acceptance of people in professions which do not correspond to the stereotype.

References

- Bem, S. L. (1981). *Bem Sex-Role Inventory: Professional manual*. Palo Alto, CA: Consulting Psychologists Press. Retrieved from <http://worldcatlibraries.org/wcpa/oclc/9335096>
- Glick, P., & Fiske, S. T. F. (1996). The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *Journal of Personality and Social Psychology*, 70, 491–512.
- Glick, P., & Fiske, S. T. (1999). The Ambivalence Toward Men Inventory. *Psychology of Women Quarterly*, 23(3), 519–536. <https://doi.org/10.1111/j.1471-6402.1999.tb00379.x>
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2009). Some grammatical rules are more difficult than others: The case of the generic interpretation of the masculine. *European Journal of Psychology of Education*, 14(2), 235–246.
- Hellinger, M., & Bußmann, H. (Eds.). (2003). *Gender across languages: Volume III. Impact: Studies in language and society*. Amsterdam: Benjamins.
- Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., . . . Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods*, 46(3), 841–871. <https://doi.org/10.3758/s13428-013-0409-z>
- Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68(2), 199–214. <https://doi.org/10.1037/0022-3514.68.2.199>

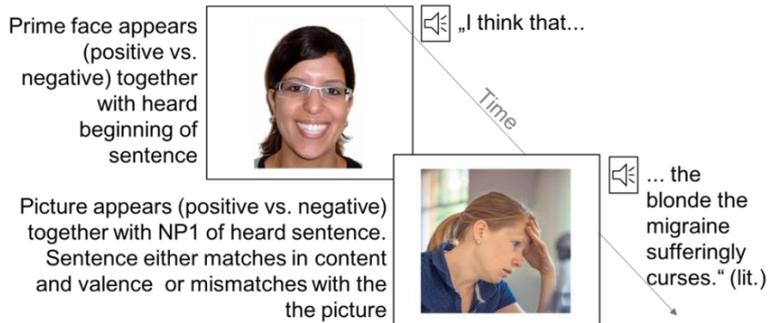
**Effects of Emotional Speaker Facial Expressions on Sentence Processing:
An ERP Study**

Katja Münster (Humboldt-Universität zu Berlin), Johanna Kissler (Bielefeld University) & Pia Knoefler (Humboldt-Universität zu Berlin, Einstein Center for Cognitive Neuroscience, Berlin School of Mind and Brain)
Katja.muenster@hu-berlin.de

Previous eye-tracking research suggests that a speaker’s emotional facial expression can rapidly modulate real-time language processing. Participants saw either an emotionally positive or negative prime face followed by two event photographs of opposite emotional valence presented side by side. Shortly after display onset, participants listened to an emotionally positive or negative German sentence related to one of these photographs. Participants preferred to inspect the photo described by the sentence more when the speaker’s prime face was negative and matched (vs. was positive and mismatched) the sentence valence. This effect emerged as soon as the first disambiguating information became available in the sentence, i.e., lit. transl. from German: *the blonde in I think that the blonde the migraine...*

Building on this research, we investigated the brain responses associated with the integration of speaker facial emotion, picture, and sentence valence. In this EEG experiment (Figure 1), young German adults ($N=25$, female=12) first inspected the face of a speaker and listened to the beginning of a spoken German sentence (e.g., “Ich denke, dass...”, lit. transl: *I think that...*). The speaker’s facial expression was either emotionally positive or negative. Subsequently participants saw an emotionally positive or negative IAPS (International Affective Picture System) picture and heard the sentence continuation (e.g., “die Blonde die Migräne leidend verflucht.”, lit. transl.: *the blonde the migraine sufferingly curses.*) whereby the adjective/adverb and NP2 conveyed the strongest emotional valence. The sentence meaning matched (vs. mismatched) the IAPS picture. After sentence end, participants performed a sentence-valence judgement task.

Figure 1: Experimental procedure exemplifying a trial in which the speaker’s positive emotional prime face mismatches in valence the picture and the sentence (the literal translation of the German sentence is presented). Note: The IAPS picture has been substituted due to the database’s user agreements.



ERP Results (descriptive): Grand average ERPs showed larger broadly distributed negative mean amplitudes for trials in which a positive speaker prime face was followed by a negative picture - negative sentence combination compared to negatively congruent speaker prime face - picture-sentence trials (mismatching speaker prime face, Figure 2). This effect emerged after the onset of the adjective/adverb (see Figure 2). ERPs were not modulated substantially by prime face valence time-locked to the onset of the first noun phrase (Figure 3).

Discussion: These effects corroborate and extend previous eye-tracking results for younger adults, specifying how a speaker’s emotional facial expression can rapidly impact real-time sentence processing. As Figure 2 suggests, this is particularly the case at the adjective/adverb region, which carries clear emotional valence in the sentence.

Figure 2: larger mean amplitudes for mismatching (red) vs. matching (black) speaker prime face trials. Negative is plotted up. 19/25 participants show this pattern (Cz, adjective/adverb onset)

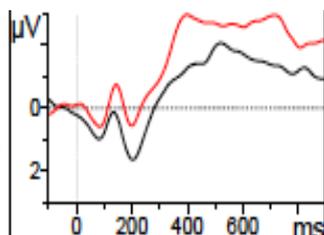
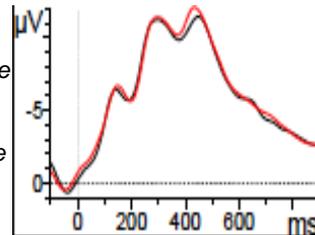


Figure 3: less ERP modulation by speaker prime face mismatch (red) vs. match (black). Negative is plotted up. (Cz, NP1 onset)



Saturday Poster.34

WHEN SWITCHING LANGUAGE IS COST-FREE

Michela Mosca (University of Potsdam), Chaya Manawamma (University of Groningen) & Kees de Bot (University of Pannonia)
mosca@uni-potsdam.de

The goal of this study is to determine to what extent language switching in bilingual speakers is a costly process. It is widely agreed that when bilinguals use one language, the other one is inhibited. In order to switch language, this inhibition needs to be overcome [1] and the new language has to be reconfigured [2]. Both processes take some time to be fulfilled, leading to the so-called "language switching costs". While inhibition decay is a passive process during which inhibition gradually dissipates, reconfiguration is an active process that dynamically prepares the system for the new task. Hence, reconfiguration can be considered the "mental gear" that allows language switch: It includes processes such as shifting the attention to the new language, retrieving the rules of that language (such as grammatical and phonological rules) and inhibiting the non-intended language.

A way to examine language reconfiguration is by informing the speaker about the language to use beforehand. Previous studies have shown that when speakers are given time to prepare for the upcoming language, switching costs are reduced [3]. However, longer preparation can foster not only the active preparation of the new language but also the passive dissipation of the inhibition from the previous task. This can happen when the interval between trials is relatively short [4] or is left uncontrolled [5]. In fact, in order to investigate the reconfiguration process only, the interval between trials needs to be relatively long so to allow complete dissipation of the inhibition coming from the previous utterance. Specifically, it has been shown that when the interval between trials is relatively long, a preparation time of 800ms allows bilinguals to switch cost-free [6]. However, while this result indicates that it is possible to eliminate switching costs, it leaves unclear how much preparation time is necessary to fully prepare for a language switch and so to switch cost-free.

To address this issue, we tested 30 native speakers of Dutch (mean age: 22 years; 6 males) with a good proficiency of English (81.5% of the LexTale; L2 mean AoA: 10.4 years) in a picture naming task involving language switching. The interval between trials was held relatively long (> 3000ms) so to allow the complete dissipation of the inhibition coming from the previous trial, whilst preparation time was manipulated by displaying the language cue *before* the stimuli (Cue to Stimulus Interval, CSI= 800ms, 500ms, and 250ms) and *together* with the stimuli (CSI= 0ms). Results revealed that language switching was costly when speakers were given no time to prepare but not when some preparation time was provided. Precisely, language switching became cost-free when preparation time was relatively long (800ms and 500ms) but also when preparation time was as short as 250ms. This finding suggests that language switching can be a cost-free process and that bilinguals require less than 250ms to fully prepare for a language switch.

References

- [1] Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67–81.
- [2] Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207–231.
- [3] Khateb, A., Shamsoum, R., & Prior, A. (2017). Modulation of language switching by cue timing: Implications for models of bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1239-1253.
- [4] Declerck, M., Koch, I., & Philipp, A. M. (2013). Digits vs. Pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition*, 15, 896–904.
- [5] Verhoef, K., Roelofs, A., & Chwilla, D. (2009). Role of inhibition in language switching: Evidence from event related brain potentials in overt picture naming. *Cognition*, 110, 84–99.
- [6] Mosca M., & Clahsen H. (2016). Examining language switching in bilinguals: The role of task demand. *Bilingualism: Language and Cognition*.

RELATIONSHIP BETWEEN THE PROCESSING OF SEMANTICALLY ANOMALOUS INTERPRETATION AND SUPPRESSION MECHANISM

Manabu Arai (Seijo University)

manabu-arai@seijo.ac.jp

An influential sentence processing theory assumes that the cost of processing input is inversely correlated with the probability that comprehenders estimated based on prior information (Levy, 2008). Our previous study, however, showed that the input evoking an utterly anomalous interpretation caused greater processing cost than the input evoking an implausible but intelligible interpretation even when both were equally unlikely to be predicted. One possibility is that in the former comprehenders may have experienced processing deadlock, finding the initially constructed analysis uninterpretable and necessary to be suppressed before seeking for an alternative analysis. In the latter, they could potentially consider the correct analysis without the initial interpretation deactivated (Christianson et al., 2001). This suggests that the ability to suppress inappropriate information, or inhibitory control, may help processing these costly sentences (cf. Gernsbacher & Faust, 1991).

We tested this possibility with Japanese sentences such as (1). The accusative NP was followed by a transitive verb in (1a) and by an intransitive verb in (1b). The intransitive verb should cause great difficulty as readers most likely predict a transitive verb following the accusative NP. Since the expected difference between (1a) and (1b) may partly be due to the facilitation caused by the prediction of a transitive verb in (1a), we also tested (1c), in which an accusative NP appeared after the intransitive verb and the RC head and readers wouldn't predict a transitive verb. Crucially, readers would be required to suppress a semantically anomalous interpretation in (1b) and we predict individuals' ability to comprehend these sentences would be linked to their ability to suppress inappropriate information.

Our participants (N=27) participated an eye-tracking reading experiment as well as a Stroop test. All analyses were conducted using Linear Mixed-Effects models with backward selection approach. The results from a Stroop task showed a robust Stroop effect ($t=7.01$). The results from the eye-tracking study showed a reliable difference between (1a) and (1b) and also between (1b) and (1c) in second pass times at the RC verb region ($t=3.32$ and $t=2.84$). The results demonstrate that readers spent longer time in rereading the intransitive verb when it appeared following the accusative NP. Similarly in the following RC head region, there was a reliable difference between (1a) and (1b) in regression path and second pass times ($t=2.05$ and $t=2.97$) and between (1b) and (1c) in regression path times ($t=4.76$).

We conducted a further analysis in which the z-scores of individuals' Stroop effects were added to the models on reading times. The analysis showed a significant interaction between Stroop Score and second pass times in the RC head region between (1a) and (1b) ($t=2.50$) as well as a marginal interaction between Stroop Score and regression path times between (1b) and (1c) ($t=1.71$). Our results suggested that comprehenders' ability to suppress inappropriate information helps them to process these costly sentences.

Examples

1a. *Seinenjitsugyoka-ga, koukana wain-o nondeita daijoyu-ni hohoenda.*

Young businessman-NOM_{RC}[expensive wine-ACC drinking] famous actress smiled at

1b. *Seinenjitsugyoka-ga, koukana wain-o aruiteita daijoyu-ni tewatashita.*

Young businessman-NOM expensive wine-ACC_{RC}[walking] famous actress-DAT handed

1c. *Seinenjitsugyoka-ga, aruiteita daijoyu-ni koukana wain-o tewatashita.*

Young businessman-NOM_{RC}[walking] famous actress-DAT expensive wine-ACC handed

References

- Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42, 368-407.
 Gernsbacher M.A., Faust M.E. (1991). The mechanism of suppression: A component of general comprehension skill. *JEP: LMC*, 17, 245-262.
 Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.

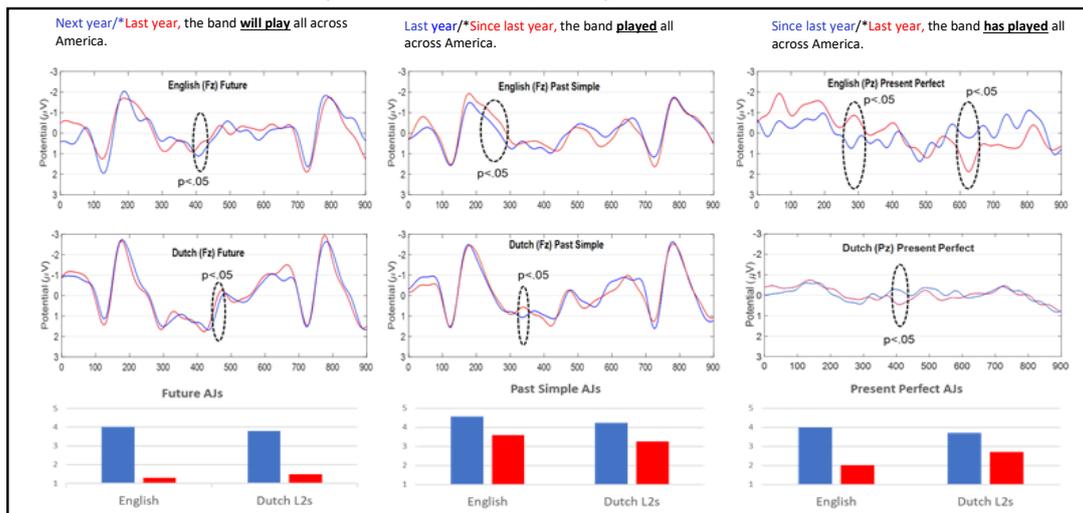
On-line sensitivity to tense and tense/aspect mismatches in L1 and L2 English

Leah Roberts, Norbert Vanek and Josje Verhagen (leah.roberts@york.ac.uk)

We investigated English natives' and Dutch L2 learners' acceptability judgments (AJs) and on-line sensitivity to tense and tense/aspect violations, created by a fronted adverbial which either matched or mismatched with a following tensed verb, focusing on the past simple, present perfect and future (e.g., Last year,/*Since last year, the band played..). The aim was twofold: firstly, ERP evidence in Dutch shows that processing a past simple verb involves discourse-referential processing (rather than local binding) supporting the Past Discourse Linking Theory (Dragoy et al. 2012), and the authors argue this should extend to the present perfect. We examine this in English. Secondly, given that Dutch has obligatory tense-marking but not grammaticalized aspect, the L2s were predicted to be native-like in their processing of the past simple and future, but not the present perfect tense (c.f., Roberts & Liszka, 2013).

Method: ERP responses were examined, time-locked to the tensed verb (*played/has played/will play*) of 20 English and 20 Dutch participants as they read 40 sentences per condition (with 120 fillers), answering yes/no comprehension questions after 50% of the items. The L2s were selected with a cloze test on tense/aspect violations (all above 70%). The items were later rated in an AJT (1=least acceptable/6=most acceptable).

Results: No differences were found between the learners' and natives' AJs: all showed a grammaticality effect, demonstrating that the L2s have explicit knowledge of tense/aspect. On-line, the L2s patterned like the NSs in the future and past simple conditions, though the effects were smaller and later. The future mismatch condition elicited a short-lived negativity similar to that found in Hindi and French (Dillon et al, 2012; Fonteneau, et al. 1998) although more frontal. The past simple condition elicited a 'referential negativity' at about 300ms, over anterior sites (c.f., Kaan et al. 2000), supporting the view that processing a past simple verb involves establishing a discourse referent (Dragony et al. 2012). In contrast, a biphasic LAN/P600 was seen for the present perfect for the native speakers, suggesting that it is treated as an agreement violation. Thus, for English speakers, processing the present perfect differs from the past simple going against the Past Discourse Linking Theory. Specifically, present perfect violations elicit more 'traditional' agreement violation responses, akin to those that have been found for present tense violations in Dutch (Baggio, 2008). As regards the L2 findings, the learners were not sensitive on-line to the present perfect mismatch condition, supporting the view that even advanced L2s may not show implicit knowledge of grammatical phenomena not instantiated in their L1 (Roberts & Liszka, 2013).



Baggio, G. (2008). Processing temporal constraint. An ERP study. *Language Learning*, 58, 35-55.
 Dragoy, O., Stowe, L., Bos, L. & Bastiaanse, R. (2012). From time to time: Processing time reference violations in Dutch. *Journal of Memory and Language*, 66.
 Dillon, B., v Nevins, A., Austin, A. & Phillips, C. (2011). Syntactic and semantic predictors of tense in Hindi: An ERP investigation. *Language and Cog. Processes*.
 Kaan, E., Harris, A., Gibson, E. and Holcomb, P. 2000. The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15: 159-201
 Fonteneau, E., Frauenfelder, U.H. & Rizzi, L. (1998). On the contribution of ERPs to the study of language comprehension. *Bulletin suisse de linguistique appliquée: Etudes en neurolinguistique*, 68, 111-124.
 Roberts, L. & Liszka, S.A. (2013). *Processing tense/aspect agreement violations in L2 English*. *Second Language Research*, 29, 413-439.

DISTINGUISHING COMPETITION FROM TASK DEMANDS: AN ERP STUDY OF NAME AGREEMENT IN TIMED PICTURE NAMING

Evangelia Balatsou, Guillaume Thierry & Gary M. Oppenheim
 Bangor University, UK
 e.balatsou@bangor.ac.uk

Pictures can be named in multiple ways. Name agreement is an empirical measure of how often people produce a picture's dominant name in picture naming norms. It correlates strongly with naming latencies (Alario et al., 2004), produces robust effects in neuroimaging (Kan & Thompson-Schill, 2004) and has been associated with both early and late effects in ERPs (Cheng et al., 2010; Shao et al., 2014; Valente et al., 2014). Such effects are almost invariably taken to reflect endogenous competition between alternative lexical entries for a given picture. However, it has recently been shown that people name pictures with strong alternatives faster than name-agreement-matched pictures with weaker alternatives, challenging the competitive selection view. Interestingly, studies manipulating picture–name agreement often include a pre-experiment familiarisation phase, requiring participants to name pictures using their norm-assessed dominant names. That may be a problem, because speakers can develop stable preferences even for non-dominant names. If speakers have previously established a preference for an alternative name, then such familiarisation creates a general task demand to override that preference. This would disproportionately affect low-agreement pictures, because fewer participants would spontaneously volunteer their dominant names, and it may thereby account for some effects attributed to endogenous competition.

In the current study, we measured naming latencies and electrophysiological activity as participants named high- and low-agreement pictures before and after familiarisation. Crucially, we used each participant's initial naming preferences to select desired names for their familiarisation: half of the pictures were familiarised using the name they initially produced, and the other half were familiarised with an alternative. We thus deconfounded name agreement from the demands inherent to externally directed name switching. As Figure 1 illustrates, the initially strong effect of name agreement on naming latencies largely dissipated after familiarisation, replaced by a name switching cost. Interestingly, it appears that weak “competitors” are just as hard to override as strong ones (Panel d).

Preliminary ERP analyses suggest that although there are some differences between high- and low-agreement pictures during initial naming, differences observed after familiarisation are largely introduced by the demand to switch names. For instance, for *both* low and high name agreement pictures in the name switching condition we find a much greater late positivity (often associated in the literature with components that reflect processes of reanalysis).

Such ERP patterns, suggest that differences previously considered as evidence for endogenous competition between alternative labels may instead reflect a form of exogenous competition, that is, the process of overriding one's naming preferences in response to explicit task demands.

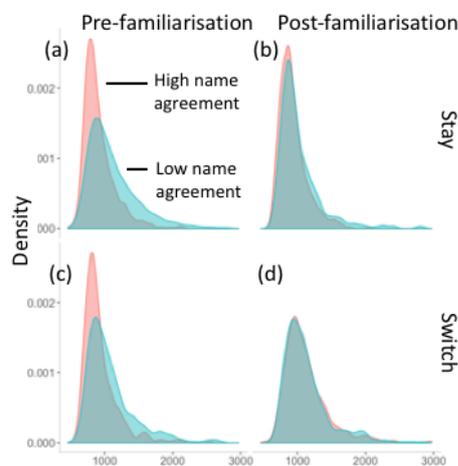


Figure 1. Picture naming latency densities.

Saturday Poster.38

IS THE TENDENCY TO LEXICALLY ENTRAIN STABLE ACROSS TIME AND INTERLOCUTORS?

Anita Tobar (University of Edinburgh), Hugh Rabagliati (University of Edinburgh), Holly Branigan (University of Edinburgh)
anita.tobar@ed.ac.uk

It is well documented that in dialogue speakers tend to use the same linguistic expressions as their interlocutor (lexical entrainment). This is specially evident when referring to objects that can be named with several lexical labels (e.g., *umbrella* vs *brolly*). Despite decades of research on lexical entrainment (e.g., Clark, 1996; Pickering and Garrod, 2004, Giles and Ogay, 2007), it is unknown whether individual speakers are consistent in their propensity to lexically entrain. Exploring this issue is crucial for evaluating the effect of transient factors, such as cognitive load, where it is imperative to ensure that, *across population*, this tendency is stable when those factors are not in play. Investigating this issue is also crucial for understanding how individual differences, such as gender, may affect speakers' tendency to lexically entrain. Expressly, in order to suggest that lexical entrainment is affected by individual differences, it is first necessary to ensure that measures are stable *within individuals*.

We investigated whether speakers' tendency to lexically entrain is stable across time and interlocutors. We conducted two experiments in which native speakers of British English engaged in two sessions of an interactive online picture-matching and -naming task. Participants alternated turns with an alleged remote player to select and name a target. Although participants believed they played with a different partner in each session, they actually played with a pre-programmed computer. Experimental items comprised a target that could be named with both a highly favoured label, e.g. *umbrella*, and a disfavoured, but highly acceptable, label, e.g., *brolly*. (To determine favoured and disfavoured labels, the materials were pre-tested in a different sample from the same population.) Importantly, the 'partner' always named the experimental targets before the participants, using the disfavoured name exclusively. Participants' subsequent use of the disfavoured word when naming the same object was interpreted as lexical entrainment; their use of the favoured, or any other, word was interpreted as non-entrainment. In Exp 1, participants played both sessions immediately consecutively. In Exp 2, participants played the second session a week after the first session. In both experiments, they played the two sessions with 'different partners'.

Overall, paired Wilcoxon tests showed that participants aligned above chance in both Exp 1 ($V=1$, $p<.0001$) and Exp 2 ($V=1$, $p<.0001$). Participants' tendency to align during the first and the second session was highly correlated not only when they played both games in a row (Exp 1: $r=.81$, $p<.0001$, 95% CI [0.63, 0.91]), but also when there was a one-week gap between sessions (Exp 2: $r=.71$, $p<.0001$, 95% CI [0.44, 0.86]). These results suggest that individual speakers' tendency to lexically entrain is stable across time and interlocutors. Confirming that this tendency is stable across individuals validates results of previous experiments exploring the effects of transient factors on lexical alignment. Moreover, confirming that the tendency to lexical entrain is consistent within individuals provides a solid basis for the study of individual differences. A possible effect of type of task on the strength of the correlations found is discussed.

AVOIDING GAPS IN ROMANCE

Francesca Foppolo (UNIMIB), Ingrid Konrad (Paris VII), Massimo Burattin (UNIMIB), Adrian Staub (UMASS), Carlo Cecchetto (CNRS, Paris VIII), Caterina Donati (Paris VII)
francesca.foppolo@unimib.it

In a recent study, Staub, Foppolo, Donati and Cecchetto (2018) provided evidence that structural principles guide the processing of filler-gap dependencies (i.e. sentences in which an element (*filler*) has moved leaving a trace in its base-generated position, the *gap*). One limitation of Staub et al.'s study was the fact that the two structures tested differed for the attachment site of the disambiguating clause (complement vs. adjunct) thus their results could be interpreted in terms of the Minimal Attachment principle (Clifton & Frazier, 1989) or consistently with the first clause of the Minimal Chain Principle (MCP, De Vincenzi, 1991), which states that filler-gap dependencies should not be posited except when necessary. To provide a further test of the first clause of the MCP, we extended this line of research by investigating two temporarily ambiguous structures in Italian and in French that only differ with respect to the presence/absence of the filler-gap dependency.

In Italian, we investigated sentences with *che* corresponding either to a *that* in a declarative complement clause (DC) or to a *what* in an indirect question (IQ), as in "Ho capito che fare gli esami è difficile (DC)/agli esami difficili (IQ-amb)" (I've understood that doing exams is difficult/what to do in difficult exams). We also tested a control sentence of the same form in which the verb unambiguously selects an IQ ("Ho chiesto che fare agli esami difficili" (I've asked what to do in difficult exams, IQ-amb). In French, we investigated a similar ambiguous construction (*à ce que*) that can correspond to a DC or a Free Relative (FR) in sentences with an ambiguous verb like "Yasmina s'habitue à ce que sa mère regarde (la télé) chaque matin" (Yasmina gets used to the fact that her mother watches TV (DC)/what her mother watches in TV every morning (FR-amb)) and compared that to a control sentence with a verb like "voir", (to see) that unambiguously selects a FR. We conducted two acceptability judgment studies with 67 adult Italian participants and 70 adult French participants, in which they had to rate (on a 7-point Likert scale for Italian and on a 10-point Likert scale for French, as it is standard in French) 24 test sentences, presented in one of the 3 conditions exemplified above. Materials also included 24 fillers that ranged from being fully grammatical to fully ungrammatical. A significant difference across conditions was found (all $p < .01$). Pairwise comparisons showed a preference for DC over IQ-amb sentences in Italian, significant by subjects and items ($t(66) = 19.19, p < .001$; $t(23) = 10.23, p < .001$) and for DC over FR-amb sentences in French, significant for subjects ($t(69) = 2.82, p = .006$; $t(22) = 0.69, p = .49$). Interestingly, a difference is shown between the ambiguous/unambiguous constructions in Italian ($t(66) = -3.21, p = 0.002$; $t(23) = -1.79, p = .087$) and French ($t(69) = -11.06, p < .001$; $t(22) = -4.64, p < .001$), providing evidence for a garden path in the case of the ambiguous verb, when a simple DC was a possible continuation. We further ran a self-paced reading experiment in Italian (N=50) using the same material and found a penalty in the disambiguating word in (b) vs. (a) (e.g. *agli* vs. *gli*), while no difference was detected on *che* nor on the word preceding the disambiguation point between (a) and (b), Figure 1. Evidence for difficulty was also detected in (c), but immediately after *che*, when a sentence with gap is the only possible continuation. We show that (i) a structure that does not involve a gap is preferred over one that does and (ii) this effect is modulated by the availability of an alternative parsing up to a certain point, and it does not solely depend on sentence complexity.

We interpret our results as evidence for structural parsing strategies of gap avoidance during sentence processing in line with the MCP.

Italian (scale 1-7) N=67		French (scale 1-10) N=70	
Condition	Mean (SD)	Condition	Mean (SD)
(a) DC	5.54 (0.91)	(a) DC	6.53 (1.68)
(b) IQ-amb	2.81 (1.16)	(b) FR-amb	5.95 (1.54)
(c) IQ-unamb	3.16 (1.21)	(c) FR-unamb	7.93 (1.28)

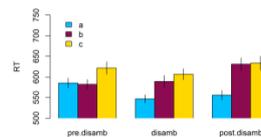


Figure 1. Self-Paced Reading

Retrieval Errors as Common Source of Misinterpretations and Repetition Errors

Michael Meng (Uni. of Applied Sciences, Merseburg) & Markus Bader (Goethe Uni. Frankfurt)

Since their discovery by Ferreira (2003), misinterpretation errors for sentences with non-canonical word order have had a major impact on models of sentence comprehension. According to the good-enough approach to sentence comprehension (Ferreira, 2003; Ferreira and Patson, 2007), misinterpretation errors result from the application of heuristics, which provide a shortcut to a full algorithmic analysis at the price of being error prone. We propose an alternative account of misinterpretation errors according to which such errors do not occur during the parsing of a sentence, but only when information has to be retrieved at a later point in time.

In two experiments we investigated German main clauses varying according to two factors: word order (subject-before-object/SO or object-before-subject/OS) and plausibility (plausible or implausible). In Experiment 1, participants first read a sentence and then had to do one of two tasks—either judging the plausibility of the sentence (plausible versus implausible) or retrieving a phrase from the sentence corresponding to a cue given after sentence presentation (“do-er” or “acted-on”). Task was a within-subject and within-sentence factor. The task required on each trial was signaled only after sentence presentation was complete. Participants could therefore not adapt their parsing strategies to the demands of the particular task. As Figure 1 shows, participants judged the plausibility of the sentences with high accuracy and without any significant effects of order or plausibility. When the actor or patient had to be retrieved in response to the respective cue, accuracy was high for plausible sentences but a substantial number of errors occurred for implausible sentences, even more so when the sentence had OS order. Since this striking task difference cannot be ascribed to differential parsing strategies, we hypothesize that the parser always computes the correct structure and meaning (resulting in high accuracy with plausibility judgments) but that the memory operations necessary for cue-based retrieval are fallible.

In Experiment 2, participants had to repeat sentences that were presented to them auditorily. After sentence presentation and before recall, participants had to add two two-digit numbers and say the result aloud. The results for Experiment 2 are shown in Figure 1. Accuracy was generally high with the exception of implausible OS sentences, which showed a substantial number of repetition errors. For most errors, the clause-initial object was repeated as subject and the clause-final subject as object, thus resulting in plausible SO sentence.

We conclude from Experiment 1 that the performance differences between plausibility judgments and agent/patient naming do not result from adapting processing strategies to task demands, thus supporting an account which ascribes misinterpretation errors to retrieval processes that contact both semantic and structural memory representations. Experiment 2 suggests that such retrieval processes are also operative in short-term sentence recall, contrary to the predictions of regeneration accounts of sentence memory.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, 47(2):164–203.

Ferreira, F. and Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass*, 1(1-2):71–83.

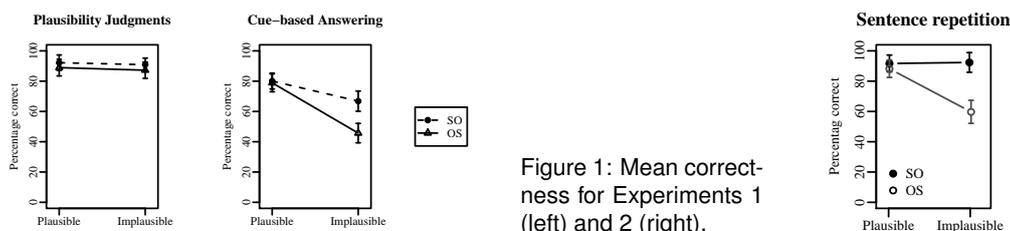


Figure 1: Mean correctness for Experiments 1 (left) and 2 (right).

PROCESSING CORRELATES OF ACTION VERB SPECIFICITY

Margit Scheibel (Heinrich-Heine-Universität Düsseldorf)

scheibel@phil-fak.uni-duesseldorf.de

Previous research on verb processing has given little attention to effects of semantic specificity. Findings from the interface make two hypotheses conceivable on how semantic specificity influences verb processing: (i) Semantic specificity on action details is an instance of semantic complexity and slows down verb processing (similar to complexity effects of complex event structures, cf. findings in [1]). (ii) Semantic specificity correlates with a multiple-coding of the verb's concept and additional image codes and/or stored action simulations speed up verb processing (cf. findings in [2], [3]). These hypotheses on verb specificity effects were tested in two experiments. We investigated processing times of specific and unspecific German action verbs in isolated presentation (Exp. 1) and in contextual embedding (Exp. 2).

Exp. 1 (n subj=27, n items=48) investigated single-word processing times of specific and unspecific action verbs, e.g., *besticken* (to embroider) vs. *verziern* (to ornament), in a lexical decision task. Verb pairs were selected such that the specific verb semantically entails the action described by its unspecific counterpart. Specific verbs specify a method by which the action is carried out, whereas unspecific verbs leave the method of the action unspecified. Verb conditions did not differ significantly in other confounding variables (e.g., word length, frequency, familiarity). The results revealed longer response latencies for specific verbs (17ms longer than unspecific verbs, main effect in a LR test of LME-models). The finding demonstrated that processing correlates of verb specificity resemble processing correlates of lexical complexity as found in [1]. Thus, we argue in the same line that specific verbs are costlier to encode due to their increased method information (vs. unspecific verbs with almost no method information).

In Exp. 2 (n subj=40, n items=24), short contexts such as (1) with specific vs. unspecific action verbs were presented for self-paced reading (word-by-word in a stationary window). Verb conditions were again matched for confounding variables such as frequency. The short contexts were identical for both verb conditions except for the critical action verb in the first sentence. The agent of the action was always named initially by a proper name and continued as agent. Comprehension questions after each context about the action mentioned in the first sentence or the pronoun resolution in the second sentence ensured reading for comprehension.

(1) *Jasper bestickt | verziert das Sofakissen. Er hat dafür ein edles Design entworfen. (Jasper is embroidering | ornamenting the scatter cushion. He has created a fancy design for it.)*

The results revealed that processing correlates of verb specificity go into reverse when verbs are embedded in sentences. We found a trend for shorter reading times for specific verbs compared to unspecific verbs (slightly shorter for low frequent verbs, about 30ms shorter for high frequent verbs, $p=.08$). The results suggested that additional non-semantic codes are part of the mental representation of specific verbs and these additional codings facilitate verb processing. We will discuss the knowledge about the agent as a reason for the facilitating effect of specificity in Exp. 2. Knowledge about the agent at the time when specific verbs are encountered enables the reader to apply the method information immediately to the agent. A densely connected agent-action representation might enhance the activation and prominence of image and sensorimotor codings of the action concept. As a result, processing of specific action verbs become facilitated as compared to processing of unspecific verbs (in contrast to their processing in isolation).

In sum, the findings provide evidence that verb specificity can cause two distinct specificity effects in on-line comprehension. Semantic specificity as such seems to be an instance of lexical complexity correlating with higher processing costs. However, when verbs are processed in contexts (introducing an event participant like the agent prior to the verb) the additional codings of specific verbs seem to make an impact and facilitate verb processing.

References: [1] Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, 89(1), B27-B41; [2] Marino, B. F., Gallese, V., Buccino, G., & Riggio, L. (2012). Language sensorimotor specificity modulates the motor system. *Cortex*, 48(7), 849-856. [3] Van Dam, W. O., Rueschemeyer, S. A., & Bekkering, H. (2010). How specifically are action verbs represented in the neural motor system: an fMRI study. *Neuroimage*, 53(4), 1318-1325.

Do cross-linguistic patterns of morpheme order reflect a cognitive bias?

Carmen Saldana & Jennifer Culbertson (University of Edinburgh)

carmen.saldana@ed.ac.uk & jennifer.culbertson@ed.ac.uk

Human languages are incredibly diverse; however, certain linguistic patterns are found frequently in the world's languages, while others are rare or even unattested. A foundational goal of linguistics is to investigate whether shared features of the human cognitive system can explain these skewed distributions of linguistic patterns. In this study we use Artificial Language Learning techniques (e.g., Culbertson & Adger, 2014) to test a hypothesized link between cognitive biases active during language learning and morpheme order cross-linguistic regularities; in particular, we investigate the tendency for number morphemes (e.g., singular or plural markers) to be ordered closer to the noun stem than case morphemes (e.g., nominative or accusative markers) (Greenberg, 1963, Universal 39). We hypothesize that this universal tendency might be driven by learners' biases towards orders that match semantic scope relationships (Bybee, 1985; Rice, 2000). We taught English-native participants (N=40) an artificial language with noun stems, and case and number morphemes. In the language, singular and nominative were both zero-marked; plural and accusative were both overtly marked. Crucially, the input language indicated only that each morphemes preceded (N=20) or followed (N=20) noun stems; examples in which the two morphemes co-occurred were held out—i.e., no instances of plural accusatives. The frequency of each morpheme in the input was identical. At test, participants were asked to produce utterances, including the held-out examples. If learners have an a priori preference for placing number closest to the noun, then they should produce Case Number Noun or Noun Number Case. Our results confirm this: learners consistently produced number morphology closer to the noun stem than case—whether the morphology followed or preceded the noun stem (post: 19/20 participants at 100%; pre: 19/20 participants at 100%). This provides evidence of a cognitive bias towards scope-isomorphic patterns which may play a causal role in this strong typological generalization.

References

- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16), 5842–5847.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (p. 73–113). Cambridge: The MIT press.
- Rice, K. (2000). *Morpheme order and semantic scope*. Cambridge: Cambridge University Press.

BILINGUAL – AND MONOLINGUAL? – LANGUAGE CONTROL

Iva Ivanova (University of Texas at El Paso)

imivanova@utep.edu

Bilinguals possess efficient control mechanisms to avoid wrong-language intrusions. But even within a single language, speakers need to control which words they say when, to avoid unwanted semantically-related or wrong-register intrusions. These control mechanisms might be similar to those used by bilinguals, or different in kind: bilinguals outperform monolinguals in some non-linguistic executive control tasks (Bialystok et al., 2004), a difference assumed to originate from bilinguals' frequent need for language control that is lacking for monolinguals.

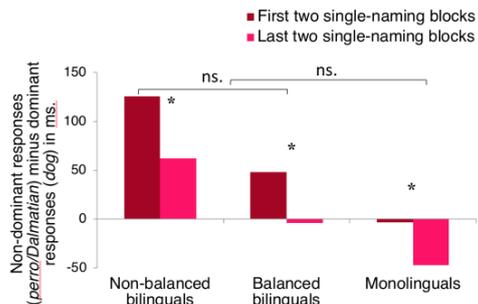
We investigated bilingual and monolingual language control mechanisms with a picture-naming task involving switching between a dominant and a non-dominant name. Two groups of Spanish-English bilinguals (non-balanced and relatively balanced) switched between naming 12 pictures in their two languages (*dog* [dominant] or *perro* [non-dominant]). To approximate the bilingual task, English monolinguals switched between naming 12 pictures with a basic-level (*dog* [dominant]) and a subordinate name (*Dalmatian* [non-dominant]). Sessions began with two single-naming blocks (dominant-only, non-dominant-only) in counterbalanced order, followed by three mixed blocks in which participants switched between languages or name types based on color cues, followed by two further single-naming blocks.

Of main interest was whether monolinguals would exhibit similar or different signatures of language control to the bilingual groups. We looked at four such signatures. In bilinguals, the first three of these reflect de-prioritizing of the dominant language (e.g., by means of greater inhibition: Green, 1998) to reduce its interference in non-dominant language production.

1. *Mixing costs* (non-switch trials in mixed blocks slower than single-block trials) – the global costs of maintaining readiness to produce two different name types in mixed blocks. In our study, all groups showed mixing costs, larger for non-dominant than dominant names for non-balanced bilinguals and monolinguals (an asymmetry of similar magnitude), and equivalent for the two name types for balanced bilinguals. 2. *Switching costs* (switch trials slower than non-switch trials in mixed blocks) – trial-level costs of switching between two different name types. All groups showed switching costs of similar magnitude, symmetrical for dominant and non-dominant names. 3. *Dominance effects*. The last two single-naming blocks were overall slower than the first two single-naming blocks, but dominant names were slowed down more than non-dominant names. Of note, this greater dominant slowing was of equivalent magnitude for all three groups (see Figure, plotting the speed difference between dominant and non-dominant names in the first two and last two single-naming blocks). A between-group difference was monolinguals' reversed dominance: Subordinate names were globally faster than basic-level names, except in the first single-naming blocks for those monolinguals who began the experiment with basic-level names. The reversed dominance was at least in part because some of our pictures more readily elicited subordinate than basic-level names. 4. *Practice effects*. In single-naming blocks, non-dominant responses sped-up with each repetition more than dominant responses, to a similar extent for all three groups (cf. larger practice benefits for low-frequency than high-frequency names, Griffin & Bock, 1998).

In sum, monolinguals showed equivalent (asymmetrical) mixing costs to non-balanced bilinguals, and equivalent (symmetrical) switching costs, equivalent dominance-effect changes and equivalent practice benefits to both bilingual groups. We note that being bilingual is different in many respects to being monolingual, and subordinate and basic-level names are in some ways different from translation equivalents. Still, our results suggest that control mechanisms (possibly involving inhibition) might be similar when selecting between highly-activated dominant and non-dominant names across two languages and within a single language.

Dominance Effect Over Time



**THE INFLUENCE OF THE ANIMACY OF DIRECT OBJECTS
IN BRAZILIAN PORTUGUESE**

João Vieira (Universidade Federal do Ceará), Brenda Sousa (Universidade Federal do Ceará) & Elisângela Teixeira (Universidade Federal do Ceará)
elisteixeira@letras.ufc.br

Introduction. It is known that the animacy status of nouns directly affects the costs of language comprehension. Animated direct objects always yield longer fixations (Czypionka, 2013) when observed in sentences with dative and accusative verbs, and tend to present higher N400 potentials (Paczynski & Kuperberg, 2011) in sentences that have both subject and objects animate. These studies tend to compare objects which are animate or inanimate. Also, Paczynski and Kuperberg (2011) found results suggesting that the linear position, if the animate noun appears as a subject or object, is more influential in facilitating processing costs of the animacy trait than other aspects, such as Thematic Roles.

Materials & Methods. Still, experiments that control the verbal restriction for the animacy are necessary to clarify and disentangle this matter. We conducted an eye-tracking experiment to investigate in Brazilian Portuguese (BP) if there were effects of animacy for verbs with and without restriction for the thematic roles of their objects. We constructed 48 experimental items, 12 per condition. The experiment had a total of 28 participants (6 males, age average of 24,85). We selected 12 verbs with restriction for animate objects, 12 verbs with restriction for inanimate objects and 12 verbs without restrictions, which allow both animate and inanimate objects, as in this table below:

Restriction & Animacy	Sentences				
With [+Anim]	O músico	[alegrou]	[o garçom]	[que andou]	exaustivamente.
With [+Anim]	The musician	[cheered]	[the waiter]	[who walked]	exhaustively.
With [-Anim]	O músico	[editou]	[o ofício]	[que molhou]	na chuva.
With [-Anim]	The musician	[edited]	[the document]	[that wet]	in the rain.
Without [+Anim]	O músico	[empurrou]	[o garçom]	[que gritou]	subitamente.
Without [+Anim]	The musician	[pushed]	[the waiter]	[who cried]	suddenly.
Without [-Anim]	O músico	[empurrou]	[a cadeira]	[que quebrou]	bruscamente.
Without [-Anim]	The musician	[pushed]	[the chair]	[that broke]	abruptly.

Results & Conclusion. We explored many different dependent variables on the main regions (verb, object, and spill over regions - in BP it was always a relative clause), but we only find main effects on the durations of total fixations (first and second pass plus go-past measures), as well as the total number of fixations. We did not find interaction between verb' restriction and animacy. In the comparison between pairs, the experimental conditions that indicated the main effect in the total fixation time ($F = 2,752$, $p = 0.04$) and number of fixations ($F = 3,9$, $p = 0.03$) were those with verbal restriction for the [+ animated] trait and without verbal restriction accompanied by objects with a [+ animated] trait on the relative clause, after the direct object. According to our interpretation of the results, the conditions without verbal restriction for direct object animacy apparently potentiate the influence of animacy, since the parser doesn't know beforehand the animacy of the following object, increasing the processing costs.

References

Czypionka, A. (2013) The interplay of object animacy and verb class in representation building. PhD dissertation, Berlin.
Paczynski, M. & Kuperberg, G. (2011) Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Lang Cogn Process*, 26(9), 1402-1456.

RIGHT-LATERALIZATION OF VERBAL COLLOCATIONS

Shohini Bhattasali, Murielle Fabre & John Hale (Cornell University)
sb2295@cornell.edu

Human sentence comprehension involves both memory retrieval and structural composition. Frequent collocations like *break the ice*, *boa constrictor*, *safe and sound*, *see to it*, *in spite of* can help us address the neural bases of these processes. These collocations, known as Multiword Expressions (MWEs) in computational linguistics, are considered a single unit, rather than a structurally composed combination (Sag et al., 2002). Thus, these word sequences are likely to be stored as a unit, rather than built-up compositionally, where their syntactic or semantic properties as a whole cannot be derived from their parts.

This study investigates whether different types of MWEs evoke different patterns of activation in the brain using fMRI. Specifically, we investigate if in these kinds of lexicalized expressions, the presence of verbal elements is observable at the cerebral level. By comparing verbal MWEs and non-verbal MWEs featuring no argumental structure, we ask if establishing verb-argument selectional relations is determining any difference in the processing of these expressions.

Participants (n=51) listened to *The Little Prince's* audiobook for 1h38min. Participants' comprehension was confirmed through multiple-choice questions. MWEs were identified using a statistical tagger trained on Children's Book Test dataset. Presence/absence of verbal expression yielded two categories of MWEs (56% verbal vs. 44% non-verbal). To account for sentence-level compositional processes, we included a regressor formalizing syntactic structure building. Contrasts were inclusively masked with the main effect of all MWEs and FWE voxel-corrected (T-score > 5.32).

Presence of MWEs elicited activation mainly in bilateral Supramarginal Gyrus, right Angular Gyrus, right MFG, and right Precuneus Cortex (Fig1A). Verbal MWEs appear right-lateralized compared to non-verbal ones in IPL and in IFG triangularis (Fig.1B). The opposite contrast yielded a mostly right-lateralized and wider pattern of activation, including bilateral Supramarginal Gyrus extending to STG and right SMA together with smaller activation clusters in Pars

Opercularis and MTG.

These findings suggest firstly that these non-compositional expressions elicit a surprisingly right-lateralized network. Secondly, we observe that although these word sequences are lexicalized, the presence of verbal argument structure appears to be processed in different brain areas.

Our findings corroborate that bilateral Supramarginal Gyrus is sensitive to co-occurrence frequency of word combinations (Graves et al., 2010; Price et al. 2015).

Significant clusters for verbal and non-verbal MWEs illustrate spatially distinct patterns of activation and a dorso-ventral gradient is observed in Broca's area. Finally, activation patterns for verbal-MWEs indicate that verb-argument selectional relations involve right hemisphere activity in Broca's area and IPL.

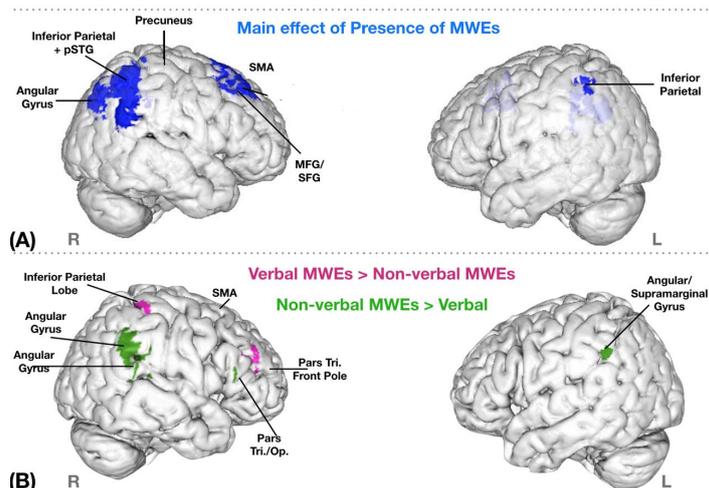


Figure 3 (A): Whole-brain main effect for the presence of MWEs in blue. (B): Contrast images with significant clusters for [Verbal MWEs > Non-verbal MWEs] in purple. (C): [Non-verbal MWEs > Verbal] contrast in green. Brain maps are FWE-corrected for multiple comparisons at voxel-level, $p < 0.05$.

Saturday Poster.46

IT'S TIME TO PRIME TIME! STRUCTURAL PRIMING SHOWS INTERRELATION BETWEEN VIEWPOINT ASPECT AND EVENT STRUCTURE

Monique Flecken and Johannes Gerwien
(monique.flecken@mpi.nl)

In the literature on syntactic priming, a longstanding open question concerns the exact interrelation between semantic-conceptual information, as encoded in a preverbal message, and the build-up of syntactic structure (Bock 1986, Dell & Ferreira, 2016). Specifically, in how far do conceptual specifications in the message affect syntactic encoding? How much conceptual information is preserved at the syntactic level? (Branigan & Pickering, 2018). On the one hand, there are studies that show a relatively strong impact of different types of conceptual information on the generation of syntax, e.g., conceptual accessibility, animacy, thematic role structure and order, discourse status of referents (Bock & Warren, 1985, Pappert & Pechmann, 2014, Griffin & Weinstein-Tull, 2003, Prat-Sala & Branigan, 2000). On the other hand, there is work showing that the impact of conceptual specifications on syntactic encoding is fairly minimal (e.g. Bock & Loebell, 1990; Huang et al., 2016).

This question has not been resolved to date, because, first of all, the types of structures used to study syntactic priming is fairly limited, and, second of all, the structures used almost exclusively involve the mapping of referents or event participants onto syntactic structure (e.g., active/passive; PO-DO alternations). Importantly, besides information on referents (and their assigned thematic roles), a preverbal message also contains other conceptual features; it contains, for example, a specific temporal viewpoint, which is typically conveyed by *aspect*. In the present study we investigate the priming of grammatical aspect in Dutch, a language in which progressive and non-progressive verbs can be used (to a large extent) interchangeably when describing ongoing events. Importantly, the Dutch progressive/non-progressive alternation is reflected in different syntactic structures and a different ordering of the elements in the verb phrase: *wij zijn [een artikel] aan het schrijven* vs. *wij schrijven [een artikel]* (~ 'We are an article at the write' (progressive) vs. 'We write an article' (non-progressive). Progressive aspect in Dutch thus allows an investigation of the role of conceptual representations in structure repetition.

We have shown previously that an *aan het* aspectual viewpoint can be primed in native speakers, but not second language learners of Dutch speaking a non-aspectual first language (German). This was taken as (some) evidence for the role of conceptual representations during aspect priming (authors, 2015). In the present study, we explore this issue in more detail. Using the classic sentence-picture recognition paradigm (Bock, 1986), we tested a large sample of participants (N=90 Dutch native speakers). Four priming conditions were included: 1) neutral prime (*Peter typt een smsje* 'Peter types a text message'); 2) transitive progressive prime (*Peter is een smsje aan het typen* 'Peter is typing a text message'); 3) intransitive progressive prime (*Peter is aan het gapen* 'Peter is yawning'); 4) form prime (*Peter werkt aan het artikel* 'Peter works on the paper'). All target pictures (N=30) showed an animate agent acting upon an inanimate patient-object (e.g. man knitting a scarf, woman doing dishes).

Results show a reliable priming effect for the *transitive progressive prime* condition only: Relative to the neutral prime, only in the transitive prime condition more (transitive) *aan het* constructions were produced. The absence of an effect in the form prime condition shows that superficial form features (the PP *aan het*) alone are not sufficient to elicit structure repetition. More importantly, the absence of a priming effect in the *intransitive progressive* condition suggests that viewpoint aspect and information on event structure (number and type of thematic roles) form a bundle of features, represented together tightly. Thus, viewpoint aspect, as a conceptual feature, cannot be primed in isolation. We provide an in-depth discussion of the implications of these findings for current views on syntactic processing in language production.

Saturday Poster.47

ANTI-LOCALITY EFFECT WITHOUT VERB-FINAL DEPENDENCIES

Mingya Liu (Osnabrück University) & Ming Xiang (University of Chicago)

liu.mingya@uni-osnabrueck.de

Anti-locality effects ([2]; [4]) provide strong evidence for expectation-based sentence parsing models [3]. Previous discussion of the anti-locality effect, however, largely focused on the argument-verb dependencies in verb-final constructions. It is argued in [6] that memory retrieval based account is equally adequate in accounting for the anti-locality effect in verb-final constructions. In the current study, we investigate whether anti-locality effect is present in constructions that do not involve verb-final dependencies.

Our study compares two different determiners in German: the morphologically complex determiner *derjenige* ‘the-jenig’ and the bare determiner *der* ‘the’. The complex determiner *derjenige* obligatorily requires a relative clause (RC), whereas the bare determiner doesn’t trigger such expectations ([1]; [5]). A self-paced reading experiment was conducted based on a 2x2 design: either a bare or a complex determiner (e.g. *den* vs. *denjenigen*) preceded the object NP; the RC that modifies the object NP is located either right after the object NP or at a distant position (see an example in (1)). There were a total of 24 experiment items, plus 48 fillers. After each trial participants rated the naturalness of the sentence (1-7 scale). The critical region (CW) for the SPR data analysis is the relative pronominal phrase *dessen Mutter*. Anti-locality would predict longer RT on the CW for the complex-distant (1d) compared to the complex-local condition (1c). No anti-locality effect was predicted for the bare determiner (1b vs. 1a), since the bare determiner doesn’t trigger a prediction for a RC.

- (1) a. *Maria Richter hat den Mitarbeiter, [dessen Mutter] ein großes Haus in Spanien besitzt, in einem Cafe gesehen.* **bare-local**
 (M. R. has the colleague, whose mother a big house in Spain owns, in a café seen)
 b. *Maria Richter hat den Mitarbeiter in einem Cafe gesehen, [dessen Mutter] ein großes Haus in Spanien besitzt.* **bare-distant**
 (M. R. has the colleague in a café seen, whose mother a big house in Spain owns)
 c. *Maria Richter hat denjenigen Mitarbeiter, [dessen Mutter] ein großes Haus in Spanien besitzt, in einem Cafe gesehen.* **complex-local**
 d. *Maria Richter hat denjenigen Mitarbeiter in einem Cafe gesehen, [dessen Mutter] ein großes Haus in Spanien besitzt.* **complex-distant**

Results and Discussion (subj N=54): All analyses were done with mixed effects models. Rating results found that bare *der*-DPs with a distant RC (1b) were rated less natural than with a local RC (1a) ($p < .01$), possibly suggesting a locality effect; in contrast, complex *derjenige*-DPs with local or distant RCs received the same ratings (Determiner x Locality interaction $p < .01$).

For the RT results, there is an anti-locality effect on the CW (*dessen Mutter*) under only the complex determiner *denjenigen*: the RT was shorter when the RC is more distant ($p < .01$). There was no such effect for the bare determiner conditions (Locality x Determiner interaction $p < .05$). However, for the complex determiner, a strong anti-locality effect appeared also before the CW: RTs on the phrase “*gesehen/seen*” in (1d) is shorter than on “*denjenigen Mitarbeiter/the colleague*” in (1c) ($p < .01$). No such difference appeared for the bare determiner conditions. On the one hand, this effect on the pre-RC region may simply be due to the word length and the low frequency of the complex determiner *denjenigen*, introducing a confound and therefore undermining any conclusion we can draw on the CW. On the other hand, it is also possible that since the comma on the pre-RC region serves as a strong cue to signal the upcoming RC, any anti-locality effect on the relative pronominal phrase may shift to this earlier position. We are currently conducting corpus studies to examine these possibilities.

References: [1] Alexiadou et al (2000); [2] Konieczny (2000); [3] Levy (2008); [4] Levy & Keller (2013); [5] Roehrs (2006); [6] Vasisht & Lewis (2006)

ACQUISITION OF RESULTATIVE EVENT REPRESENTATIONS IN DUTCH: DOES DESCRIBING EVENTS AID MEMORY OF EVENT CULMINATION?

Ciara Hobbelink, Miguel Santin & Angeliek van Hout (University of Groningen)
c.hobbelink@student.rug.nl

Thinking-for-speaking effects arise when the use of language to encode information focuses the attention of speakers on particular features of the world (Slobin 1996). These effects have been found for event representation in adults across languages. However, the acquisition question has hardly been raised (except for e.g. Bunger et al. 2016). If language can indeed influence cognitive representations, how many years of active language use does it take for this effect to emerge in learners? Our study addresses this question for the domain of resultative events (i.e. events progressing towards the achievement of a goal or result) in Dutch children. As a so-called *satellite-framed language* (Talmy 2000), Dutch typically encodes actions in simple verbs (*roeren* ‘stir; *puzzelen* ‘do a puzzle’), while the result of events are encoded in particle verbs (verb + satellite; e.g. *door-knippen* ‘cut through’; *vol-schenken* ‘pour full’). Does the use of such particle verbs to describe event results affect their representation in adults’ and children’s memory?

Dutch children (n=45, mean 4;6) and adults (n=48) watched short video clips of resultative events progressing towards a culmination point (peeling tangerine) and non-resultative events (stirring in a pan). The actions either came to an end (ceased) or were still in progress at video offset (ongoing). In a between-subjects design, participants in the Verbal Experiment first described what happened in the videos, while those in the Non-Verbal Experiment first performed a visual attention task. Subsequently, participants performed the same memory task: judging whether screenshots correctly showed the event endings (ceased/ongoing) they had just watched (half of the items were a match, the other half a mismatch).

A binomial mixed-effects regression in the Verbal experiment showed that ceased-resultative events were recognized more accurately when participants had first described the events ($p < 0.01$). However, this effect was modulated by age: the model showed a negative effect in the recognition of ceased-resultative events for children ($p < 0.05$). Recognition accuracy in the Non-verbal experiment was equally poor in adults and children. Linguistic analyses of the event descriptions revealed that adults, but not children, produced particle verbs mostly for describing ceased-resultative events (i.e. events that reached a culmination point).

The results show *thinking for speaking* effects on the representation of events in memory of adults. The absence of any pattern in the use of particle verbs by children suggests that Dutch 4-year-olds have not yet developed specific language to encode event results. Possibly, this is why they did not show any *thinking-for-speaking* effects.



Figure 1: Verbal structure types produced by Dutch adults and children in the verbal experiment.

Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: implications for language production and attention. *Language, Cognition and Neuroscience*, 31(8), 1015–1037. <https://doi.org/10.1080/23273798.2016.1175649>

Slobin, D. (1996). From “Thought and Language” to “Thinking for Speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Towards a Cognitive Semantics: Typology and Process in Concept Structuring* (Vol. 2). Massachusetts: MIT Press. <https://doi.org/10.1353/lan.2002.0176>

DISCOURSE INERT?

IMPLICIT OBJECTS CAN BE AS ACCESSIBLE AND AS PERSISTENT AS OVERT ONES

Ana Besserman & Elsi Kaiser (University of Southern California)

pianibes@usc.edu

Optionally transitive verbs such as ‘read’, ‘eat’ and ‘play’ may appear with (i) an *overt* direct object (“Lisa was reading a book”) or (ii) an *implicit* object (“Lisa was reading \emptyset ”). These **Implicit Arguments** (henceforth **IAs**, denoted with \emptyset) are considered **discourse inert** (Koenig & Mauner’99, Williams’05), partially because they are not felicitous antecedents for pronouns (“Lisa finished reading \emptyset ; and left ~~it~~, on the table”). But are IAs really completely inert in the discourse? We take a closer look at IAs’ discourse-level behavior by testing **(i) IA accessibility**, i.e., under which conditions anaphoric expressions can felicitously refer back to these implicit objects, as well as **(ii) IA persistence**, i.e., how likely they are to be mentioned again in subsequent discourse.

To do this, we use **sluicing**, constructions in which an entire clause is elided except for a wh- phrase/remnant (“Lisa was reading {a book / \emptyset } but I’m not sure which book ~~Lisa was reading~~”). This wh- phrase (which book) can have either an overt antecedent (*a book*) or an implicit one (\emptyset) in the prior clause. We use sluices as a tool to probe IAs’ discourse status because, unlike pronouns, sluices exhibit a preference for local, i.e. *object* antecedents (Frazier & Clifton’98, Harris’15) and are thus well-suited for investigating the accessibility of implicit *objects*.

To explore (i) **IA’s accessibility** we look at whether the acceptability of implicit objects as antecedents is modulated by the **form** of the retrieval cue, the wh-phrase. Less accessible antecedents are generally referred to with *more informative* expressions and vice-versa (e.g., Givon’83, Ariel’90, Gundel et al.’93). Thus, **we compare sluices with informative wh-phrases (“which book”) to uninformative ones (“which one”) to assess the accessibility of IA antecedents**. In **Study 1**, we manipulated (i) object type (overt/implicit) and (ii) wh-phrase (which noun/one). 45 native English speakers rated 20 items/30 fillers on a 5-point scale. **Results:** Two main effects and an interaction (p ’s<.05, lmer on z-scores): With uninformative remnants (*which one*), overt objects are more acceptable than implicit objects (p <.001). With informative remnants (*which book*), implicit vs. overt object conditions *do not differ* (p >.08): implicit objects are as accessible antecedents as overt objects when the wh-phrase is informative. In **Study 2** we also manipulated **context set complexity** (see Hofmeister’11) to test whether representational complexity ameliorates IAs’ acceptability as antecedents in discourse-linked constructions, but found no significant effects of context set complexity.

Based on these findings, **Study 3** tested **(ii) IA’s discourse persistence**. Prior work suggests that comprehenders use information about a referent (e.g. whether it was introduced by a definite or indefinite noun phrase) to make assumptions about whether this referent will be mentioned again (Gernsbacher et al’89, Brocher et al.’16). Comprehenders may be less likely to expect re-mention of object if it was introduced implicitly rather than overtly. Thus, in Study 3, we manipulated (i) object type (overt/implicit) as well as (ii) QUD (question under discussion, manipulated by means of a preceding overt question/no question). The question manipulation aims to boost objects’ salience with *explicit* discourse goals. 32 native English speakers answered a forced-choice completion questionnaire in which they decided whether a sluice’s wh- phrase referred back to a subject (always overt) or an object (overt or implicit). **Results:** No sig. diff. between conditions; object responses were the majority (~65%) for all conditions, **confirming the previously established preference for object correlates in sluicing even when the object is implicit**.

Summary: Despite prior claims that implicit objects are discourse inert, we find that they can be **as accessible as overt objects when the retrieval phrase is informative**, and **as discourse persistent as overt objects**: Comprehenders were not any less likely to think an implicit object will be re-mentioned relative to an overt one. Under some conditions, such as sluicing, IAs can participate in discourse relations despite their apparent ‘inertness’.

PROCESSING QUANTITY IMPLICATURES UNDER QUDS

Eszter Ronai & Ming Xiang (The University of Chicago)

ronai@uchicago.edu

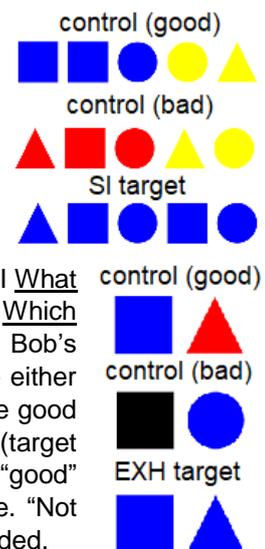
Background: Implicatures are an important testing ground for examining the time-course of integrating semantic and pragmatic information. Earlier discussion on implicature processing has focused on whether semantic information is privileged as the listener reasons about intended meaning, predicting implicature calculation to be costly (Bott & Noveck, 2004 a.o.), or whether implicature calculation is effortless and default (Grodner, et al., 2010). However, instead of making a categorical distinction between “costly” vs. “cost free”, a constraint-based approach views implicature calculation and processing as resulting from the interaction of multiple cues and constraints (Degen & Tanenhaus, 2015). A potential cue is Question Under Discussion (QUD), which, manipulated via explicit questions (Zondervan, et al., 2008), a background story (Degen, 2013) or focus prosody (Cummins & Rohde, 2015), has been shown to affect the rate of implicature calculation. We go beyond previous work by conducting an elicitation study establishing likely QUDs. We present novel evidence that QUD affects not only calculation rates, but also the reaction time (RT) cost of two types of quantity implicature.

Experiments: The two quantity implicatures investigated were scalar inference (SI: *Some of the shapes are blue* +> *Not all of the shapes are blue*) and it-cleft exhaustivity (EXH: *It is the square that is blue* +> *Only the square is blue*). In **Exp.1**, participants saw SI and EXH sentences paired with pictures, and were asked to provide QUDs. They were given a background story: Anne is asking questions from Bob, about pictures that only Bob can see. The task was to guess Anne’s questions, given Bob’s answer and the picture.

Exp.2 used the most frequent elicited QUDs from Exp.1 in a sentence-picture verification task. Participants saw dialogues between Anne and Bob, and had to decide whether Bob gave a good answer to Anne’s question, given the picture he saw. The QUD manipulation included: SI What color are the shapes?/Are any shapes blue?/Are all shapes blue?, EXH Which shape is blue?/Are there any blue shapes?/Are both shapes blue?. Bob’s answers were the SI and EXH target sentences. The sentences were either unambiguously good/bad descriptions of the (control) pictures, or were good descriptions on their literal, but not on their inference-enriched reading (target pictures) - see images for SI/EXH. Thus if a participant says Bob gave a “good” description of a target picture, she has not calculated the implicature. “Not good” indicates implicature calculation. Responses and RT were recorded.

Results: For SI, any QUDs resulted in fewer implicatures calculated than all ($p < 0.001$) or what ($p < 0.05$). For EXH, both QUDs resulted in more implicatures calculated than any ($p < 0.001$) or which ($p < 0.001$). Therefore we can see that any and which are Literal-biasing, while all, what and both are Implicature-biasing QUDs. These differences in question type predict speed of processing. In SI, we find a significant interaction of Question and Response ($p < 0.01$) and in EXH, a significant interaction of Question and Condition ($p < 0.001$), such that with Literal-biasing QUDs, making an SI/EXH-enriched judgement takes longer than responding to the relevant literal control. With Implicature-biasing QUDs, there is no such difference. In other words, QUDs that bias against implicature derivation make that derivation incur a reaction time cost. Contrarily, under QUDs that bias towards implicature derivation, there is no cost.

Conclusion: We present novel data showing that QUDs modulate implicature calculation rates and crucially also affect processing cost. This challenges previous discussions of there being a uniform cost (or lack of cost) for deriving scalar or quantity implicatures. Instead, our findings are most compatible with a constraint-based account of implicature, and language processing more generally, where QUD is one of many cues.



Saturday Poster.51

CROSS-VARIETAL LEXICAL ALIGNMENT

Marie-Anne Morand (University of Bern) & Constanze Vorweg (University of Bern)

constanze.vorweg@csls.unibe.ch

Interlocutors tend to converge on a wide range of linguistic levels and features, including word choices. Speakers have been found to repeat previously heard or read lexical choices not only when different labels reflect different conceptualizations [1], but also when there is a choice between roughly synonymous words [2]. To inquire whether lexical alignment can also occur when the source of priming is an utterance in a different language variety, we investigated lexical priming in bi-varietal speakers of Swiss German (dialect) and Standard German. The linguistic situation in German-speaking Switzerland can be described as *diglossia*, i.e., a juxtaposition of two varieties (instead of a continuum, cf. [3]), with both varieties being used in different situations, each in their own right. Swiss German (SwG, umbrella term for all of the dialects) is used for most spoken communication, Standard German (StG) is used for most written language, in education, in many national TV programs, and with non-dialect speakers. Accordingly, both varieties are clearly distinct, speakers use either one or the other, without intermediate forms. They are deliberately kept apart.

To explore whether there is lexical priming between StG and SwG, we used a lexical alignment paradigm [2] in which participants alternate between selecting a picture that matches a given description and naming a picture themselves. Forty students participated in the experiment. In one condition, the prime descriptions (matching turn) were given in SwG (n=20); in one condition, they were given in StG (n=20). Participants always named pictures (naming turns) in SwG. On critical trials, a prime picture that has two acceptable names (e.g., 'flat iron', SwG: *Büguyse* vs. *Glettyse*, StG: *Bügeleisen* vs. *Glätteisen*) was named with one of them. After two intervening filler trials, the participant named the same picture (now target). Based on several pretests (naming and acceptability rating), 16 critical pictures were selected, which appeared twice each (once as prime, one a target). Altogether, there were 480 filler trials (240 matching turns, 240 naming turns).

In both the within-variety and the cross-variety conditions, participants tended to repeat the lexical choice that they had heard before in the matching trial. It should be noted that in the cross-variety condition, a lexical repetition may involve a phonetic difference (as in the examples given above). Moreover, priming ratios (Fleiss' Kappa) did not differ significantly between conditions (SwG: $M=.69$, $SD=.10$; StG: $M=.64$, $SD=.17$; $t(38)=1.20$, $p=.239$, $d=.36$).

The results demonstrate strong lexical alignment between Swiss German and Standard German, even though the default preferences may be very different for both varieties. They are compatible with the idea put forward by [4] that cognate words have shared lexical representations across SwG and StG. Lexical alignment across language varieties provides a mechanism for both interactive adjustment and long-term contact effects. Furthermore, the presented results suggest that language production models need to take into account lexical choices that speakers have in particular naming situations. Lexical variants might be the result of geographic mobility and language contact, coupled with pluricentricity and regional variation, or of the (phonologically adapted) inclusion of dialect words in the standard language, or vice versa. We propose that there is a mechanism that allows the language production system to choose between lexical alternatives.

References

- [1] Brennan & Clark (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 1482–1493.
- [2] Branigan, Pickering, Pearson, McLean & Brown (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121, 41–57.
- [3] Ammon et al. (2016). *Variante Wörterbuch des Deutschen*. De Gruyter.
- [4] Lüthi, Vorweg, Pickering & Branigan (2016). *The morphosyntactic representation of language varieties: Bivarietal syntactic priming*. Poster presented at CUNY 2016.

First Language Processing of Compounds in Late Bilinguals
Serkan Uygun (Potsdam University) & Ayşe Gürel (Boğaziçi University)
serkanuygun2014@gmail.com

It has been suggested that native speakers develop different L1 representations as they become L2 users. Accordingly, L2 users' L1 knowledge may diverge from that of monolinguals in various linguistic domains. Previous research also suggests that L1 problems bilinguals experience may be linked with difficulties in accessing/integrating grammatical information in comprehension and production.

The present study investigates potential changes in late bilinguals' processing of L1 compounds, as this will be revealing for the representation of morphologically complex forms in the bilingual mental lexicon. Two main hypotheses pertaining to this issue assume either decomposition or full-listing in word recognition. In between these models, there are also hybrid (dual-route) models that predict both decomposition and full-listing in complex word processing depending on various factors such as familiarity, transparency, and frequency. Previous cross-linguistic work on compounds reveals semantic transparency and headedness as important factors influencing the processing pattern.

In a masked priming experiment, monolinguals ($n=73$) and high-proficiency Turkish-English adult bilinguals residing in Turkey ($n=34$) were tested on Turkish noun-noun compounds, which are mostly right-headed. The stimuli consisted of 10 transparent-transparent, 'kuzeydoğu' (northeast) (*kuzey*=north, *doğu*=east); 10 partially transparent, 'büyükelçi' (ambassador) (*büyük*=big, *elçi*=delegate), 10 pseudocompounds ('fesleğen', 'basil', *fes*=fez; *leğen*=bowl/pelvis), and 60 monomorphemic words, 'kaplumbağa' (turtle), together with 90 nonwords. The prime-target pairs were presented in three conditions: (i) Constituent 1 (*kuzey*-KUZEYDOĞU), (ii) Constituent 2 (*doğu*-KUZEYDOĞU), and (iii) Unrelated (*çanta* 'bag'- KUZEYDOĞU). All items were matched on length and frequency.

A 2 x 3 x 2 Mixed ANOVA for the RTs revealed a significant main effect of word type ($F=66.731$; $p<.001$), prime type ($F=11.114$; $p<.001$), and the interaction of word type and prime type ($F=7.014$; $p<.002$). Both groups processed compound words significantly more slowly than noncompounds ($p<.001$). A further analysis of compounds indicated significant differences between constituent 1 and unrelated prime ($p<.001$) and constituent 2 (head) and unrelated prime ($p<.001$), suggesting constituency-independent decomposition for both groups. No priming effects were observed for noncompounds. In order to investigate the effect of semantic transparency in compound processing, another 2 x 3 x 2 Mixed ANOVA was conducted and the results showed a significant main effect of prime type ($F=10.863$; $p<.001$) and the interaction among word type, prime type, and group ($F=5.219$; $p<.007$). A further analysis revealed significant differences between constituent 1 and unrelated prime ($p<.001$) and constituent 2 (head) and unrelated prime ($p<.001$). A pairwise comparison of the significant interaction indicated that, in monolinguals, while both constituent 1 ($p<.001$) and constituent 2 ($p<.001$) triggered significantly faster RTs than the unrelated prime in partially transparent compounds, it was only constituent 2 (the head) ($p<.001$) that revealed the same result in transparent-transparent compounds. As for the bilingual group, no priming effects were observed in partially transparent compounds, but a significant difference was found between constituent 1 and unrelated prime ($p<.008$) in transparent-transparent compounds.

Overall, the results reveal that while Turkish monolinguals employ decomposition regardless of semantic transparency, the bilinguals employ a dual-route as they display a semantic transparency-based decomposition in processing compounds. These findings suggest that qualitative and quantitative changes may occur in L1 morphological processing of late but highly proficient L2 users residing in the L1 country.

Saturday Poster.53

HOW MUCH DOES VERB SEMANTICS DETERMINE VERB SYNTAX?

Mariela V Jennings, Boston College
Martha Palmer, University of Colorado Boulder
Joshua K Hartshorne, Boston College

Verbs vary in terms of which syntactic frames they can appear in. In principle, this could be an unpredictable fact about the verb that must be acquired, much like the phonological form of the verb. However, many theorists posit that there is a systematic relationship between the semantics of a verb and the syntactic frames in which it can appear (Levin and Hovav, 2005). This Semantic Consistency Hypothesis has played a central role in many theories of language acquisition and processing (e.g., Gleitman, 1990).

Despite its theoretical importance, tests of the Semantic Consistency Hypothesis are scattered. This is a function of scale: with thousands of verbs, hundreds of frames, and dozens of purportedly determinative semantic features, it is difficult to study more than a tiny fraction. While researchers have identified at least 280 verb classes in English (groups of verbs that appear in all and only the same syntactic frames) (Kipper et al., 2008), only a few have been systematically investigated.

We address the issue of scale through crowd-sourcing, recruiting large numbers of volunteers, each of whom may provide only a few annotations. Ultimately, we hope to study 8,000 verbs across 280 verb classes. To date, we have focused on 83 verb classes covering 2,167 verbs and 10,976 verb/frame combinations. For each semantic feature and for each verb/frame combination, we continue collecting data until there is unambiguous agreement across annotators, as determined by a modified entropy criterion (explained in the presentation). Analysis focuses on these “finished” items (see Table).

We selected six semantic features from those most commonly invoked in prior theoretical work. We also included one semantic feature (EVALUATION – roughly, whether an event is good or bad) that has not been invoked, which provides a control. Each task was extensively piloted to ensure annotators understood the instructions.

At time of analysis, we have obtained 568,940 annotations from 11,260 annotators, sufficient to analyze an average of 5,159 verb/frame combinations per semantic feature (see Table). We operationalized “semantic consistency” as the percentage of verbs in a class that have the same annotation. For instance, if every verb in a class involves the verb’s subject undergoing a physical change, that class would have 100% consistency for the PHYSICAL CHANGE feature. Because a verb’s semantics often depends on the syntactic frame it is in (Goldberg, 1995; Levin & Rappaport Hovav, 2005), we calculate consistency separately for each syntactic frame and then average.

Semantic consistency, averaged across verb classes, was near ceiling for our six critical semantic features (see Table). This strongly supports the Semantic Consistency hypothesis. Interestingly, semantic consistency was much lower for the control feature (EVALUATION), lending support to the claim that semantic consistency is specific to some “core” set of semantic features (Jackendoff, 1990; Pinker, 1989). In the presentation, we consider how the specific findings (which verbs involve which features) compare to prior theoretical work.

Despite being the most comprehensive investigation of the Semantic Consistency Hypothesis to date, it should be noted that our study focuses on only seven semantic features for only a fraction of verbs in a single language. These results may not generalize. We are continuing to collect data and hope to present results for more verbs and features in the future.

Task	Semantic Feature	Anns.	Anns./Item	Items Done	Consistency
Entropy	PHYSICAL CHANGE	83,568	8	56%	96%
Equilibrium	APPLICATION OF FORCE	92,799	9	39%	94%
Explode on Contact	PHYSICAL CONTACT	76,149	8	44%	98%
Fickle Folk	CHANGE OF MENTAL STATE	60,813	8	44%	99%
Philosophical Zombie Hunter	MENTAL STATE	92,030	9	27%	96%
Simon Says Freeze	LOCATION CHANGE	75,185	8	48%	97%
A Good World	EVALUATION	88,396	8	71%	76%

Table 1: Task, semantic feature, number of annotations, mean annotations/item, percentage of items fully annotated, mean consistency.

PUSHKIN: AN OPEN-SOURCE ENGINE FOR SOCIAL SCIENCE AT SCALE

Mariela V Jennings, Joshua K Hartshorne

Boston College

Half the world’s population has Internet access. In principle, researchers are no longer limited to the subjects they can recruit into the laboratory. Any study that can be run on a computer or mobile device can be run with nearly any demographic anywhere in the world and in large numbers. This has allowed scientists to effectively run hundreds of experiments at once, studying in unprecedented detail how mind and behavior vary across age (Germine et al., 2011; Halberda et al., 2012; Hartshorne & Germine, 2015; Maylor & Logie, 2010) and culture (Bleidorn et al., 2013; Hauser et al., 2007; Reinecke & Gajos, 2014), and to test theories with 10,000s of stimuli (Hartshorne et al., 2014; Keuleers et al., 2015).

Despite their transformative power, such studies remain rare for practical reasons. They require creating sophisticated new software to implement them. Although there are an increasing number of software solutions for online labor market studies (e.g., studies using Amazon Mechanical Turk), such software does not address the unique challenges of the massive online experiments described above. For this reason, laboratories that conduct such studies usually use custom-built software. Even more challenging is designing new research paradigms that take full advantage of the unique opportunities of the Internet, such as active learning and automated experimental design (Fedorov, 2010; Lindley, 1956; Settles, 2012), and experience sampling (Killingsworth & Gilbert, 2010). Finally, subject recruitment is challenging, as the standard methods (payment or course credit) are not available.

We present Pushkin, a free and open-source platform for designing and conducting massive experiments over the Internet. Pushkin allows a wide range of behavioral paradigms through integration with the intuitive and flexible jsPsych experiment engine (de Leeuw, 2015). Crucially, it addresses basic technical challenges associated with massive, worldwide studies, such as scaling and subject recruitment.

Pushkin facilitates research paradigms that can capitalize on the Internet’s massive scale, including active learning, automated experimental design, and citizen science. It allows dynamic stimulus choice (backed by real-time machine learning), and provides functionality such as social media integration, tracking individuals across sessions, and interactive forums. Crucially, Pushkin supports entry-level users by providing ready-to-use templates, while providing customizability and extensibility to advanced users.

In the presentation, we describe Pushkin’s functionality, provide an overview of usage, and address common concerns.

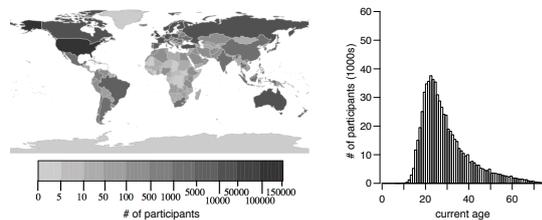


Figure 1: Country of residence (left) and ages (right) of nearly 700,000 English-speaking subjects in one massive online experiment.

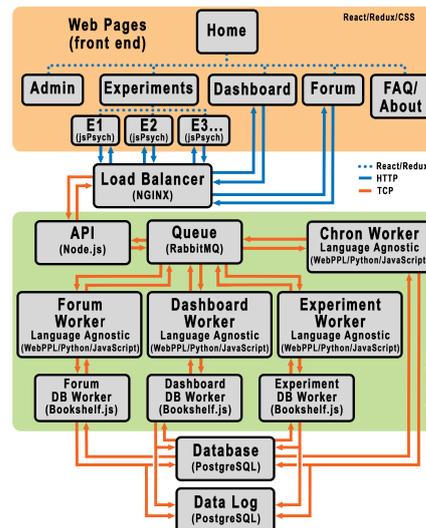


Figure 2: Schematic of an massive online experiment website implemented with Pushkin.

THE CONSTRUCTION THAT THE READER NEVER LEARNS: ORCS AND ADAPTATION

Caroline Andrews, Brian Dillon and Adrian Staub (UMass Amherst)
ceandrews@linguist.umass.edu

Syntactic adaptation is a proposed mechanism that allows comprehenders to learn to process once-difficult constructions by adjusting how expected a structure is in context, *i.e.*, the comprehension correlate of abstract priming [1,2]. Evidence for adaptation comes from decreasing reading times (RTs) for difficult MainVerb/RelVerb garden paths as a function of the number of RelV-structures participants had seen [1,2]. One of the advantages of syntactic adaptation as proposed is that it is a *general* learning mechanism for all syntactic structure; however it is currently unknown to what extent adaptation generalizes across syntactic constructions. We find that object relative clauses (ORCs) display limited evidence of adaptation in comprehension, or priming in production. Our results contrast with [3], the previous study on adaptation of ORCs. However, [3], like [1], used self-paced reading, which is known to be susceptible to task adaptation [4] and is therefore problematic when the effect of interest is also adaptation [5]. Additionally, [3] only compared pre- & post-test reading times (RTs) — not the incremental adaptation predicted by [1].

The current study tests for incremental adaptation for ORCs using eyetracking-while-reading (n=72). Subjects were shown 32 ORCs (25% of all stimuli). We measured RTs on the RC subject as a function of the number of ORCs seen, and compared this to two baseline conditions: SRCs (1b) and a matched Complement clause (1c). To increase the statistical shift in favor of ORCs, subjects were shown only 8 SRCs and Comp sentences, which also acted as baselines for any task adaptation. Thus, any syntactic adaptation would manifest as a SENTENCETYPEXORDER interaction when ORCs are compared to the baseline conditions.

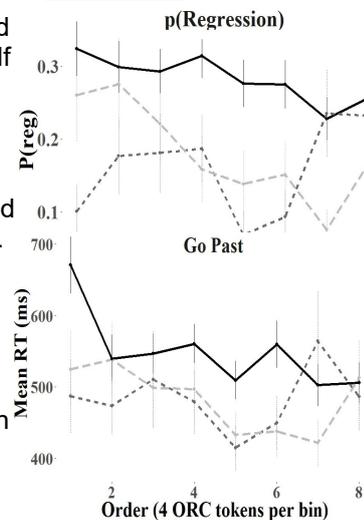
- (1) a. (n=32) ORCs: The biologist that/ the botanist/ consulted/ presented ...
- b. (n=8) SRCs: The biologist that/ consulted/ the botanist/ presented ...
- c. (n=8) Complement clause: The biologist/ believed that/ the botanist/ consulted...

Additionally, a sentence completion task was given before and after eyetracking, using the first region of (1a/b) as a prompt. If adaptation/priming is possible for ORCs, we expect an increase in ORCs in the post-test. As a control, 32 PO datives were included in the eyetracking and 28 dative preambles were also included in the sentence completion task.

Results: Pretest sentence completion indicates that ORCs had a pre-experiment surprisal of 8.38 bits (c.f. S(RC)=6.97 in [1]). Consistent with [6], we observed an ORC penalty at the relative NP. There was a main effect of ORDER for ORCs in Go Past times (*LME* $t = -3.08$). Yet, the critical interaction of SENTENCETYPEXORDER did not obtain significance in any measure when ORCs were compared to either SRCs or Comp sentences (Go-past: OvSRC $t = -1.33$; ORCvComp $t = -1.38$). In addition, sentence completion failed to show any evidence of increased ORC production between the pre- and post-tests ($p = .26$ in poisson regression). Critically, this cannot be a failure of the experiment to produce priming, because we did find priming of PO datives in the post-test: marginally for PO-only completions ($p = .06$) and significantly across PO+PO-like frames (e.g., locatives) which participate in dative priming ($p < .005$) [7].

These results suggest that neither the adaptation in [1] nor even traditional production priming extend easily to ORCs, contra predictions of a syntactic adaptation approach to priming.

[1] Fine et al., 2013. *PloSOne*. [2] Tooley & Traxler. 2018. *JML*. [3] Wells et al. 2009. *Cognitive Psych*. [4] Witzel et al. 2012. *JPR*. [5] Atkinson, 2018. *CUNY*. [6] Staub et al. 2016. *Cognitive Science*. [6] Bock and Loebell, 1990. *Cognition*.



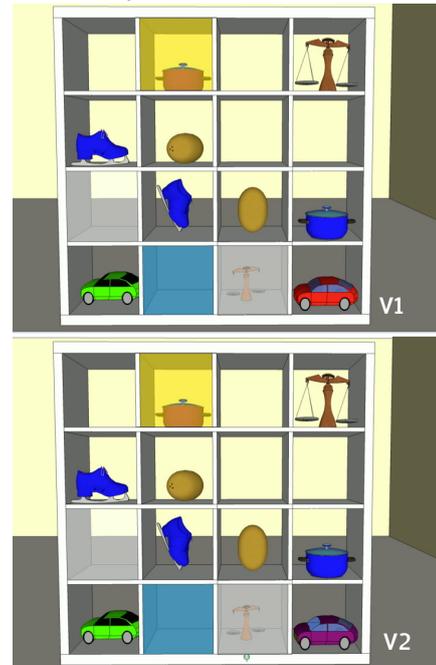
	ORC	SRC	Comp
pre:	0.3	20.5	25.5
post:	2.0	24.7	34.5
Sentence Completion %			

PERCEPTUAL UNCERTAINTY EFFECTS AND REFERENTIAL CONTRAST

Adriana Baltaretu, Craig Chambers (University of Toronto)
 adriana.baltaretu@utoronto.ca

Past studies have shown real-time referential processing is sensitive to speaker perspective and reflects expectations for lexical consistency. In these studies, the descriptions listeners encounter normally contain accurate information about entities. However, in real-world contexts, descriptions can reflect a tension between technically accurate information vs.

information that is inaccurate in-the-moment yet is simple to encode due to entrainment or perceptual accessibility. E.g., during unpacking, a person might refer to a plain empty box as "the dishes box" even though the contents had already been emptied. In the current study, we use the domain of colour to explore how listeners manage these trade-offs in their referential expectations. In our task, a confederate provided instructions ("Click on the...") for a participant-listener seated beside her. Instructions related to a virtual shelf unit containing 10 objects (two objects from five categories), presented on a large 40" display. Although speaker descriptions were scripted, no participants guessed she was a confederate. For each display, the speaker gave a sequence of instructions. On critical trials, these included reference to a particular object using a disambiguating modifier ("the [ADJ] pot") either 1 time (low entrainment: **LE**) or 3 times (high entrainment: **HE**). Crucially, this object had either a transparent panel mounted on the front of its compartment, or a translucent coloured one. The coloured panel changed this entrained object's (**EO**) actual color (e.g., a purple pot that appeared as such from the back of the display or through a transparent panel appeared red through a yellow panel--as shown in the figure--although this tinting effect was not *a priori* obvious). Speaker's descriptions reflected the **EO's** "apparent" colour in the coloured filter condition ("red pot") or its actual colour in the transparent condition ("purple pot"). After the entrainment rounds, the shelf unit rotated 180°, revealing the actual color of any objects behind colored panels, followed by a rotation back to the original position. A final instruction then occurred, where the test object (e.g., a car) belonged to a different category than the **EO**. In the first version of the expt. (**V1**, $N=20$), the test object and its description ("red car": top panel) had the same color as the **EO's** notional colour when the **EO** was behind a coloured filter. In this case we found significantly more early looks to the **EO** with the colored panel ($M=0.67$) upon hearing the test object's description, than with the transparent panel ($M=0.48$, all results tested with LME models). This suggests knowledge of the competitor's actual color had surprisingly little effect. Also, the effect was not modulated by the degree of entrainment for the colour-filtered **EO**. In the second version (**V2**, $N=19$), the test object ("purple car") was instead the same colour as the **EO's** (newly learned) actual colour. Thus comparable consideration of the **EO** across the two filter types would occur only if listeners' are using their "updated" colour knowledge. Results instead showed reliably fewer looks to the **EO** competitor in the colored filter condition ($M=0.49$) than in the transparent condition ($M=0.64$)--again reflecting listeners' apparent disregard for newly acquired information about objects' true nature--as well as insensitivity to the degree of entrainment. In sum, the findings show that, while computing referential contrast, listeners appear to strongly prioritize perceptual information available "in-the-moment" (superficial appearance) over conceptual knowledge about objects' true nature. Moreover, interactions with entrainment were not observed, suggesting the effect of past descriptions on listeners' expectations emerges only in contexts of perceptual certainty.



Saturday Poster.57

LEARNING NOVEL MORPHOSYNTACTIC FEATURES DURING VISUAL ACTION-EVENTS: EYE-TRACKING

Yang Gao, Dato Abashidze, Pavel Trofimovich & Kim McDonough
Concordia University, Montreal, Canada
dato.abashidze@concordia.ca

Recent years have witnessed increased interest in identifying how various types of exposure and individual learner profiles contribute to novel morphosyntactic pattern learning, typically using either text or static images during learning (e.g., Andringa & Curcic, 2015; Brooks, Kwoka, & Kempe, 2016; Indrarathne & Kormos, 2017) or comparing videotaped instructions with those containing no videos (e.g., Secules, Herron, & Tomasello, 1991). What has been less examined is how dynamic visual events illustrating the meaning of novel utterances can facilitate novel grammar learning. The current study extends this prior research by examining novel pattern learning from dynamic visual displays, focusing on learners' eye gaze behavior during learning.

English-speaking university students ($N = 36$), all speakers of Germanic or Romance languages with no prior knowledge of Georgian, carried out a construction learning task followed by tests. The target patterns involved completed (*bich-**ma** kocn-**ul** gogoit*, "boy kissed girl") versus ongoing (*bich-**su** kocn-**ar** gogoit*, "boy is kissing girl") events in Georgian, with affixes on the noun (subject) and the verb (-ma and -ul for completed, -su and -ar for ongoing). Participants first learned six Georgian nouns and three verbs by associating them with relevant toy characters (e.g., girl) and actions (e.g., kiss). They then watched 36 videos (about 5.5 seconds each) depicting completed and ongoing actions performed by a person holding a toy character in each hand. Participants' eye movements were recorded as they listened to an utterance describing the event, while viewing the image of the last video frame. In tests, participants heard 24 sentences featuring correct and incorrect combinations of target morphemes and selected the corresponding image of a completed or ongoing event.

Results showed that, in tests, participants tended to rely on the morphological marking on the first noun as well as the verb. They responded more accurately to sentences describing ongoing actions than those describing completed events, which was accompanied by longer gaze fixations to images of ongoing than completed events. Eye gaze patterns further revealed how learning progressed. Early in the utterance, participants tended to look at the subject character for completed actions more than at the subject character in ongoing actions. However, after verb onset, they inspected the object character more than the subject character, and in the last word region more looks went to the object character for ongoing actions than to the object character for completed actions. This early preference for the object character suggests participants' reliance on the morphological marking in the beginning stage of learning. Implications of findings for morphosyntactic learning are discussed.

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition.

Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning.

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study.

Secules, T. Herron, C., & Tomasello, M. (1992). The effect of video context on foreign language learning.

THE ROLE OF VARIABILITY IN LINGUISTIC GENERALIZATION: EVIDENCE FROM A COMPUTERIZED LANGUAGE TRAINING GAME WITH 7-YEAR-OLDS

Elizabeth Wonnacott, Masa Vujovic and Chantal Miller
(masa.vujovic.15@ucl.ac.uk)

Successful language learning involves acquiring abstract structures that operate across particular linguistic items. *Statistical learning* approaches suggest that such generalization is driven by the linguistic input, with the prediction that generalization should be more likely when learners are exposed to more varied instances, since this allows linguistic structures to be disassociated from trained instances (Bybee 1995, *LCP*; Ramscar et al. 2010, *Cognitive Science*; Wonnacott et al. 2012, *JML*). The current work explores this hypothesis in a language training experiment with child learners.

Methods: Forty 7-8 year-olds played a computerized game exposing them to instructions in an unfamiliar language (Japanese). All exposure sentences involved: (i) one of the case markers *no ue ni* (above) or *no shita ni* (below) (ii) the object marker *o* (iii) the verb *oku* (put). Nouns were all English cognates (avoiding lengthy vocabulary training). Sentences were pre-recorded by a native speaker of Japanese. Children heard the sentences whilst viewing a grid (Figure 1), and attempted to follow the instruction by moving the pictures. Importantly, incorrect responses were followed by a demonstration of the correct response (though no explicit teaching was provided); this process was repeated until the child made the correct response for that trial. Following training (over two 20-30 minute sessions), children were tested on their comprehension of untrained sentences containing untrained nouns (more English cognates) in order to assess generalization. The test mirrored training although without feedback.

Critically, there were two between-participant training conditions: (1) high-variability exposure: *28-above* and *28-below* sentences, each encountered once; (2) low variability exposure: *2-above* and *2-below* sentences, each repeated 14 times; (number of nouns (8) and total exposure matched across conditions). We predicted high-variability exposure would lead to greater generalization due to better dissociation of the structures from trained instances.

Results: Both participant groups improved through training, although children in the low-variability condition showed overall strongest performance during training (Cohen’s $d=.84$), presumably due to repeated testing on identical items (see also Hsu & Bishop, 2014, *PeerJ*). Critically, however, children in the high-variability condition showed stronger performance in the generalization test (Cohen’s $d=.64$). This supports the hypothesis that exemplar variability plays a key role in driving linguistic generalization.



Figure 1. Screenshot of a training trial. Participant may hear:
banana o chokorēto no ue ni oku
the banana OBJ the chocolate ABOVE put
meaning: “Put the banana above the chocolate”.

MORPHOLOGICAL REGULARITY AND PROCESSING DIFFICULTY IN AN FMRI STUDY ON RUSSIAN

Natalia Slioussar (HSE, Moscow, & SPBU), Maxim Kireev (IHB, RAS, & SPBU), Alexander Korotkov (IHB, RAS), & Svyatoslav Medvedev (IHB, RAS)
slioussar@gmail.com

We aim to find out if there are qualitative differences in the processing of morphologically regular and irregular forms in the brain, and, if yes, what the nature of these differences is. Previous studies explored mostly English and German, where the regular / irregular distinction is clear-cut (e.g. Beretta et al., 2003; Desai et al., 2006; Jaeger et al., 1996; Marslen-Wilson & Tyler, 1998; Sach et al., 2004; Sahin et al., 2006; Ullman et al., 1997). In Russian, verbs cannot be simply divided into regular and irregular — there are many inflectional classes that differ in frequency, productivity etc., which allows exploring this question.

A recent fMRI study (Slioussar et al., 2014; Kireev et al., 2015) took two groups of real and nonce Russian verbs: the most frequent productive AJ class, which was observed to behave as the default class in previous behavioral studies (Gor & Chernigovskaya 2001, 2003), and the least frequent non-productive classes, which can be called the most regular and the most irregular, and demonstrated that morphological regularity and processing complexity effects can be teased apart. Activity of the left inferior frontal gyrus (LIFG) was greater for production of irregular verbs (compared to regular ones) and real verbs (compared to nonce ones), so this pattern was explained by processing difficulty (Slioussar et al., 2014). But the functional connectivity of the LIFG with temporal lobe was relatively increased when regular verbs were produced (compared to irregular ones). Nothing similar was found for the real/nonce distinction, so this was concluded to be a genuine regularity effect (Kireev et al., 2015).

However, the two verb groups used in this study differed by many properties (type frequency, productivity, defaultness), so it was impossible to find out which one was responsible for the observed effect. We conducted a new fMRI study adding a third verb class — the I class (very frequent, but less frequent than the AJ class, productive, but not default). We also wanted to find out whether the observed effect would be replicated in comprehension.

The only published fMRI study where regular vs. irregular verbs were compared in comprehension (Stamatakis et al., 2005; Tyler et al., 2005) used English language and an ingenious design (an auditory same-different judgment task) that drew participants' attention to the morphological features of stimuli. This design could not be replicated in Russian, so we came up with the following one. In every trial, participants first saw a real or nonce verb in the infinitive form and a pronoun *ja* 'I' or *on* 'he' below it (600 ms). After an interval, two present tense forms of the previously shown verb appeared on the left and on the right of the screen (1500 ms). One of them agreed with the pronoun, the other did not. Participants were asked to select the correct form. We preferred this design to showing only one form (agreeing or not) because this task would involve agreement violations, and would focus on error detection.

Participants were 24 native Russian speakers. We analyzed BOLD signal changes associated both with the 1st stimulus (an infinitive and a pronoun) and the 2nd stimulus (two present tense forms). Based on these analyses we selected ROIs for the analysis of psychophysiological interactions (PPI) revealing changes in functional connectivity. All non-trivial results were associated with the 2nd stimulus. We showed that the effects of interest are the same in production and in comprehension. The subtractive analysis showed that the activity of the LIFG gradually increased from the AJ class to I class and then to irregular verbs. The effect was analogous to the processing difficulty effect from (Slioussar et al., 2014).

The PPI analysis revealed a connectivity pattern that was very similar to the ones reported in (Kireev et al., 2015) and (Stamatakis et al., 2005). The fact that it was found in the two languages with relatively poor and relatively rich inflectional morphology and in the studies using three very different tasks proves that this effect is reliable. Moreover, we found out whether it can be associated with type frequency, productivity or defaultness. The latter was true: an increase in functional connectivity of the LIFG was observed for the AJ class (as opposed to the I class and irregular verbs). Notably, this can be explained only in the dual route approach to inflectional morphology (e.g. Pinker 1991) postulating a categorical distinction between the default class and the other classes. In the other approaches, regularity effects, if present at all, are expected to correlate with type frequency and productivity.

The study was supported by the grant 16-18-00041 from the Russian Science Foundation.

Saturday Poster.60

**PHONEMIC PREDICTION AT THE ROOT FOR
ARABIC WORDS WITH PREFIXES AND INFIXES**

Samantha Wray (New York University Abu Dhabi)
Alec Marantz (New York University, New York University Abu Dhabi)
samantha.wray@nyu.edu

How unexpected a phoneme is, given the previous phonemic string (formalized as *surprisal*) is sensitive to morphological structure; hence phonemic surprisal is heightened at morphemic boundaries (Ettinger et al. 2014). For words of Arabic, which exhibit classic nonconcatenative morphology in which a root morpheme C-C-C is interleaved with vowels to produce a wordform CVCVC, phoneme surprisal at the level of the root morpheme is a more reliable correlate of neural activity than surprisal at the linear word level including every phoneme of the word (Gwilliams & Marantz 2015).

The current study focuses on a dialect of spoken Arabic (Emirati) and demonstrates that root surprisal is a correlate of neural activity even when a consonantal prefix precedes the root (nCVCVC), or when a consonantal infix intervenes between two root consonants (CtVCVC).

Participants (n=13) performed a lexical decision task on aurally presented words (190 target items, plus filler and nonwords) that differed significantly in their root and linear surprisal values word-medially (the second consonant of the root and third consonant of the word). Magneto-encephalography was recorded concurrently and auditory evoked responses were investigated for significant clusters of activity in time that correlated with root and linear surprisal values within the primary auditory cortex. Clusters were determined by performing a mass-univariate ANOVA with root surprisal and linear surprisal as factors predicting average distributed source activity.

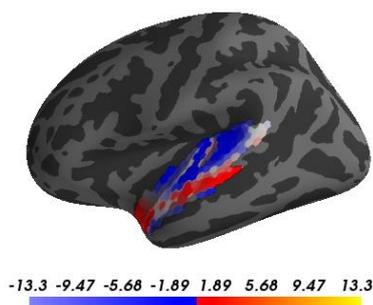


Figure 1: Activity (dSPM) correlated with an interaction between root and linear surprisal 76-99ms after surprising phoneme in primary auditory cortex.

Three clusters of activity in time were significant: from 55-70ms, from 76-99ms, and from 129-141ms (all time locked to the surprising phoneme) with all clusters correlating with interactions between root surprisal and linear surprisal. Activity correlating with interactions with root surprisal implicates the role of online decomposition in auditory word recognition given that root surprisal operates on a purely morphemic level. Furthermore, the results lend further evidence for the primacy of the Semitic root in lexical access, even in auditory processing of a dialect that's not standardly written.

GENDER ATTRACTION IN MODERN GREEK

Anastasia Paspali (Humboldt University of Berlin)
 anastasia.paspali@hu-berlin.de

This study explores the effects of gender attraction (gender agreement errors due to gender interference from a local phrase) in auditory comprehension. Representational accounts support that attraction occurs when the attractor is marked, while memory accounts claim that attraction occurs with unmarked attractors too. Previous comprehension studies have shown mixed results with regards to the experimental manipulations and the methods used: a) gender attraction was found for marked attractors [1, 2], b) attraction was found with unmarked attractors as well [3, 4, 5], (c) lack of gender attraction [6]. In Greek, neuter is unmarked and case-ambiguous (*to.NOM-to.ACC*) on definite articles, while feminine is marked and case-unambiguous (*i.NOM-ti.ACC*). To date, there is only one study on attraction in Greek, which tested number instead of gender attraction and it shows that number attraction occurs in Greek but only with ambiguous/neuter attractors [7]. In this study, we tested gender attraction in Greek participles (Ex1) and object-clitics (Ex2). A 2x2 repeated measures design with *Grammaticality* and *Attractor* as within-subjects variables was used. Neuter and feminine heads and attractors were matched in length and frequency. 52 adult Greek Native Speakers engaged in both experiments (in counterbalanced order & two-week break in between). They completed a self-paced listening task with comprehension questions (Figure 1). Prosody was eliminated through splicing. The analysis revealed significant interactions indicating attraction with feminine heads and unmarked-neuter attractors (Figure 2) in both Experiments; there was a facilitation in reaction times for the ungrammatical mismatch condition. No attraction was observed with neuter heads and marked-feminine attractors. These findings are in line with memory accounts of agreement attraction and the aforementioned Greek study on attraction [7].

Head-Attractor MATCH

- the recipeFEM for the pizzaFEM (**Ex1**) **tornFEM/*NEUT** / (**Ex2**) and found **itFEM/*NEUT**
- the spoonNEUT for the dessertNEUT (**Ex1**) **stainedNEUT/*FEM** / (**Ex2**) and found **itNEUT/*FEM**

Head-Attractor MISMATCH

- the recipeFEM for the breadNEUT (**Ex1**) **tornFEM /*NEUT** / (**Ex2**) and found **itFEM/*NEUT**
- the spoonNEUT for the soupFEM (**Ex1**) **stainedFEM/*NEUT** / (**Ex2**) and found **itNEUT/*FEM**

Figure 1. Experimental stimuli in Experiment 1 (Ex1) and Experiment 2 (Ex2).

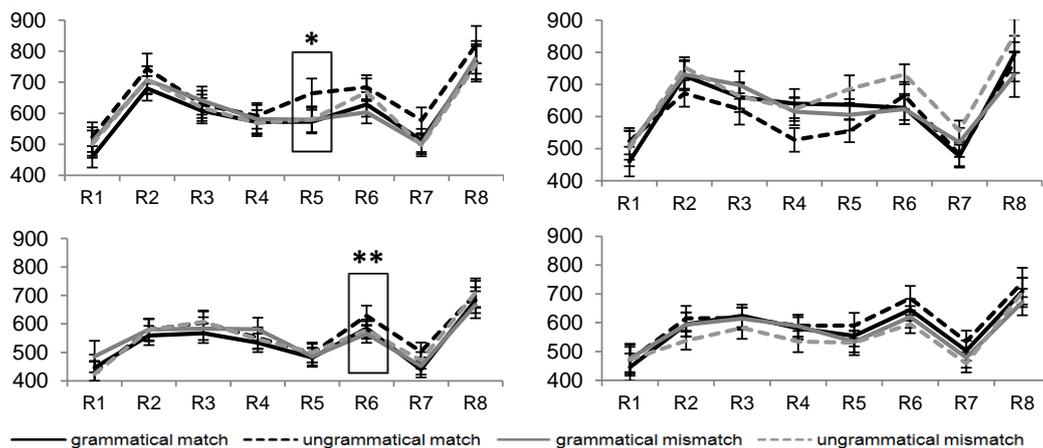


Figure 2. Upper left: Experiment 1 - feminine heads, Upper right: Experiment 1 - neuter heads, Bottom left: Experiment 2 - feminine heads, Bottom right: Experiment 2 - neuter heads. R5 is the critical region of the participle (Ex1) / object-clitic (Ex2).

References: [1] Acuña-Fariña et al (2014), *Lingua*, 143, 108-128; [2] Villata & Franck (2016), 29th *CUNY*; [3] Tucker et al. (2016), 21st *AMLaP*; [4] Slioussar & Malko (2016), *Front Psychol* 7, 1651-1671; [5] Cunnings et al. (2017), 23rd *AMLaP*; [6] Scontras et al. (2018), *Glossa* 3, 1-29, [7] Iraklidou et al. (2011), 17th *AMLaP*

Saturday Poster.62

THE DEVELOPMENT OF IDIOM KNOWLEDGE ACROSS THE LIFE SPAN

Simone Sprenger (University of Groningen, Center for Language and Cognition),
Jacolien van Rij (University of Groningen, Department of Artificial Intelligence)
s.a.sprenger@rug.nl

Idioms are figurative expressions whose meaning is not a function of their component parts (e.g., 'to kick the bucket', to die). Due to the ambiguity that arises from the difference between their literal and figurative meanings, idioms play a major role in psycholinguistic models of multi-word item processing and production. However, little is known about the number of idioms that we know, or how they are acquired: developmental research on idioms has focused on the age at which children are able to understand the difference between an idiom's literal and figurative interpretation, rather than the age at which children acquire specific expressions. Further, the possible variation in the idiom repertoire in native speaker populations has not yet been studied systematically. Finally, the lack of corpus studies on idioms makes it difficult to estimate specific type and token frequencies. In this study we investigate the L1 acquisition of idiom knowledge from adolescence to old age. Given the assumption that idioms can be considered entries in the mental lexicon, our hypothesis is that idiom acquisition follows a similar pattern as single word acquisition, with an early period of rapid expansion, followed by gradual flattening (e.g., Brysbaert et al., 2016). If the idiom vocabulary proceeds to grow until old age in the same way in which the L1 single word vocabulary expands, we would expect that the representations of idioms should be affected by factors such as item frequency, the context of occurrence, and concreteness (or rather idiom transparency).

To get an overview of idiom knowledge across the lifespan, we analyzed the familiarity of 189 Dutch idioms (one or two NPs) and their variability in a sample of 412 native speakers of Dutch (96 men, 316 women, 16-86 years old, mean: 40, sd: 16). Familiarity (on a scale from 1-5) was assessed by means of an online questionnaire and analyzed using Generalized Additive Mixed Modeling (Wood, 2006) for ordered categorical data. Idiom frequencies were obtained from the Lassy Large corpus (van Noord et al., 2012). Transparency ratings for 104 idioms were collected in online questionnaires for English and German speakers (23 participants), who were asked to select the best one-word description (out of three words) for each (translated) Dutch idiom. The three possible answers described the figurative meaning, the literal meaning, and an incorrect but related meaning. The proportion of idiomatic interpretations per idiom was used as an estimate of transparency.

Results of our GAMM analyses convincingly show a significant increase in idiom familiarity with age. The pattern shows a relatively sharp increase before the age of 30, which may be attributed to the increase in language exposure during study time (Brysbaert et al., 2016), followed by a less steep but continuous increase until old age. There is a large variation between items, which decreases with age. Part of this variation can be explained by the item frequencies, which interact with age. The analysis does not show a significant influence of our transparency measure on the idioms' familiarity. Taken together, our findings show that the Dutch idiom vocabulary is a dynamic collection of items that continually changes across the lifespan, with significant development past the age of adolescence. We will discuss these findings in light of theoretical and experimental approaches to idiom comprehension, production, and L2 acquisition.

References

- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in psychology*, 7, 1116.
- van Noord, G.J., Bouma, G., van Eynde, F., de Kok, D., van der Linde, J., Schuurman, I., Tjong Kim Sang, E., Vandeghinste, V. (2012). Large Scale Syntactic Annotation of Written Dutch: Lassy. In: *Essential Speech and Language Technology for Dutch*. Springer.
- Wood SN (2006) *Generalized additive models: An introduction with R*. Florida: Chapman and Hall/CRC.

MENTAL STATE VERBS, ENORSEMENT READINGS, AND THEORY OF MIND

Natalia Talmina and Kyle Rawlins
(natalia.talmina@gmail.com)

Lewis, Hacquard & Lidz (2017) argue that children’s difficulty with processing belief reports is pragmatic: children struggle with inferring the correct Question Under Discussion (i.e., about someone else’s, rather than their own, belief). They argue this difficulty results from the prevalent use of mental state verbs like *think* in the “endorsement” meaning in the course of child-directed speech. For instance, by uttering “*I think it’s time to go to bed*” the caretaker endorses and reinforces the truth of the complement clause. This is different from the prototypical meaning of clause-embedding verbs, where the embedded clause usually carries the main point of the utterance, while the main clause carries some information about the source or reliability of the embedded claim (Simons 2007).

We hypothesize that the ability to have endorsement meanings emerges as a factor of several semantic properties of mental state verbs, such as whether the verb is factive or not, whether it is felicitous in slifting (sentence lifting) constructions (Ross 1973, Simons 2007), and whether it references emotive states (i.e. verbs such as *amaze*, *baffle*).

We explore this hypothesis by eliciting acceptability judgments for sentences involving endorsement in the context of some Question Under Discussion (QUD). Participants are provided with a context and then asked whether a statement involving a verb in the endorsement meaning can be used as an answer to the relevant question under discussion. Statements are rated on a 7-point scale. For example, participants are asked to read the following passage:

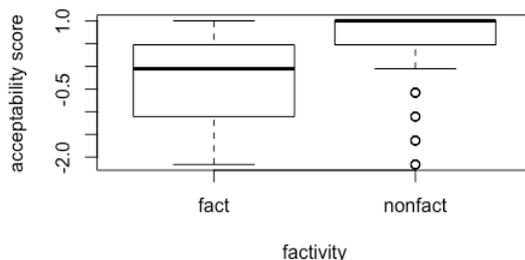
Paul overheard Bill saying that he has a doctor’s appointment today. He also saw Sue preparing some slides. Paul’s boss is asking him: “Is Bill going to attend the meeting, or Sue?”

In this example, “Is Bill going to attend the meeting, or Sue?” is the question under discussion. Then the participants are asked:

Paul replies: “I think that Sue is going to attend the meeting.” Please indicate how natural this answer sounds.

We have tested 15 verbs whose meanings refer to mental states: factive (*understand*, *realize*, *remember*, *find out*, *discover*), non-factive (*think*, *believe*, *suppose*, *guess*, *imagine*) and emotive (*alarm*, *amaze*, *amuse*, *baffle*, *please*). We further manipulated the type of the question under discussion and the knowledge state of the speaker uttering the endorsement statement. QUDs varied between *alternative* and *polar interrogative* questions. In the *knowledge* condition, participants knew that the characters from the scenarios had first-hand information about the QUD (e.g., Paul heard Bill say he couldn’t go to the meeting). In the *ignorance* condition, the characters only had indirect information about the QUD (e.g., Paul knows that Sue usually goes to meetings).

We used R (R Core Team 2012) and lme4 (Bates, Maechler & Bolker 2012) to perform a linear model with acceptability score as the dependent variable and factivity, knowledge state of the speaker, and slifting score (obtained from the MegaAttitude dataset, White & Rawlins 2016) as predictors. Only factivity was a significant factor ($F(3, 551) = 26.25, p < 0.001$). However, there is a possibility that the relationship between slifting and endorsement is non-linear. To explore this possibility further, we are running a large-scale study with ≈ 500 verbs.



DISSECTING STRUCTURAL PRIMING: DIFFERENTIAL PRIMING OF STRUCTURAL FEATURES IN TRANSLATION AND REPEATING

Robert M Maier (Universität Augsburg)

robert.maier@phil.uni-augsburg.de, rob.m.maier@web.de

Reports about fully identical structures as requirement for cross-linguistic structural priming (Jacob, Katsika, Family & Allen, 2016) suggest closer examination of presumable results of structural priming in translation (Maier, Pickering & Hartsuiker, 2017). In a new study, German ditransitive main clauses were translated by native speakers of German into one of two legitimate English target constructions (1a, 1b) from one of three legitimate German source constructions (2a, 2b, 2c).

- (1a) The little boy sent the letter to the neighbour. (PO construction)
- (1b) The little boy sent the neighbour the letter. (DO construction)

- (2a) Der kleine Junge schickte den Brief an den Nachbarn. (prepositional Recipient)
- (2b) Der kleine Junge schickte dem Nachbarn den Brief. (dative Recipient-first)
- (2c) Der kleine Junge schickte den Brief dem Nachbarn. (dative Recipient-last)

According to Jacob et al., we may expect priming to support and possibly facilitate translations between fully identical structures (2a)>(1a) and (2b)>(1b), but not between non-identical (2a)>(1b),(2b)>(1a) or (2c)>(1a/b).

In line with these predictions, analysis of construction choice – cf. col.A in the table below – discovers significant differences between construction choice in translation of (2a) and (2b), similar to an earlier study (Maier, Pickering & Hartsuiker, 2017) that employed only those stimuli (cf. col.D). However, col.A also shows that there are only marginal differences between translations of (2b) and non-identical (2c).

In a control task, informants were requested to merely repeat constructions verbatim in German. Here, the earlier study had reported reliable repetition of (2a) as (2a) and of (2b) as (2b), with only spurious productions of (2c) – cf. col.C. In the new study, however, full reliability of repetition is observed only for (2b). Other constructions are frequently rendered as non-identical constructions, i.e. type (2a) as (2c), and vice versa (cf. col. B). Contextual availability of an alternative construction here appears to interfere with verbatim repetition, similar to reports in Potter & Lombardi (1998). Analyses will attempt to trace this effect in translation responses. These findings suggest that structural priming is not triggered by – linguistically complex – "constructions" in a wholistic fashion. Rather, the influence appears to be exerted by more simple features of structure (here: order of arguments, phrasal status of recipient argument). As these may be involved in functionally comparable constructions of different languages to different degrees, this suggests a more nuanced revision for Jacob et al.'s proposal.

	this study		MPH (2017)	
	A	B	C	D
	trl. (1a)	rep. (2a / 2c)	rep. (2a / 2c)	trl (1a)
(2a)	93%	72% / 27%	99% / 1%	98%
(2b)	63%	0.5% / 0%	0% / 0.5%	67%
(2c)	84%	24% / 72%	---	---

DELAYED ONLINE ATTACHMENT FOR PARENTHETICAL RELATIVE CLAUSES

Marju Kaps, Alexandra Lawn, and Jesse Harris (UCLA)
mkaps@ucla.edu

Sentence processing research has only begun to address when not-at-issue-content is integrated into broader discourse context [1]. Parenthetical relative clauses (PRC) offer a rich test case for investigating the time course of discourse integration, as they allow for attachment ambiguities much like restrictive relative clauses (RRC). We propose that the factors guiding the online integration of RRCs and PRCs are distinct in English: whereas the processor follows structural biases (e.g., *Late closure* or *Recency*) to resolve attachment ambiguities for RRCs early in processing, it delays when integrating PRCs into the structure, possibly constructing a preliminary or underspecified parse before consulting discourse properties, such as discourse salience which may favour high attachment. Further, PRCs are said to contribute not-at-issue content, independent from the main assertion [2]. We predicted that their independence allows the processor to withhold immediate syntactic integration, thereby circumventing locality preferences during incremental sentence parsing.

High and low RC attachment were disambiguated with grammatical number, forcing the RC-internal verb to agree with one of two nouns in a complex object NP, e.g., *brother* (high) or *hosts* (low) in (1). We found no differences in attachment preferences between RRCs and PRCs in two offline norming tasks [vs. 1]. However, an eye-tracking experiment showed an online asymmetry in the time course of RRC and PRC attachment. RRCs exhibited an early high attachment penalty, as previously observed for English [e.g., 3]. In contrast, PRCs showed an online high attachment preference with a delayed time course.

(1) **Sample item for all experiments.** Plurality of NP1 vs NP2 was balanced across items.

RRC High/Low: Everybody met the brother of the hosts who was / were really tall,

PRC High/Low: Everybody met the brother of the hosts (who was / were really tall),

... although the party was really crowded.

Offline norming: A **naturalness ratings** task (N=44) found that items with PRCs ($M=5.03$, $SE=0.08$) were rated as more natural than those with RRCs ($M=4.70$, $SE=0.08$) in a linear mixed effect regression model (as below), $p<.01$, but there was no effect of attachment, $p=.59$, and no interaction between clause type and attachment, $p=.94$. In a **fill-in-the-blank task** (N=38) where the RC verb (*was/were*) was replaced with a blank, roughly equal high attachment completions were provided for PRCs (58%) and RRCs (54%), $p=.22$.

Online processing: In an **eye-tracking experiment** (N=36), an interaction between clause type and attachment in first pass times on the disambiguating RC verb region (*who was/were*), $p<.05$. The overall pattern suggests an early processing asymmetry in attachment sensitivity between PRCs and RRCs. RRCs showed a greater penalty for high attachment on the RC verb region (diff=63ms) than PRCs (diff=21ms). However, there was a crossed interaction for first pass times on the sentence-final region (*was really crowded*), with a low attachment penalty for PRCs (diff=49ms) and a high attachment penalty for RRCs (diff=67ms), $p<.001$. The effects plausibly reflect post-syntactic discourse-integration processes associated with “wrap up” [4], at which point we hypothesize that PRCs are fully integrated into the main structure.

The results suggest that RRCs and PRCs yield distinct processing commitments at different stages in sentence processing. We posit that this finding is linked to the discourse status of the two clause types, so that attachment of independent / non-at-issue content (PRCs) is delayed, a finding that is broadly compatible with Construal Theory and other frameworks in which some elements of a parse may go underspecified in early stages of comprehension.

References: [1] Dillon et al. (2018). No longer an orphan: evidence for appositive attachment from sentence comprehension. *Glossa* 3. [2] Griffiths, & de Vries (2014). Parenthesis and presupposition in discourse. *Linguistics in the Netherlands*. [3] De Vincenzi & Job (1993). Some observations on the universality of the late-closure strategy. *JPR* 22. [4] Hirotsani et al. (2006). Punctuation and intonation effects on clause and sentence wrap-up... *JML* 54.

Saturday Poster.66

AGE OF ACQUISITION RATINGS VALIDATED BY ACTUAL VOCABULARY SCORES

Shalom Zuckerman & Manuela Pinto (Utrecht University)

S.Zuckerman@uu.nl

We present new empirical support for the claim that Age of Acquisition (AoA) word ratings are a valid measurement of the age at which a word has actually been acquired. We show that these AoA values are the best predictors of lexical knowledge as tested with 139 Dutch speaking children.

The Age of Acquisition (AoA) effect (Morrison and Ellis 1997, Brysbeart et-al 2000, Juhasz 2005) has received much attention in recent literature. The basic finding is that AoA ratings of words strongly correlate with accessibility values – the earlier a word has been acquired, the faster it is accessed.

Objections to the AoA phenomenon divide into two types. The first claims that the AoA effect is a secondary effect that can be reduced to frequency. The second type of objections cast doubt on the methodology through which the AoA ratings are collected. These ratings are obtained through subjective evaluations of a (usually small) group of respondents who are asked to estimate in which age they think they have acquired various words. Apart from the subjectivity problem, claims have been made that participants' estimations are affected exactly by the ease in which they can access a word and are therefore not a valid measurement of the words' real AoA. However, recent findings (Brysbeart and Biemiller, 2017) have shown that subjective English AoA ratings correlate with actual test-based values of English words from 37 years ago, as measured in an vocabulary test conducted in 1981. This supports the claim that AoA ratings are real and cannot be explained by the method of collection. We report additional evidence for the relation between AoA ratings and actual current test-based values.

Using a new digital coloring method, which allows for a fast and reliable measurement of children's vocabulary, we tested 139 Dutch speaking children, all of the same age-group (5-6 years old) on their knowledge of 74 Dutch words. This "knowledge-value" of the words was correlated with six different measurements which are currently used for vocabulary assessment in the Netherlands. Five of them are frequency-based and the sixth is the subjective AoA ratings (Brysbeart et al 2014). The frequency-based lists were: (1 and 2) Schrooten and Vermeer 1994 estimation of childrens input, based on 15000 words taken from children books (we used both the general frequency and the specific 4-6 year old frequency). (3) The BAK list (Kuiken and Droge 2010) which is the official list used as a guideline by Dutch schools. (4) SubtlexNL (Keuleers et-al 2010) – a general frequency measurement based on TV and movies subtitles. (5) BASILEX (Tellings et-al 2014) – a list based on 11,4 million words collected from children books.

Surprisingly, children's actual knowledge values correlated the strongest with the AoA ratings ($r=0,63$; $p<0,01$). Partial correlation analysis shows that this relation between AoA ratings and children's current vocabulary knowledge remains significant even when the frequency-based values are controlled for. We claim that these results strengthen the view of AoA as a real phenomenon, and of AoA rating as a valid measurement of the age in which a word has actually been acquired. Further, we claim that guidelines and predictions for vocabulary development in children, should be not based on frequency of words in the input, but rather on empirical measurements of actual word knowledge.

Saturday Poster.67

WH-QUESTIONS ARE UNDERSTOOD BEFORE POLARS

Sara Moradlou (Université Paris-Diderot), Xiaobei Zheng (Shenzhen University), Ye Tian (Amazon Research, Cambridge) & Jonathan Ginzburg (Université Paris-Diderot)
sara.moradlou@gmail.com

In earlier corpus work using the Providence corpus we observed that the ability to answer wh-questions emerges before the ability to answer polar questions. We equate this ability with *question comprehension*. This pattern cannot be explained on the basis of either syntactic or semantic complexity (Hamblin, 1973; Groenendijk and Stokhof, 1984; Ginzburg and Sag, 2000). Caregivers tend to answer their own questions at a very high rate (Moradlou and Ginzburg, 2014), however in our corpus studies we found that self-answering rate is much higher for wh-questions compared to polars. In addition, for certain classes of wh-questions, the answer involves reference to a concrete and sharable entity; for (non-request) polars, this is not the case. Given this, order of understanding is hypothesized to be correlated with questions whose answers are easier to understand. Additionally, it is possible that children answer polar questions later, because the use of polar particles requires a representation of propositions. To better understand different factors in children's ability to answer different types of early questions, we conducted elicitation studies in the context of shared picture book reading. The book consisted of pictures of colorful objects and animals (ten pages, two pictures per page) chosen based on CDI scores in the age group of our study (18-27 months). The motivation of our study is that by limiting the range of questions we would get a better picture of what makes a question hard or easy to answer. We chose a set of three common question types (identity: "what is this?", "is this a cat?", location/existence: "where is the flower?", "is there a flower there?", and animal sounds: "what does the cow say?", "does the cow say moo?") each question type is asked at some point in the experiment in three different conditions, wh, 'yes' polar, and 'no' polar (yes or no being the correct answer to the polar question). The aim here is to control for the contextual setup and truth bias of the questions. The experiment was done in daycares in Germany and China. The experimenters looked at the pictures with each child and asked questions about each picture after establishing joint attention, gave feedback on the child's answers and repeated the questions when needed. Children's answers were then coded based on their type correctness, correctness, and pragmatic congruity. A point was given for matching each of these dimensions and a score out of three was calculated.

We used linear regression to assess the role of each variable (wh vs polar, question type, and question bias) in predicting the answer scores. For German, the model using all three variables predicted the scores significantly better (nested model comparison using anova function in R), and polars significantly decreased, question type location, and identity significantly increased, and bias 'no' significantly increased the score at p-values at least < 0.05 (reference levels: filler, wh, no bias). For Chinese, the model with variables wh vs polar and question type predicted the scores best. Similarly to German, polars, and question type sound significantly decreased, and question type location significantly increased the score at p-values at least < 0.05 . No child did better on polars compared to wh; some children did equally well on both. Our experiments show that wh-questions are easier, even where answering is prompted using truth bias and the context is relatively restricted compared to naturally occurring talk.

References

- Ginzburg, J. and Sag, I. A. (2000). *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.
- Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.
- Hamblin, C. L. (1973). Questions in Montague English. In Partee, B., editor, *Montague Grammar*. Academic Press, New York.
- Moradlou, S. and Ginzburg, J. (2014). Learning to understand questions. In *SemDial 2014 (DialWatt)*.

UNIFORM INFORMATION DENSITY CONSTRAINS OMISSIONS IN FRAGMENTS

Robin Lemke, Lisa Schäfer & Ingo Reich (Saarland University)

robin.lemke@uni-saarland.de

Background Instead of a full sentence (1a), speakers frequently use fragments (Morgan, 1973) as (1b) to get a message across. Despite much theoretical work on the syntax of fragments, the question of why we use fragments is relatively unexplored. Bergen and Goodman (2015) argue based on data from RSA models (Frank and Goodman, 2012) that fragments are used more often the more likely they are understood correctly. However, their data are based on a very small (two question-answer pairs) and relatively artificial data set.

- (1) a. [Conductor to passenger:] “Show me your ticket, please.”
- b. [Conductor to passenger:] “Your ticket, please.”

We present two experiments that suggest that Uniform Information Density (UID, Levy and Jaeger (2007)) determines whether we use a fragment or a sentence. UID states that speakers prefer tend toward distributing information ($-\log_2 p(\text{word}|\text{context})$) (Shannon, 1948), or surprisal (Hale, 2001), uniformly across the utterance. Since omissions contribute to this purpose. UID predicts a preference for fragments in predictive context and for omitting predictable words.

- (2) a. [Conductor to passenger:] “Put your suitcase in the overhead compartment, please.”
- b. [Conductor to passenger:] “Your suitcase, please.”

Rating study We first compared predictable (1) to unpredictable (2) sentences and fragments in a 2×2 (SENTENTIALITY \times PREDICTABILITY) acceptability rating study. Materials were based on event chains (Manshadi et al., 2008) extracted from DeScript (Wanzare et al., 2016), a corpus of script knowledge (Schank and Abelson, 1977). The target utterance (1)–(2) was always preceded by a context story (three events likely to follow each other). In the predictable condition, the target utterance referred to the most likely event given the context story. 48 subjects rated 24 items on a 7-point Likert scale (7=completely acceptable) and completed a survey on familiarity with the tested scripts (5-pt scale, 5=very familiar). The analysis with CLMMs (R, `ordinal` (Christensen, 2015)) reveals significant main effects of both DVs (predictable > unpredictable, sentences > fragments) and a significant interaction between them ($|z| = 2.97, p < .01$): Fragments are specifically dispreferred in the unpredictable condition. Furthermore, models fit to sentences and fragments separately show that the script knowledge scores improve ratings for predictable fragments ($|z| = 3.34, p < 0.001$), but not for sentences.

Production study The rating study supports our general hypothesis, but it does not show whether the less informative expressions are indeed preferably omitted. In a production study, we asked subjects (N=100, German) to read the context stories of exp. 1 and to enter the utterance they considered most likely in this context. Responses were pre-processed by normalizing synonyms and by inserting those nouns and verbs that are required in a grammatical sentence (omitting function words). This yields representations as *show hearer ticket* for (1). For each word we separately annotated whether it was omitted in the original string. We then calculated the bigram surprisal of each word based on the enriched representations (SRILM, (Stolcke, 2002)). We then performed a logistic regression analysis (R, `lme4` (Bates et al. 2015)) with random intercepts and by-scenario random slopes considering SURPRISAL as a predictor for a word’s OMISSION. The analysis confirms that, as predicted by UID, a word is indeed more likely to be omitted the lower its surprisal is ($\chi^2 = 12.31, p < .001$).

Selected references Levy, R. P. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In Schlökopf, B., Platt, J., and Hoffman, T., editors, *Advances in neural information processing systems*, 849–856. MIT Press. • Wanzare, L. D. A., Zarcone, A., Thater, S., and Pinkal, M. (2016). DeScript: A crowdsourced corpus for the acquisition of high-quality script knowledge. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 16)*, 3494–3501.

VOICE MISMATCHES IN VP ELLIPSIS ARE LICENSED BY SYNTACTIC CUES

Robin Lemke, Lisa Schäfer & Ingo Reich

robin.lemke@uni-saarland.de

Background. Verb phrase ellipsis (VPE) in principle allows for a voice mismatch between both conjuncts (1a), but this is not equally possible with all connectives (1b) (Kehler 2002).

- (1) a. This problem was to have been looked into, but obviously nobody did (look into this problem).
 b. #This problem was looked into by John, and Bob did (look into this problem), too.

(Kehler 2002:53 & Kehler 2000:551)

We pursue the hypothesis that this is due to processing. Processing predictable expressions requires less effort than processing unpredictable ones (Levy 2008). In (1) a structurally parallel clause to the first one is probably more likely following *and* than following *but* or *because*. Consequently, a mismatch will be less likely after *and* and require more processing effort, what in turn reduces acceptability. This predicts that (i) mismatches are more acceptable the weaker this *parallelism bias* of the connective is, and that (ii) other cues can further modulate this bias.

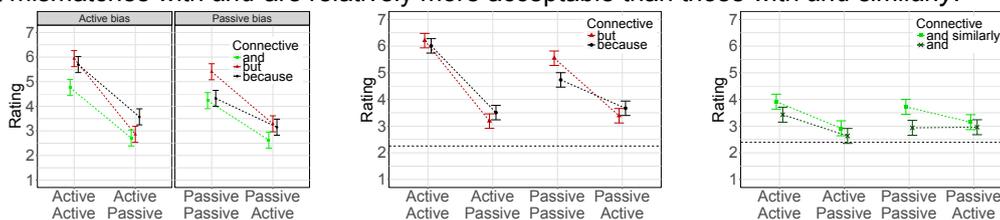
Exp. 1. We first tested all possible combinations of **BIASING** first conjuncts (active (2a)/passive (2b)), voice **MISMATCH** and **CONNECTIVE** (*and*, *but*, *because*) in a 2×2×3 design.

- (2) a. Joshua didn't give Sarah private lessons in Mandarin (and | but | because) Jacob (did | was).
 b. Sarah wasn't given private lessons in Mandarin by Joshua (and | but | because) Jacob (did | was).

96 subjects (half saw the active and half the passive bias conditions) rated 30 items on a 7-point Likert scale (7 = fully acceptable). The analysis with CLMMs (R, *ordinal* (Christensen 2015)) partially confirms our hypothesis (fig. 1): Significant main effects of **CONNECTIVE** show that *but* is overall more acceptable than *because*, but significant **CONNECTIVE:MISMATCH** interactions show that voice mismatches improve with *because* compared to *but* ($z_{\text{active bias}}=4.897$, $z_{\text{passive bias}}=5.74$, both $p<0.001$). Still though, the same holds for *and*. Possibly *and* triggers the expectation of parallel polarity between the conjuncts, which our materials violate.

Exp. 2. To account for a possible floor effect we tested the *but* and *because* conditions with more ungrammatical fillers (2×2×2 within subjects design, **BIAS** × **CONNECTIVE** × **MISMATCH**, n=64). The significant **CONNECTIVE:MISMATCH** interaction was replicated ($z=-7.38$, $p<0.001$). All conditions were more acceptable than the ungrammatical fillers (the dashed lines in fig. 2,3).

Exp. 3. We investigated the effect of the adverb *similarly* on mismatches with *and* (2×2×2 within subjects design, n=64). Both conjuncts had positive polarity in order to make the *and similarly* conditions plausible. While *and similarly* was significantly preferred over *and* ($z=9.05$, $p<0.001$), a significant **CONNECTIVE:MISMATCH** interaction ($z=-4.56$, $p<0.001$) showed that mismatches with *and* are relatively more acceptable than those with *and similarly*.



Exp 1: Means and 95% CIs.

Exp 2: Means and 95% CIs.

Exp 3: Means and 95% CIs.

Discussion. Exp. 1 and 2 showed that mismatches improve significantly with the subordinating *because* compared to *but* as predicted by our and Kehler's (2002) account. Exp. 3 showed that mismatches with *and* are further degraded if the parallelism bias is strengthened by the additional adverb *similarly* before *and*. This is also predicted by our account, but does not directly follow from the categorial split predicted by Kehler's (2002) theory.

Selected references. Kehler, A. (2002). 'Coherence and the resolution of ellipsis'. *Linguistics and Philosophy* 23, 533–575 • Levy, R. (2008). 'Expectation-based syntactic comprehension'. *Cognition* 106, 1126–1177

Saturday Poster.70

NULL PRONOUN IS ALWAYS BETTER THAN OVERT. BEHAVIORAL AND EYE-TRACKING EVIDENCE ON ANAPHORA RESOLUTION IN POLISH LANGUAGE

Agata Wolna, Joanna Durlik, Jakub Szewczyk, Michał Remiszewski, Zofia Wodniecka (Jagiellonian University)
(agatawuwu@gmail.com)

Only recently has psycholinguistics started to investigate the question of how universal the mechanisms of anaphora resolution are and to what extent they are language-specific. For example, in case of the pronominal anaphora, a question arises to what extent pro-drop languages differ in the interpretation of null and overt pronouns. According to the Accessibility Theory (Ariel, 1990), possibility to use null pronouns as an alternative to overt pronouns may be modified by both contextual and language-specific conditions, but at least some of these principles are expected to have universal validity across languages. Still, majority of data showing how pronominal anaphora is resolved in pro-drop languages comes from studies on Spanish and Italian. Here, we aimed to verify whether pronominal anaphora resolution in Polish is similar to that observed for Roman languages in which the null pronoun subordinate clause is interpreted as referring to the subject antecedent of a main clause, while overt pronouns are typically linked to the object of the preceding clause (Chamorro, Sorace, Sturt, 2016). We conducted two studies with native Polish speakers. In Study 1, participants ($n = 70$) read ambiguous sentences, in which the subordinate clause might refer to both subject and object of the main clause (e.g. “A **woman** waved to a **girl** when **she** was crossing the street”). Participants were asked to choose between two interpretations of the sentence by answering a question “Who crossed the street?”. This study aimed to assess the natural pattern of anaphora resolution in Polish, and should be treated as a preliminary study before the study proper (2), since no empirical data on the topic were available so far. In the second experiment, a new group of participants ($n = 27$) was presented with unambiguous sentences, which forced only one reading of the subordinate clause (as referring to either subject or object of the main clause as antecedent); in null or overt pronoun condition, meeting or not the antecedent preferences in Polish (e.g. “A **woman** waved to the girls when **she** was crossing the street” vs. “The women waved to the **girl** when **she** was crossing the street”). We recorded participants eye-movements to properly assess the process of sentence interpretation. This study aimed to examine how is anaphora resolution affected by antecedent preferences in Polish. Additionally, in both experiments, participants were asked to rate the naturalness of each sentence. For behavioral data in each experiment we conducted separate ANOVAs, comparing sentence conditions. For Study 1, a t-test run on interpretation of the sentences and naturalness judgments showed that in ambiguous sentences Polish speakers tended to interpret the object as the antecedent of the overt pronoun and the subject as the antecedent of the null pronoun. For the behavioral results of Study 2, we run a repeated-measures ANOVA, which showed that for unambiguous sentences with overt pronoun the object-match is considered more natural than a subject-match. Repeated-measures ANOVAs for the eye-tracking data showed interaction between antecedent (subject or object) and pronoun (null or overt); which demonstrated that in the subject-match condition the interpretation of the subordinate clause was hindered compared to the object-match condition. The results of Study 1 and Study 2 show that speakers of Polish clearly prefer to match null pronouns to subject antecedents and overt pronouns to object antecedents. Moreover, the results suggest that contextual disambiguation and linking the overt pronoun with the object antecedent does not facilitate processing subordinate clauses containing overt pronouns, compared to null pronoun sentences with object antecedent.

LEARNING CONSISTENT GENDER ERRORS IN NON-NATIVE SPEECH

Thomas St. Pierre (University at Buffalo) & Jean-Pierre Koenig (University at Buffalo)
 tastpier@buffalo.edu

Listeners use grammatical gender information in the determiner to help process upcoming nouns (e.g., Dahan et al., 2000; Hopp, 2012). However, when gender is unreliable, it is no longer used as a cue (van Heugten & Christophe, 2013). A visual world experiment tested whether German listeners take advantage of incorrect gender-marking when used *consistently* (e.g., *die*_[FEM] *Kuh* ‘the cow’ labelled **der*_[MASC] *Kuh*). Results indicate that listeners use incorrect gender marking, but only in cases where incorrect gender can help reduce ambiguity.

L1 German speakers (N = 28) heard 12 stories from an L1 or L2 speaker. Crucially, in the L2 stories, one noun was always incorrectly gender-marked. After each story, participants were asked to locate these nouns on a computer screen (*Wo befindet sich* DET + ADJ + NOUN? ‘Where is X located?’). In critical arrays, the target could either be uniquely identified by its gender (Unambiguous) or shared gender with another object (Ambiguous). In the Non-native condition, a normally unambiguous target (*die*_[FEM] *alte Kuh* ‘the old cow’ occurring with a dog_[MASC]) could be ‘ambiguous’ if listeners remembered the non-native error (**die*_[FEM] *alte Hund* ‘the old dog’) and vice versa (see Figure 1 for an overview of the conditions).

Figure 2 plots the proportion of looks to the pictures across time. A logistic regression comparing looks to the target vs. the “competitor” was conducted starting 180 ms after noun onset (before eye gaze could be influenced by the noun), with Time, Nativeness and Ambiguity as predictors (Barr, 2008). Crucially, there was a significant interaction between Nativeness and Ambiguity ($b_1 = -1.67, t_1 = -2.40, p < .05$). Specifically, while there were no differences in looks in the L1 Ambiguous condition ($b_1 = 0.03, t_1 = 11, p = .91$), there were significantly more looks to the target in the L2 Ambiguous condition ($b_1 = 1.23, t_1 = 3.47, p = .001$), suggesting listeners were taking into account the L2 gender errors to anticipate the target noun. In the Unambiguous condition, however, there were significantly more looks to the target in both the Native ($b_1 = 1.41, t_1 = 3.85, p < .001$), and Non-native conditions ($b_1 = 0.93, t_1 = 2.58, p = .01$), suggesting listeners had not learned to treat the incorrectly marked nouns as true competitors.

Listeners thus learned consistent, incorrect gender-marking, but only where knowledge of the incorrect gender could help constrain possible candidates. The failure to learn errors in the Unambiguous condition might mean that the influence of the L2 representations was weak. In the Non-native Ambiguous condition, the determiner highly activates two nouns in the listener’s native grammar, and the weak influence of incorrect gender helps inhibit the non-target. In the Non-native Unambiguous condition, however, only one noun is highly activated in the listener’s grammar, and the influence of incorrect gender is not strong enough to increase the activation of a gender-mismatching noun.

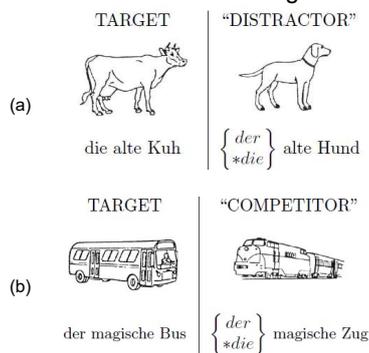


Figure 1: (a) L1 Unambig. / L2 'Ambig.'
 (b) L1 Ambig. / L2 'Unambig.'

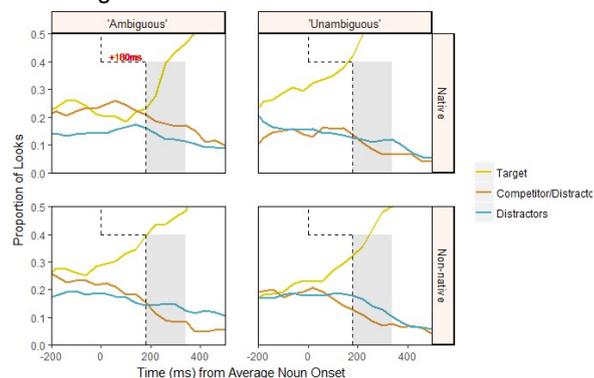


Figure 2: Proportion of looks to object (average noun onset = 0)

Saturday Poster.72

PASSIVE SENTENCE DIFFICULTY: IT'S NOT ABOUT ARGUMENT ORDER BUT A STATE OF MIND

Caterina Paolazzi (UCL), Claudia Cera (UCL), Nino Grillo (University of York), Artemis Alexiadou (Humboldt-Universität zu Berlin) & Andrea Santi (UCL)
caterina.paolazzi.11@ucl.ac.uk

Passive sentence difficulty has played a key role in theories of language acquisition, language breakdown and adult processing. Its difficulty has been argued to be (at least) greater with stative than eventive predicates for multiple reasons: (1) temporary ambiguity (adjectival vs. verbal interpretation), (2) raising an unaffected argument, and (3) coercion of the underlying state to an eventive reading [1]. Recent work in the healthy adult population provides contradictory evidence for this interaction: acceptability judgements [3] support it, but self-paced reading (SPR), and comprehension accuracy data do not [2]. Experiment 1 further explored an online effect of predicate semantics interacting with voice using the Eye-Tracking while reading technique that, unlike SPR, is not limited to early aspects of processing. An interaction emerged in late measures. Experiment 2, further investigated an interaction in the offline accuracy effect considering task biases across conditions. As the voice of the comprehension question was always in the active, it could have biased processing in favour of the active by generating interference in memory (from the linear order of NPs) for the passive. Experiment 2 manipulated the voice of the question to (mis)match the sentence in a web-based comprehension task. An interaction between predicate semantics and voice was observed.

Experiment 1. 48 British native English speakers participated in the eye-tracking while reading study using a 2 Voice (passive active) x 2 Predicate Class (event – A, state – B) design. (A) The guitarist (was) rejected (by) the attractive and talented singer on Tuesday morning. (B) The guitarist (was) admired (by) the attractive and talented singer throughout the tour. Pre-norming studies ensured equivalent plausibility of thematic role assignment. At the verb and post-verbal noun-phrase, a significant interaction between predicate and voice was observed in Total Time (verb: $\beta=188.62$, $t=3.38$, $p<.001$; post-verbal noun: $\beta=98.83$, $t=2$, $p=.045$). In both regions, stative predicates had longer fixation durations in the passive than active voice, while eventives showed no difference or a difference in the opposite direction. This effect is commensurate with resolving the temporary ambiguity and/or coercion effects.

Experiment 2. 160 native British English speakers were recruited from the web-based platform *Prolific Academic* for this 2 Voice (active, passive) x 2 Predicate Class (event, state) x 2 Question Match (yes, no) experiment. Participants rated the acceptability of the sentence and then answered a comprehension question targeting thematic role assignment. No significant effect emerged in the acceptability data, but in the accuracy data there were 2 significant interactions: Voice*Question Match ($\beta=0.452$, $z=2.15$, $p=.03$) and Voice*Predicate Class ($\beta=-0.982$, $z=-4.70$, $p<.001$). More errors arose when the question mismatched the sentence in voice and this difference was greater for active sentences ($\beta=-0.982$, $z=-4.70$, $p<.001$). This confirms that previous effects of voice on comprehension accuracy likely arose from memory demands induced by the mismatch. More importantly, with eventive predicates accuracy did not differ across voice, but with stative predicates, significantly more errors arose with passive sentences than active ones ($\beta=-0.77$, $z=-5.51$, $p<.001$).

Collectively, the results converge on passive sentences being difficult *only* in the case of stative predicates. We argue the difficulty lies in representing the eventive reading of a stative predicate that requires both ambiguity resolution and coercion. The results also support two important themes in sentence processing: (1) task interactions [1] and (2) predicate semantics' interaction with structure [6]. The broader implication is that there is no general difficulty with noncanonical argument orders [5] counter the decades of work in this direction.

[1] Messenger et al. (2012). *Journal of Memory and Language*, 66, 568-587. [2] Paolazzi et al. (2015). *AMLaP*. [3] Ambridge et al (2016). *Cognitive Science*, 1435-1459. [4] Gehrke & Grillo (2009). *Interface Explorations*, 231-268. [5] Ferreira (2003). *Cognitive Psychology*, 164–203. [6] Grillo et al. (2015). *Cognition*, 116-122.

NEURAL MECHANISM FOR PRONOUN RESOLUTION IN CHINESE DURING NATURALISTIC LISTENING

Jixing Li, Murielle Fabre, Wen-Ming Luh and John Hale
(jl2939@cornell.edu)

Introduction Pronoun resolution is influenced by syntactic and morpho-syntactic constraints as put forward in the Binding Theory (Chomsky, 1981), but discourse factors such as prominence of the antecedents (Grosz et al., 1995) are also relevant in pronoun resolution. Chinese pronouns bear no gender marking on their spoken forms, thus providing little morpho-syntactic cues for pronoun resolution. In addition, Chinese allows *pro*-drop and is considered as a discourse-oriented language where topic-comment articulation defines the basic sentence structure (Li & Thompson, 1976; Tsao, 1977; Xu, 2006). It is therefore hypothesized that Chinese speakers would rely more on discourse factors to identify the reference of the pronouns. The current study compared brain activity associated with complexity metrics for pronoun resolution derived from a syntax-sensitive and a discourse-based computational model while Chinese participants listened to an audiobook during fMRI recording. The syntax-sensitive Hobbs algorithm (Hobbs, 1977) follows Binding Theory, whereas the discourse-based ACT-R model (van Rij et al., 2013) calculates how salient the antecedent is given the grammatical role, frequency and recency of the entity in the discourse context. Compared with the Hobbs metric, the ACT-R metric elicited a larger left lateralized frontoparietal network including the left Angular Gyrus (AG), consistent with previous work on the main effects of anaphora (e.g., Matchin et al., 2014; Santi & Grodzinsky, 2012). The results supported our hypothesis that pronoun resolution in Chinese relies more on salience-based discourse factors.

Methods 35 Chinese speakers (15 female, mean age=19.3) listened to a Chinese audiobook version of "The Little Prince" for about 100 minutes. BOLD functional scans were acquired using a multi-echo planar imaging (ME-EPI) sequence with online reconstruction (TR=2000 ms; TE's=12.8, 27.5, 43 ms; FA=77; matrix size=72x72; FOV=240.0x240.0 mm; 2 x image acceleration; 33 axial slices, voxel size=3.75x3.75x3.8 mm). Preprocessing was carried out with AFNI and ME-ICA (Kundu et al., 2011). The audiobook contains 388 third person pronouns, and for each of the third person pronouns, we computed the Hobbs distance, namely the number of noun phrases that the algorithm skips while searching for the antecedent. Our hypothesis is that a higher Hobbs distance yields a processing effort due to higher syntactic and morphological constraints. We also computed the activation level of the antecedent for the third person pronouns using the ACT-R model. We then took the negative of the ACT-R scores to indicate processing difficulty for pronoun resolution based on the salience of the antecedent. The observed BOLD signal was modeled by the two complexity metrics for pronoun resolution, time-locked at the offsets of the third person pronouns in the audiobook. We also included a binary regressor which marks the presence of the third person pronouns, and three control regressors: RMS intensity at every 10 ms of the audio; word-rate at the offset of each word, and log-frequency of each word in Google Books.

Results and Conclusion The Hobbs metric correlates with significant clusters in the left Precuneus. The ACT-R model is associated with significant activation in a left lateralized network including the AG, MTG, SFG, MFG, IFG, and STG ($p < 0.05$ FWE; see Figure 1). The larger ACT-R effect suggests that the salience-based model may be more accurate for pronoun resolution in Chinese.

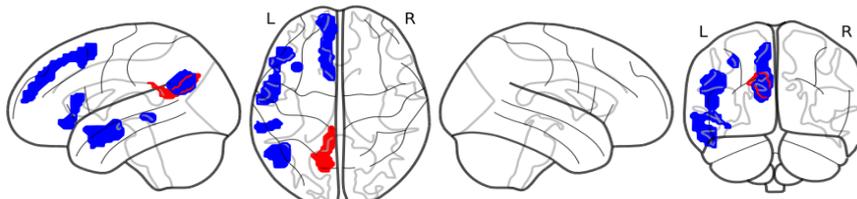


Figure 1. Activation map for the main effect of the Hobbs and ACT-R metric for third person pronouns in Chinese ($p < 0.05$ FWE, $k > 50$). Red color represents the Hobbs metric effect, and blue color represents the ACT-R metric effect.

UNDERSPECIFICATION IN RC ATTACHMENT

Pavel Logačev and Noyan Dokudan (Bogazici University, Istanbul)
 pavel.logacev@boun.edu.tr

An important aspect of how we process ambiguity in sentence comprehension is how the parser handles attachment of ambiguous RCs. One possibility is that readers are able to underspecify ambiguous RC attachment. That is, the mental representation of an ambiguous sentence faithfully represents the ambiguity in the input. Underspecification has been proposed as part of Frazier & Clifton's (1997) Construal theory. Swets et al. (2008) have also argued for underspecification of RC attachment as an explanation of Traxler et al.'s (1998) finding of the *ambiguity advantage* in sentences like (1), i.e. faster processing relative to an unambiguous baseline. A similar assumption is made by Surprisal (Levy, 2008), which predicts that the parser does not attempt to disambiguate ambiguous sentences, unless a disambiguating cue is provided, thus resulting in faster reading in ambiguous sentences.

(1) *The son of the driver who had a moustache was pretty cool.*

In order to test the hypothesis that the parser can maintain underspecified representations of ambiguous sentences, we conducted an experiment in Turkish (N = 68), in which participants read two sentences on every trial, and judged the acceptability of the second sentence in the context of the first. The experiment consisted of 116 sentence pairs: 56 experimental items, and 60 fillers. Because Turkish is head-final, RCs preceded the complex noun phrases they modified. Experimental items followed the structure of (2), and were presented in three attachment conditions: *local*, *non-local* and *ambiguous*. In unambiguous conditions, the attachment was disambiguated by animacy. After reading the first sentence in a self-paced manner, participants saw a probe sentence, which was a Turkish equivalent of (3a) or (3b). They judged whether the probe was felicitous in the context of the first sentence.

(2) Bugün mahkemede [_{RC} dün tutuklandığı iddia edilen] sekreterin/okulun müdürü/cüzdanı görülmüş.
 today in court [yesterday arrested supposedly] secretary's/school's manager/wallet was seen
 'The manager/wallet of the secretary/school who was supposedly arrested yesterday was seen in court today.'

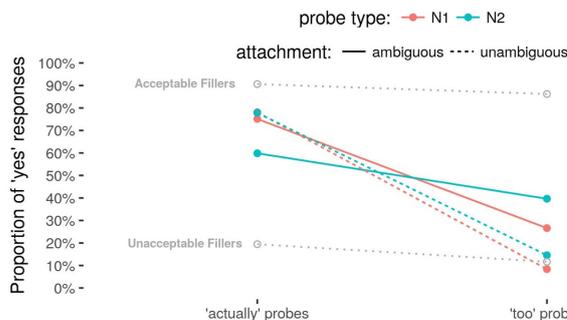
(3a) *Ali claims that the secretary/manager was actually arrested last week. ['actually' probe, N1 and N2]*

(3b) *Ali claims that the secretary/manager, too, was arrested.*

['too' probe, N1 and N2]

Importantly, the acceptability of the probe given the previous sentence depended on one's interpretation of RC attachment in the first sentence. While (3a) presupposes attachment to the NP mentioned in the probe, (3b) presupposes attachment to the other NP. A violation of these presuppositions should lead to 'unacceptable' responses. Importantly, in the case of ambiguous RC attachment, the first sentence always has *one* reading which is compatible with any of the probe sentences. Thus, if readers underspecify RC attachment, we expect that probe sentences will be classified as 'acceptable' as often as unambiguous acceptable filler sentences.

However, as illustrated in Fig. 1, the percentage of 'acceptable' responses in ambiguous attachment conditions was lower than the percentage of such responses in acceptable filler sentences.



This was confirmed by a Bayesian linear-mixed model with a logit link (lower boundaries of all 95% CrI > 1.1). This finding is not compatible with underspecified representations. Importantly, it is also not compatible with the maintenance of a single structure. Instead, we argue that this finding is compatible with simultaneous storage of multiple structures.

Saturday Poster.75

**DEVELOPMENT OF A LANGUAGE-INDEPENDENT SYSTEM
FOR AUTOMATIC EVALUATION OF L2 ORAL REPRODUCTION TASKS
USING A DEEP LEARNING ALGORITHM**

Yutaka Yamauchi (Soka University), Nobuaki Minematsu (The University of Tokyo),
Kayoko Ito (Koyasan University), Megumi Nishikawa (Tokai University),
Kay Husky (Tokyo International University) & Aki Kunikoshi (University of Groningen)
yutaka@soka.ac.jp

In this study, a new deep learning-based algorithm has been developed to automatically evaluate L2 oral reproduction tasks. This DNN (Deep Neural Network)-based algorithm can convert model voices and learners' utterances into sequences of posterior vectors which are phoneme-state posteriors with a few thousands of sound classes, directly compare the two utterances based on the distance between them using the DTW (Dynamic Time Warping) technique, and give scored feedback.

Since an average language possesses several dozens of phonemes, a few thousands of sound classes can be considered to contain and cover almost all phonemes of any language. When phonemes of a particular language are used for posterior vectors, the obtained representations are strongly dependent on the language. On the other hand, when phoneme states with a few thousands of sound classes are used for posterior vectors, the representations will be much more language-independent, or multi-lingual oriented.

An experiment was conducted to test two predictions: (1) Automatic scores derived from the distance between model voices and learners' utterances will have negative correlation with manual ratings based on accuracy of L2 oral reproduction tasks. (2) Automatic scores obtained from posterior vectors using the target language and automatic scores derived from posterior vectors using a language different from the target language will have almost the same correlation with manual ratings based on accuracy of L2 oral reproduction tasks.

A total of 124 L2 learners of English were requested to orally reproduce the target passages. Their recorded utterances were assessed in three ways: (1) automatic evaluation using the target language (English in this experiment) in converting model and learners' utterances into posterior vectors, (2) automatic assessment using a language other than the target language (Japanese) in the posterior vector converting, and (3) manual assessment by two veteran language instructors with the five-point Likert scale focusing on accuracy of pronunciation, prosody and accessibility of one's mental lexicon.

The experimental results showed that automatic scores and manual ratings had a significant negative correlation. The correlation value between automatic scores using the target language (=English) and manual ratings ($r=-.79$) was as high as that between automatic scores using the different language (=Japanese) and manual scores ($r=-.74$). The difference between the two correlation values was not significant ($p<.001$).

After the 124 participants were divided into three groups (high, middle and low, based on the manual ratings), significantly high correlation values were observed in the high and low groups, but not in the middle group. This implies that gathering participants with different ranges of proficiency might be crucial for this algorithm to work well.

The correlation value between automatic scores using the target language and those using the different language was .87 ($p<.001$). After the 124 participants were divided into three groups, significantly high correlation values were observed across the three groups: $r=.75$, $r=.70$, and $r=.83$, respectively ($p<.001$).

Therefore, it was shown that the two predictions above were true, and that the algorithm in this study could directly compare utterances in a language-independent way and be used as a multi-lingual system for automatic evaluation of L2 oral reproduction tasks.

- G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath & B. Kingsbury. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol.29, no.6, pp.82-97, 2012.
- J. Yue, F. Shiozawa, S. Toyama, Y. Yamauchi, K. Ito, D. Saito & N. Minematsu. "Automatic scoring of shadowing speech based on DNN Posteriors and their DTW," *Proceedings of INTERSPEECH*, pp.1422-1426, 2017. DOI: 10.21437/Interspeech, 2017-728.

Saturday Poster.76

PROCESSING DYNAMICS OF SUBJECT-VERB AGREEMENT: A SPEED-ACCURACY TRADEOFF ANALYSIS

Ryan King & Brian McElree (New York University)
ryan.king@nyu.edu

Resolving dependencies during sentence comprehension requires access to previously encoded information. Constraints on the amount and type of information that can occupy the focus of attention during processing, and properties of the retrieval operation that mediates access to this information when it has been displaced from attention, crucially determine the success of establishing dependencies. Previous research employing speed-of-access measures to investigate subject-verb dependencies show fast dynamics for adjacent dependencies and those separated by a prepositional phrase (PP), but slower dynamics for those separated by a relative clause (RC), indicating retrieval (McElree et al. 2003, Wagers & McElree, 2013). Furthermore, this retrieval operation is susceptible to interference: comprehension errors can result from retrieving an inappropriate representation as a consequence of a better or stronger match to the retrieval cues (Gordon et al., 2002; Van Dyke & McElree, 2011).

Previous research measuring the accessibility of information during sentence processing has primarily focused on dependencies based on constituent-level information. Here, we extend this approach to feature-based dependencies, which involve unifying or matching features between constituents, by investigating sentences containing subject-verb agreement. Marked features, such as the plural, may be explicitly encoded, hence more available in memory when forming dependencies based on number (as in agreement). As a consequence, marked features on inappropriate (non-agreement controlling) nouns may engender more interference, as suggested by prior studies on agreement attraction in comprehension (Tanner et al., 2014; Wagers et al., 2009). The aims of the study were two-fold: 1) to investigate the accessibility of number information across different types of intervening material and 2) to determine how the availability of required information influences establishing dependencies when carried by the target head noun or intervening lure nouns.

We report the results of two experiments (both $n = 20$) that manipulated the type of intervening material (**Experiment 1**: Adjacent vs SRC; **Experiment 2**: Adjacent vs Adverb vs PP) to estimate the accessibility of number information, and we fully crossed number marking on the subject noun and the non-agreement controlling lure noun (Singular vs Plural) to explore the effect of availability. Forty sets of sentences like the ones shown below were created for each experiment:

Experiment 1: *The doctor(s) [who lost the scalpel(s)] complain(s)*

Experiment 2: *The doctor(s) [abruptly/with the new scalpel(s)] complain(s)*

We employed the multiple-response speed-accuracy tradeoff (MR-SAT) procedure, which provides separate measures of discriminability and processing speed. Participants had to discriminate acceptable sentences where the verb agrees with its subject (*The doctor complains*) from unacceptable sentences where the verb violates agreement (*The doctor complain*) in response to a series of tones beginning concurrently with the onset of the final word.

An ANOVA on asymptotic values revealed a main effect of Intervening Material and no interaction with Number in both experiments: In Exp 1, d' values were significantly higher for Adjacent versus SRCs. In Exp 2, d' values were higher for Adjacent and Adverbs versus PPs, suggesting that the presence of a lure noun reduced discriminability. Differences in dynamics were assessed with model fits, which supported a 3-way rate distinction in both experiments (**Exp1**: Adjacent > RCs > RC-Target Singular Lure Plural; **Exp2**: Adjacent > [Adverbs + PPs] > PP-Target Singular Lure Plural). These results suggest that, unlike constituent-based dependencies, feature-based dependencies require retrieval in order to integrate the verb with the head noun whenever they are nonadjacent. More generally, these results indicate that memory operations are more ubiquitous in comprehension than previously recognized.

PRIMING UNGRAMMATICAL STRUCTURES ACROSS LANGUAGES

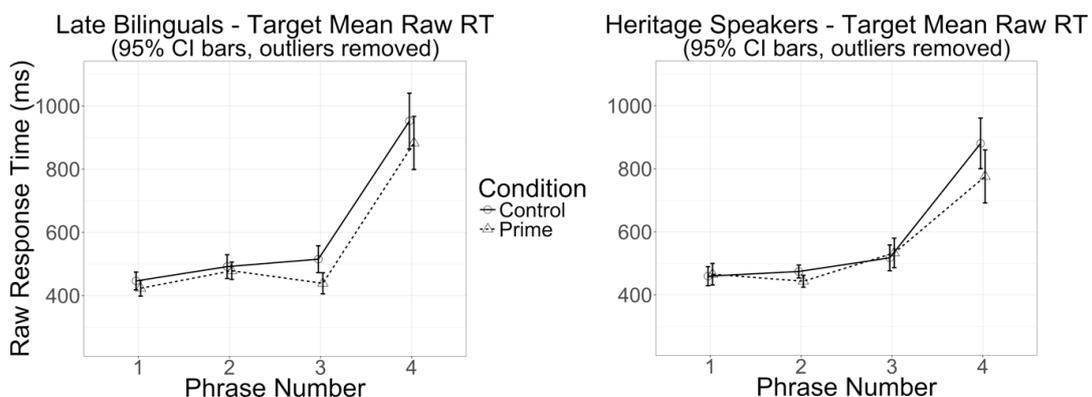
Ian Phillips (City University of New York Graduate Center)
iphillips@gradcenter.cuny.edu

Growing evidence suggests that *cross-linguistic structural priming* may play a potentially important role in driving contact-induced language change (Fernández, Souza, & Carando, 2017; Kootstra & Doedens, 2016). While the available evidence shows that priming from a later-learned language (L2) can subtly shape the first-learned language (L1) by increasing the frequency with which bilinguals produce syntactically licit but dispreferred L1 structures (Carando, 2015) and by extending licit L1 structures to new L1 environments (Torres Cacoullos & Travis, 2011), little is known about whether cross-linguistic priming can shift L1 core syntactic processing routines. To address this gap, the present study tests whether, for Spanish-English bilinguals, exposure to licit preposition stranding in L2 English sentences (Ex. 1) facilitates processing of structurally-parallel but illicit L1 Spanish sentences (Ex. 2).

- (1) *These are the scissors | that Mary | cut the paper with.*
- (2) **Este es el serrucho | que Eduardo | cortó la rama con | para hacer leña.*
"This is the saw that Eduardo cut the branch with to make firewood."

63 Spanish-English bilinguals (41 heritage Spanish speakers (HS), 22 late Spanish-English bilinguals (LB)) listened to sentences like (1–2) presented phrase-by-phrase in prime-target pairs in a self-paced listening task, controlling presentation rate by pressing a button to hear the next phrase in the sentence (phrases indicated with '|' in Ex. (1–2)). In prime trials, Spanish target sentences like (2) followed structurally-parallel English sentences like (1), with translation equivalents of the verb and preposition. In control trials, Spanish target sentences followed structurally different English sentences, e.g., *This is the paper that Mary cut with the scissors*. Log-transformed response time (RT) was analyzed separately for each target sentence phrase using linear mixed-effects models in R.

Phrase 3 model estimates show significantly reduced RTs in prime trials for LB ($\beta = -0.12, SE = 0.05, p < .05$) but not HS. Phrase 4 model estimates show significantly reduced RTs in prime trials for high Spanish fluency HS subjects ($\beta = -1.02, SE = 0.45, p < .05$) but not LB. LB subjects who had stronger Spanish (higher use, fluency, and exposure), but not HS, also showed cumulative priming effects from repeated exposure to Spanish sentences like (2). These findings suggest cross-linguistic priming effects for ungrammatical sentences like (2) were tied to residual activation of preposition stranding (Hartsuiker et al., 2004) for both groups, while LB also implicitly learned to process Spanish sentences like (2) (Chang et al., 2000). An analysis of task evoked pupillary responses (TEPR) collected during the same experiment is also reported, which shows pupil dilations generally track with RT results and indicates that pupillometry is sensitive to priming during comprehension. The significance of these findings for research on bilingual syntactic processing and the relevance of this work in contributing to our understanding of contact-induced language change are discussed.



WASTED POWER AND FALSE INTERACTIONS IN READING TIME ANALYSES

Zachary Burchill (University of Rochester), Wednesday Bushong (University of Rochester) & T. Florian Jaeger (University of Rochester)
 zachary.burchill@rochester.edu

Reading data constitute a widely used paradigm, and many common analysis methods (ANOVA, linear mixed models) assume the residuals of reading times (RTs) are normally distributed. This assumption is known to be false. It is unclear, however, whether this has practical consequences: some violations of modeling assumptions have consequences for Type I & II error rates; others do not. We address this question through non-parametric Type I/II error simulations over real RT data. We compare the standard approach to a simple potential fix: log-transforming the RTs. **The issue:** Like in other reaction time tasks^{1,2}, RTs often exhibit heavy skew, leading to strong correlations between means and standard deviations (SDs). Such correlations are not expected under normality, and violate the assumptions of standard analyses, but are predicted for, e.g., log-normal or ExGaussian distributions. We use real self-paced reading (SPR) data⁴ (fillers only, 70 trials, 169 subjects, 143,786 by-word RTs). We analyze both the *raw* (incl. massive outliers) and *cleaned* data after standard outlier exclusion ($100 \leq RT \leq 2000$ ms). The same issues reported here hold for length-corrected RTs, by-subject standardized RTs, and when incorrectly answered trials are excluded. RTs means and SDs were strongly correlated by trial or by subject ($0.65 < r^2 < 0.82$). **Approach:** To assess the consequences of actual RT distributions on Type I & II error, we bootstrap with replacement from the SPR data³. This avoids the circularity of *parametric*

Table 1: Study 1 comparison of linear and log-linear models when k constant in linear space. 10,000 bootstrapped datasets per cell.

Bin	Keep outliers				Exclude outliers			
	Linear		Log-linear		Linear		Log-linear	
	Power	TypeI	Power	TypeI	Power	TypeI	Power	TypeI
1	0.068	0.019	0.55	0.053	0.36	0.052	0.60	0.049
2	0.124	0.020	0.86	0.053	0.59	0.054	0.88	0.054
3	0.190	0.019	0.91	0.050	0.68	0.052	0.92	0.052

power analysis assuming normality.⁴ Each study used 10,000 bootstrapped data sets (BATA). Each data point of a BATA was randomly assigned to one of two (balanced) conditions. We then added a main effect of $\pm k_M$ (ms) between the two

conditions. Power is the proportion of BATAs for which the analysis *correctly* detects an effect ($k_M=7$ ms) at $p < .05$. Type I error is the proportion of BATAs for which the analysis (wrongly) detects an effect when $k_M=0$. We show the result for the standard approach (linear models) and a simple correction for skewed RTs (log-linear models). **Study 1—MAIN EFFECTS:** To compare these models for RT distributions of different means, we bootstrapped RTs from the first (Bin 1), middle (Bin 2), or last 5 trials (Bin 3) of ³, either from *raw* ($\mu_1=479, \sigma_1=4790; \mu_2=357, \sigma_2=775; \mu_3=327, \sigma_3=1095$) or *cleaned* data ($\mu_1=418, \sigma_1=211; \mu_2=333, \sigma_2=162; \mu_3=313, \sigma_3=145$). Log-linear models consistently outperformed the standard approach in power, without inflating Type I (Table 1). Although power for linear models increased with faster RTs, their power *deficit* increased as well. **Study 2—INTERACTIONS:** We investigate how a main effect and its interaction are affected by the presence of another much larger effect. We crossed the presence/absence of the (small) main effect from Study 1 with the presence/absence of an interaction with a large effect of Bin. Using BATA sampled from Bin 1 and 3, we assessed power/Type I for the smaller main effect ($\pm k_M$) and its interaction ($\pm k_I$). The log-linear model again had higher power, but the standard approach and the log-linear model traded off in Type I error. Without the interaction ($k_I=0$) and main effect ($k_M=0$), the standard approach inflates the main effect Type I error. When $k_I=0$ and $k_M > 0$, log-linear analyses inflate the interaction's Type I. Both inflations were small (.05-.08). **Conclusion:** With increasing emphasis on replicability, differences in power between approaches have gained importance. We find moderate-to-large power deficits for the standard approach. A simple solution to this problem—log-transforming RTs—avoids this but *can* cause (slightly) inflated Type I errors. We thus also explore more advanced solutions via Bayesian data analysis (gamma, log-normal+shift; ExGaussian). ¹Kliegl et al 2010-VisCog; ²Wagenmakers & Brown 2007-PsyRev; ³Stack et al 2018-Mem&Cog; ⁴Licalalde & Gordon 2018-CUNY

DISENTANGLING INTRA- AND INTER-TALKER VARIABILITY IN L2 PHONETIC PRODUCTION: L2 SPEECH, BUT NOT TALKERS, IS MORE VARIABLE

Xin Xie (University of Rochester) & T. Florian Jaeger (University of Rochester)
xxie13@ur.rochester.edu

It is often assumed that L2 speech is more variable than L1 speech, creating a source of perceptual difficulty.¹ Increased within-talker variability is also of theoretical interest because it would be expected given competition between L1 and L2 representations. However, previous work has confounded intra- and inter-talker variability in L2 speech.² Here we aim to tease apart these two sources of variability. We ask 1) Do L2 speakers exhibit greater *intra-talker variability* than L1 speakers? 2) Is there greater *inter-talker variability* in L2 speech than in L1 speech? We examine the similarities and differences in phonetic production between American English (L1) and Mandarin-accented English (L2).

Data. We present a novel database of voicing and vowel contrasts in American (AE) and Mandarin-accented English (ME) an order of magnitude larger than previous studies. All speakers produce the same set of words so as to control for phonetic context variability. Here, we focus on voicing in word-final stops (*tap* vs. *tab*), known to be affected in ME. Six word-final stops (2 voicing X 3 place of articulation) in CVC monosyllabic English words from 9 AE and 10 ME males resulted in 1519 tokens (76-86 tokens per talker, 5+ tokens per stop). ME talkers varied in L2 (AE) proficiency. We coded three cues important to stop voicing: duration of the preceding vowel, closure and burst. AE primarily uses the first two cues, and ME the third. Both raw and speech-rate corrected durations were analyzed (ME talkers spoke more slowly). We show results for the latter, but either supports our conclusion.

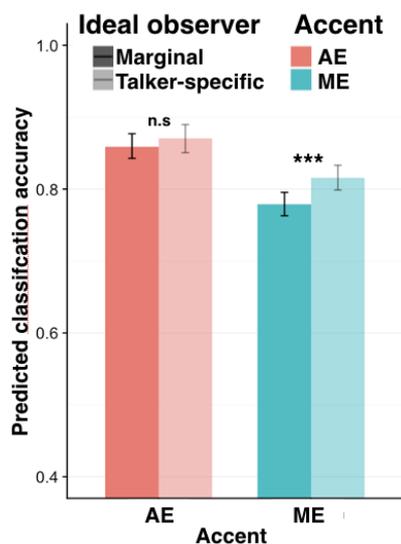
Use of cues. AE and ME speakers used similar cues to signal voicing—longer vowels, shorter closure and shorter bursts. However, the marginal means differed significantly between accents: ME speech had shorter vowels for voiced categories, longer bursts for voiceless categories, longer closure for voiced categories and shorter closure for voiceless categories ($p < .005$) than AE. Consequentially, voicing in ME is more separable (e.g., *d*-prime for /d/ vs. /t/) by burst, but less so by vowel and closure than in AE. In addition, individual ME talkers differ considerably from the accent marginal.

Intra-talker variability. The intra-talker variability (SD) of category-specific distributions did not differ between AE and ME for vowel or burst ($p > .80$). ME speakers were more variable in closure for voiced categories ($p < .01$), but less for voiceless categories ($p < .01$). On balance, voicing in ME showed no more intra-talker variability than AE speech.

Inter-talker variability. To quantify both AE and ME's inter-talker variability, we compare talker-specific ideal observers to ideal observers fit to accents' marginal distribution (both trained for the joint distribution of all three cues on a subset of the data, and assessed on held-out test data). If every speaker's distribution is identical, the talker-specific models are equivalent to the marginal model; as inter-talker variability increases (i.e., individual talker distributions deviate from the marginal), tokens will be categorized more accurately by the talker-specific, rather than marginal, ideal observer. We find a significant *talker-specific-over-marginal* advantage for ME speech ($p < .001$) but not for AE speech ($p = .33$; Fig.1).

Conclusion. L2 speakers are not more variable than L1 speakers in their production of word-final stops. We thus do not find evidence for competition-related increase in intra-talker variability. However, L2 speech does exhibit greater inter-talker variability. This finding also predicts a greater benefit from talker-specific adaptation in L2 speech than in L1 speech.

References: ¹Baese-Baerk, 2015-*JASA*; ²Wade et al., 2007-*JPhon*

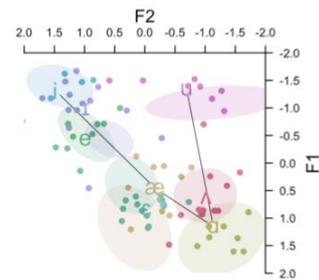


MODELING PRIOR KNOWLEDGE IN THE PERCEPTION OF NATIVE AND FOREIGN-ACCENTED SPEECH

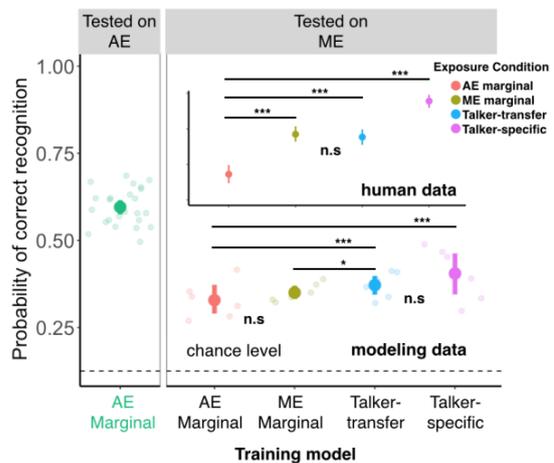
Xin Xie (University of Rochester) & T. Florian Jaeger (University of Rochester)
xxie13@ur.rochester.edu

Native listeners generally understand foreign-accented speech less well than native speech. Exemplar and other experienced-based accounts attribute this to differences in native and non-native productions: when implicit knowledge about the distribution of cues in native productions is used to recognize non-native productions, listeners experience a perceptual deficit. However, formal *models* for foreign-accented speech perception have been lacking. Focusing on vowel recognition, we explore the value of ideal observers as a simple tool with zero degrees of freedom to derive quantitatively testable predictions of experienced-based accounts. We find that ideal observers correctly capture that (1) the perception of foreign-accented speech improves after talker-specific exposure, (2) that some of this improvement can transfer to other talkers with the same accent. Our results provide a formal link between listeners' prior knowledge of cue distributions in natural speech, the actual speech input and the perceptual outcome.

Speech data. We analyzed productions of eight vowel categories from two speech datasets: American English (AE; 24 male talkers) and Mandarin-accented English (ME; 6 male talkers). Each talker produced 5-13 tokens per vowel; F1 and F2 of vowel midpoints are hand-measured and Lobanov-normalized. Fig 1 shows the vowel tokens from a representative ME talker against 95% CI ellipses determined by AE tokens. Notably, the ME productions do not deviate much from the native tokens for some categories (e.g., /u/; present in Mandarin), but are outside the native range for other categories (e.g., /æ/; absent in Mandarin). We randomly divided each talker's data into training (50%) and test (50%).



Models. We fit four different generative models to assess how well an ideal observer recognize ME vowels when conditioned on different prior knowledge: *AE Marginal* model—trained with vowel tokens drawn from six randomly selected AE talkers (blind to talker identity); *ME Marginal* model—the same but for ME talkers; *Talker-transfer* models—trained with one of the six ME talkers other than the ME test talker; *Talker-specific* models—trained with each ME talker's own speech. This held the total amount of training data constant across models. The parameters of each model (means and covariance matrix of bivariate Gaussian over F1, F2 for each model) were estimated from training data. Recognition accuracies were assessed on held-out test tokens, using Luce's choice rule over each model's posterior probabilities.



Results. Paralleling human perception, all models predict that vowels present in Mandarin have overall higher recognition accuracy than vowels absent in Mandarin (not shown). For the former, AE Marginal models yield better accuracy than models trained on ME speech (not shown). For vowels absent in Mandarin, however, performance was better for models trained on ME speech, in line with experienced-based accounts (Fig 2). Performance of the talker-specific and talker-transfer models also paralleled human data.

Conclusion. Ideal observers link cue distribution in production to predictions about perception without any degrees of freedom. They provide a powerful tool to understand basic patterns of accent perception and adaptation. Additional studies predict individual differences in talker intelligibility and transfer.

COMPARING MODELS OF UNSUPERVISED ADAPTATION IN SPEECH PERCEPTION

Shaorong Yan, T. Florian Jaeger (University of Rochester)
 <syar13, fjaeger>@ur.rochester.edu

Talkers differ in their realization of the same sounds. One mechanism that allows listeners overcome this problem is perceptual adaptation to talker-specific differences.^{1,2} Such adaptation can be understood as incremental inference of talker-specific generative models.³ For supervised learning, models within this framework provide a good fit against human responses in perceptual recalibration experiments.³ Everyday speech perception, however, often provides unlabeled input, making adaptation a problem of (partially) unsupervised learning. We develop and compare different models of how listeners might deal with the resulting uncertainty, and compare them against human responses.

Human data. We use data from an experiment⁴, which exposed listeners to /p-/b/ continua embedded in words like “p/bear”. The bimodal cue distribution along the VOT continuum was shifted to a mean VOT of 30ms for /b/s (from ~-0.3ms) and 70ms for /p/s (from 60ms).⁵ Subjects were assigned to one of two groups. In the *unsupervised* group, the input was never labeled. In the *supervised* group, the input was labeled (as /b/ or a /p/) on half the trials. Subjects shifted their categorization boundaries towards the midpoint of the bimodal distribution, as predicted by perceptual adaptation. We estimate incremental changes in listeners’ category boundaries via mixed logistic regression with lapse rate for each moving window of 5 trials (Fig. 1). **Models.** For the *supervised* exposure, we used the fully Bayesian *supervised* model of incremental belief-updating from [3]. For the unsupervised exposure, we consider two unsupervised models that differ in how they treat the uncertainty about the correct categorization of each input. The *posterior* model, performs the optimal (fully Bayesian) inference for *unlabeled* input: it updates *both* categories, weighing the input by its posterior probability under each category. The *posterior-takes-all* model, discards any uncertainty about the input categorization when updating its beliefs. It assumes that each input comes from the most probable category under the current belief, and then only updates beliefs about that category’s distribution. For comparison, we also trained a *supervised* model based on the true labels from the generative model (*not* known by subjects). This model defines the best possible learning. We used priors estimated from production data⁵ and varied the prior strengths. All models were fit to the same input data heard by subjects.

Results. All models move the boundary and slope in the same direction as human learners (Fig 1). Surprisingly, the *supervised* model had a better fit to human behavior for *unsupervised* exposure. Both the *posterior* model and the *posterior-takes-all* model did not adapt as much as the *supervised* model (true for all reasonable prior strengths we explored).

Conclusion. An unsupervised distributional learning model can qualitatively capture the trajectory of human categorization in perceptual recalibration. It does not, however, provide as good a fit to human behavior as a supervised model even for unsupervised exposure. This suggests that human subjects make use of more information than available to the models. In ongoing work, we remove simplifying assumptions (e.g., no covariance between /p/ & /b/ distributions) known to be false⁶ from our models.

REFS ¹Kraljic et al., 2008-*PsySc*; ²Norris et al., 2003-*CogPsy*; ³Kleinschmidt & Jaeger, 2015-*PsyRev*; ⁴Kleinschmidt & Jaeger, 2016-*CogSci*; ⁵Kronrod et al., 2012-*CogSci*; ⁶Chodroff et al., 2015-*JASA*

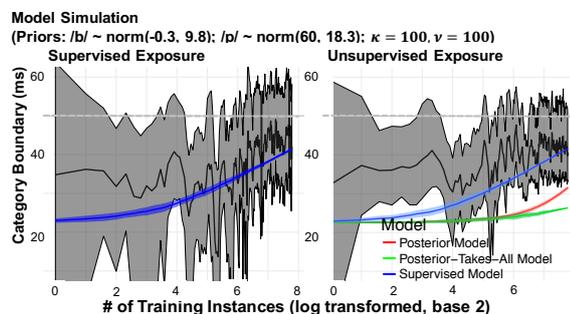


Fig. 1 Incremental changes in category boundary for supervised (left) and unsupervised condition (right) predicted by different learning models (colored) and human behavior (black). Shaded areas are 95% credible intervals. Dotted grey lines show optimal boundary if only input distribution of VOTs is considered.

Saturday Poster.82

INDIVIDUAL DIFFERENCES IN L2 SENTENCE PROCESSING: EFFECTS OF WORKING MEMORY, LANGUAGE EXPERIENCE AND INHIBITORY CONTROL

Sven Hintzen (RWTH Aachen), Elma Kerz (RWTH Aachen), Daniel Wiechmann (University of Amsterdam), Stella Neumann (RWTH Aachen)

Considerable variability has been observed in both native (L1) and non-native (L2) online sentence processing ability (see, e.g., [2] [6] for reviews). In the L1 literature, it has long been established that this variability is strongly associated with individual differences in verbal working memory capacity (vWMC) (e.g., [1] for a meta-analysis). However, prior L2 research on this relationship has been scant and has produced inconclusive results (cf., [4] for an overview). Furthermore, while accumulating evidence suggests that the relationship between vWMC and L1 sentence processing is mediated by language experience [7] [3], no attempt has been made in the L2 literature to determine to what extent vWMC is impacted by L2 experience. Finally, there is an increasing effort to integrate the role of inhibitory control in sentence processing, especially in contexts where the initial interpretation of a sentence is needed to be inhibited in place of a later alternative interpretation. [e.g., 5].

The present study sets out to replicate the results reported in a recent L1 study [3] in a group of L1 German intermediate to advanced L2 learners of English (N=48) and to extend that study by integrating the role of inhibitory control. Using a within-subject design, we could replicate the correlation between vWMC (as gauged by a reading span task) and two proxy measures of L2 experience. In a next step, we determine whether and to what extent online L2 processing ability of garden-path sentences in a self-paced reading task is affected by individual variation in vWMC and inhibitory control, measured by a Stroop color-word task. Mixed effects modeling revealed that while the magnitude of the garden path effect was not affected by vWMC, it was significantly associated with greater inhibitory control as evinced by higher Stroop interference scores. Taken together, these findings suggest that research on L2 sentence processing can benefit from the integration of L2 experience measures and from a stronger focus on inhibitory control as a locus of individual differences.

References:

- [1] Daneman & Merikle (1996). Working memory and language comprehension: A meta-analysis. *Psych. Bul. and Rev.*, 3, 422–433.
- [2] Farmer et al. (2012). Individual differences in sentence processing. In Spivey, McRae, & Joanisse (Eds.), *Cambridge handbook of psycholinguistics* (pp. 353-364). Cambridge: CUP.
- [3] Farmer et al. (2016). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Q. J. Exp. Psychol.*, 70(3),413-433.
- [4] Juffs & Harrington (2011). Aspects of working memory in L2 learning. *Lang. Teach.*, 44(2),137-166.
- [5] Novick et al. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cog., Aff., & Behavl. Neusci.*, 5(3),263-281.
- [6] Roberts (2012). Individual differences in second language sentence processing. *Lang. Learn.* 62(2), 172-188.
- [7] Wells et al. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cog. Psy.*, 58, 250–271.

Saturday Poster.83

English resumptive pronouns do not help the comprehender

Adam Milton Morgan (UC San Diego), Titus von der Malsburg (University of Potsdam),
Victor S. Ferreira (UC San Diego), Eva Wittenberg (UC San Diego)
adam.milton.morgan@gmail.com

In English, a language that primarily uses gaps to form Wh-dependencies (1a), resumptive pronouns (RPs; 1b) are not acceptable. Yet, they are regularly produced, particularly in structures where gaps are unacceptable (*islands*), and they appear to be planned. Why do speakers produce them? Some researchers suggest that RPs facilitate processing, but despite evidence of decreased reading times after RPs vs. gaps (Hofmeister & Norcliffe, 2013) and increased subjective comprehensibility ratings for RPs vs. gaps (Beltrama & Xiang, 2016), no study has tested the hallmark prediction of facilitation: more faithful interpretation. We ask: Are interpretations of RPs as correct as interpretations of gaps?

We designed stimuli devoid of pragmatic cues so that comprehenders had to rely on bottom-up syntactic processing. We manipulated (a) gap/RP and (b) clause type, so that gaps and RPs appeared in non-islands (1; the structures tested in H&N; these are almost exclusively produced with gaps), wh-islands (2; produced with ~50%gaps/RPs), and adjunct islands (3; produced with ~90% RPs; production data from Morgan & Wagers, 2018). In 4 experiments with different paradigms (sentence-picture matching, self-paced reading, visual world, sentence comprehension), participants chose one of four options (4) to answer “Who did what to whom?”. No feedback was given.

All experiments found that relative to gaps, RPs decrease rates of correct responses (4a) to comprehension questions ($p < .05$ in Exps. 1,3,4) and correspondingly increase rates of incorrect—but locally coherent—responses (4b; $p < .05$ in Exps. 1,2,3,4). In **Exp.1** (N=300), a one-shot (i.e., one-item) picture-matching task, participants saw the whole sentence for the duration of the trial and clicked on one of four pictures. In **Exp.2** (N=96), a self-paced reading task with 48 critical items (and 60 fillers designed to prevent the development of parsing strategies), participants answered the comprehension question after the sentence had disappeared. Here, we replicated H&N’s finding that RTs are faster following RPs than gaps, but we argue that this cannot constitute facilitation as previously claimed given that RPs lead to less correct interpretation. In **Exp.3**, we tracked participants’ eyes while they listened to 48 critical items (and 60 fillers) and were shown 4 potential referents for the gap/pronoun (e.g., for the example item in 1-3: dinosaur, rabbit, pig, cat). Preliminary data (N=94/96) are consistent with Exps.1 and 2: multiple-choice responses indicate that RPs are assigned a local (but incorrect) referent more than gaps. Consistent with this, eyetracking data indicate that in real-time comprehension gaps are correctly interpreted more often than not, but RPs lead to chance looking between target (dinosaur) and local (rabbit) interpretations.

Some (e.g., Asudeh, 2004) have suggested that English RPs are not syntactically identical to gaps (whose reference is fixed), but are instead ordinary pronouns. If this is the case, then their relative preference for local referents may follow from a locality preference in ordinary pronouns. In **Exp.4** (N=150), a one-shot sentence comprehension task, we ask how resumptive and ordinary pronouns’ interpretation compare to that of gaps. Contrary to our prediction, ordinary pronouns are interpreted more similarly to gaps than RPs. This suggests that the RP interpretation pattern reflects confusion, with choices approaching chance between the gender-congruent potential referents.

In sum, contrary to the facilitation hypothesis, RPs *hinder* comprehension. Further, RPs are not processed like ordinary pronouns. These results together suggest that RPs confuse comprehenders. Finally, our data reveal a common oversight in processing studies: improved behavioral measures (e.g., faster reading times; increased subjective ratings of comprehensibility) should not be interpreted as facilitation in the absence of interpretation data; they may signal any number of other underlying behaviors such as abandoning the parse and speeding through to end the trial.

1. {a,b} It was Miss Dino who Miss Rabbit *said that* Mr. Piggy tickled {__,her} with a feather.
2. It was Miss Dino who Miss Rabbit *wondered why* Mr. Piggy tickled {__,her} with a feather.
3. It was Miss Dino who Miss Rabbit *snacked while* Mr. Piggy tickled {__,her} with a feather.
4. Who did what to whom? (a) Mr. Piggy tickled Miss Dino. (b) Mr. Piggy tickled Miss Rabbit. (c) Mr. Piggy tickled Miss Cat. (d) Miss Rabbit tickled Mr. Piggy.

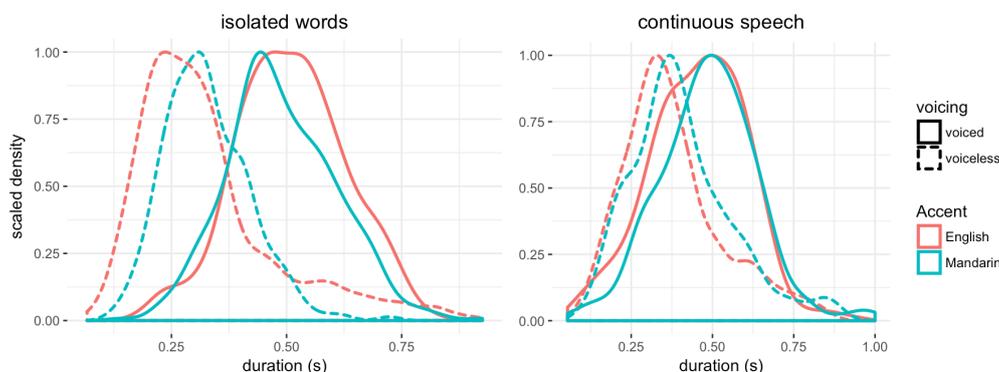
A CORPUS OF NATIVE AND NON-NATIVE SPEECH FOR SPEECH PRODUCTION RESEARCH

Ruolan Li, Xin Xie & T. Florian Jaeger (University of Rochester)
r.li@rochester.edu

To date, controlled comparisons between native- and non-native speech are largely based small samples, partly due to a paucity of non-native speech database. We introduce a corpus of speech recordings from general American English (AE) and Mandarin-accented English (ME). Both isolated and continuous speech is available from all speakers. Our aim is to create an annotated corpus of comparable speech data from a large sample of individual speakers. Here we wish to inform the research community of the availability of such a database. We begin by providing a description of the speakers and the recording materials. Then, we present some preliminary analyses to illustrate the kinds of questions that can be investigated with this corpus.

Speakers and speech materials. The corpus currently contains recordings from 30 speakers, with 15 speakers (10 male and 5 female) in each accent group. All AE speakers speak a Northeastern dialect of American English. All Mandarin-accented speakers are late L2 learners of English who acquired English in mainland China. Speakers read two lists that contained isolated words and continuous speech in a laboratory setting. These lists have been widely used in speech perception studies.^{1,2} The isolated word list contains 180 monosyllabic words sampling the entire English phonetic inventory. Specifically, this word list includes minimal pairs (e.g., tap and tab) that are confusing when read by a ME speaker. The continuous speech list contains 80 sentences, divided equally into 5 sets of phonetically balanced sentences. Each speaker is instructed to read each word (or sentence) three times. Additional recordings are made in case of mispronunciations or disfluencies.

Acoustic analyses. In this report, we present preliminary analyses on the production of word-final stops in AE and ME. We examine how voicing is distinguished by AE and ME speakers, and whether accent-specific patterns are similarly present in both isolated and continuous speech. For each speaker, we annotated and hand-measured 86 stop-final words from isolated words, and 73 words from continuous speech. Our analysis focuses on three durational measures: duration of vowel, closure, and burst. Past research shows that word-final voicing in stops is signaled by longer vowel, shorter closure and shorter burst.³ In particular, while vowel and closure are salient cues for distinguishing voiced tokens from voiceless tokens in AE, such contrast is diminished in ME. Previous work has exclusively focused on isolated speech. Extending prior research, our data suggest that for all three acoustic cues (vowel, closure and burst), the degree of separability of the voicing contrasts is affected by accents (AE vs. ME) and types of speech (isolated vs. continuous). Fig. 1 shows the distribution of vowel duration as an example (see figure). Overall, AE speakers make greater separation between voiced and voiceless stops, compared with ME speakers. This pattern is retained in continuous speech, albeit to a lesser degree. This result lends support for to the validity of generalizing findings from isolated speech to continuous speech.



References: 1. Bench, J., & Bamford, J. (Eds.). (1979). 2. Weil (2001). JASA. 3. Flege et al. (1992). JASA.

THE HANDWRITING OF CHINESE CHARACTERS: A PSYCHOLINGUISTIC DATABASE

Ruiming Wang, Shuting Huang and Zhenguang Cai
(zhenguangcai@gmail.com)

Unlike lexical access in spoken language, relatively little is known about how orthographic codes of a word are accessed in handwriting. Experimental findings often come to different conclusions about the cognitive processes underlying handwriting, for instance, regarding whether phonology is used to access orthographic codes [1,2]. Such disparity is likely due to the use of small samples of hand-picked words which may have uncontrolled confounds. In handwriting research, there have been few large-scale databases that can elucidate on the empirical debates or serve as a tool for stimulus selection for lab-based studies, though databases have become abundant in oral language production (e.g., picture naming). Here we report a large-scale psycholinguistic database of Chinese character handwriting.

A total of 203 university students handwrote 200 target characters randomly selected from a cohort of 1600 characters. They heard a phrase defining a target character (e.g., “辣椒的辣” defines the target character 辣), which they then handwrote and handwrote on a graphic tablet. We recorded (1) *writing latency* (the time between target character offset and writing onset), (2) *writing duration*, and (3) *accuracy*. We also compiled the following character-level characteristics for each character: 1) *character frequency*, 2) *stroke number*, 3) *radical number*, 4) *homophone density*, 5) *phonogram status* (if a character is dictionary-defined as a phonogram), 6) *orthographic regularity rating* (to what extent does a character contains a sound radical), 7) *sound radical writing order*, 8) *radical composition* and 9) *age of acquisition (AoA)*.

Using these lexical characteristics as predictors, we found that (1) more frequent characters have a shorter writing latency (i.e. less time needed for accessing orthographic codes), a shorter writing duration, and higher accuracy; and (2) characters with more strokes have a longer writing latency, a longer writing duration, and lower accuracy. More interestingly, we found an effect of homophone density and orthographic regularity such that (1) characters with more homophonous character neighbours need more time for accessing orthographic codes (i.e. a longer latency) and are less often correctly written, and (2) characters with more phonology-orthographic regularity require less time in accessing orthographic codes (latency) and are more often correctly written. These two findings clearly suggest a role of phonology in orthographic access. We also found AoA effects, with an advantage for early-acquired characters in writing latency, duration and accuracy. Finally, characters with typical radical composition (left-right or top-down) require less time in orthographic access than characters with less typical radical compositions. None of the other lexical variables seem to have an effect on handwriting. In terms of importance, both writing latency and accuracy can be largely explained by character frequency and AoA, while writing duration by stroke number.

These results clearly support the phonology-mediation account of handwriting (i.e. phonology is used for accessing orthographic codes) [1,3]. Also whether a character is dictionary-defined as a phonogram or not does not seem to impact handwriting while orthographic regularity as rated by participants does; this finding highlights the importance of consulting large-scale database instead of dictionaries in designing handwriting stimuli for lab-based studies. The database will be a tool for the investigation of handwriting. In particular, it can be used to study, among other things, factors that determine handwriting difficulty and access of orthographic codes in handwriting.

[1] Qu, Q., Damian, M. F., Zhang, Q., & Zhu, X. (2011). *Psychological Science*, 22, 1107-1112.

[2] Zhang, Q., & Wang, C. (2015). *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.00448.

[3] Luria, A. R. (1970). *Traumatic aphasia*. The Hague: Mouton.

**PERCEPTUAL PRIMING AND SYNTACTIC CHOICE IN ENGLISH
LANGUAGE: MULTIMODAL STUDY**

Mikhail Pokhoday, Yury Shtyrov, Christoph Scheepers and Andriy Myachykov
(mikhail.pokhoday@gmail.com)

In a fully developed production system, perception provides an input of information about the event, attention foregrounds relevant/important information for the conceptual analysis, and subsequent language production mechanisms collaborate to generate speech (Levelt, 1989). A part of this complex process is the necessity to select between simultaneously available syntactic alternatives. For example, English language provides several options that can describe the same visual event, e.g., an officer chasing a burglar. These minimally include (1) The officer is chasing the burglar and (2) The burglar is (being) chased by the officer. These active- and passive-voice alternatives differ in assigning object and subject roles to agent (officer) and patient (burglar). Existing evidence suggests that the system responsible for assigning the grammatical roles is sensitive to the distribution of the speaker's attention within the described scene (Tomlin & Myachykov, 2015, for a recent review). Specifically, a speaker of English is more likely to choose a passive-voice frame when her attention is directed to the patient of the described event and she is more likely to use an active-voice frame when the agent is in her attentional focus (e.g., Myachykov, et al., 2012). While this and other studies indicate a regular interplay between attention and syntactic choice, they also exclusively used variants of the visual cueing paradigm (Posner, 1980). As a result, the reported link between attention and syntactic choice cannot be generalized beyond the visual modality. A more ecologically valid proposal needs to take into account a multi-modal nature of attention.

Here, we report results of a series of sentence production experiments, in which English native speakers described visually presented transitive events (e.g. kick, chase, push). In half of the trials the agent appeared on the left and in the other half – on the right. Speakers' attention to the referents was manipulated by means of lateral bimodal cues. In Experiment 1 the cues were auditory (beep played monaurally) and visual (a red circle); in Experiment 2 – auditory and motor (participants were prompted to press a left or a right key depending on the color of the central fixation cross); in Experiment 3 – visual and motor. Hence, the Cued Referent (Agent/Patient) was crossed with the Cue Type (Auditory/Visual and Auditory/Motor). The proportion of the produced passive-voice sentences was the dependent variable. In Experiment 1, there were no observable effects. In Experiment 2 we registered a main effect of Cued Referent (more passive-voice sentences in Patient-Cue condition: $X^2(1) = 5.29$, $p=.02$). Also, there was a main effect of Cue Type (more passive-voice sentences in Motor-Cue condition: $X^2(1) = 6.56$, $p=.01$). There was no interaction between the two factors suggesting that only one attentional modality at a time can impact syntactic choice. In Experiment 3 there was an effect of visual, but not motor cue: $X^2(1) = 4.08$, $p=.043$. No other effects or interactions were significant. Overall these results suggest an existence of a hierarchy in effects of modality of primes on syntactic choice.

References:

- Gleitman, L.,R., January,D., Nappa,R. and Trueswell, J.C. (2007). On give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544-569.
- Kostov, K. and Janyan, A. (2012). The role of attention in the affordance effect: can we afford to ignore it? *Cognitive Processes*, 13, 215-218.
- Myachykov, A. and Tomlin, R. S. (2014). Attention and salience. In Dabrowska, E., & Divjak, D. (Eds.). (2015). *Handbook of Cognitive Linguistics*(Vol. 39). Walter de Gruyter GmbH & Co KG.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, 1191(1), 182-200.
- Tomlin, R. S. (1995). Focal attention, voice, and word order: an experimental, cross-linguistic study. *Word order in discourse*, 517-554.

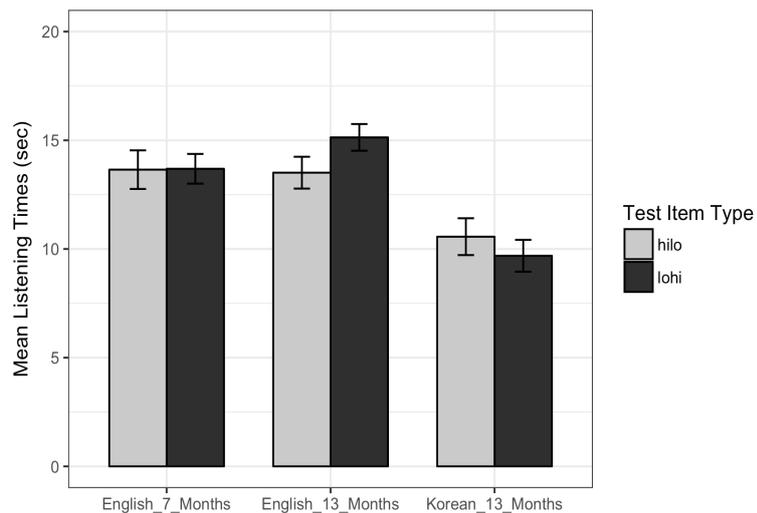
Saturday Poster.87

LANGUAGE-SPECIFIC STATISTICAL COMPUTATIONS IN ADULTS AND INFANTS
Luca Onnis, Erik Thiessen, Soo-Jong Hong and Kyung-Sook Lee

Adults' linguistic background influences their sequential statistical learning of an artificial language characterized by conflicting forward-going and backward-going transitional probabilities [1]. English-speaking adults favor backward-going transitional probabilities, consistent with the head-initial structure of English. Korean-speaking adults favor forward-going transitional probabilities, consistent with the head-final structure of Korean. Using a preferential looking paradigm, the current experiments assessed when infants develop this directional bias. Seven-month-old infants showed no preference for forward-going or backward-going regularities. By 13 months, though, English-learning infants favor backward-going transitional probabilities over forward-going transitional probabilities, consistent with English-speaking adults (Figure 1). Korean-learning infants show a (non-significant) trend in the opposite direction. This suggests that statistical learning may adapt to the predominant syntactic structure of the native language in the first years of life. Such adaptation may facilitate subsequent learning by highlighting statistical structures that are likely to be informative in the native linguistic environment.

Figure 1. Looking times to test stimuli from an artificial language learning characterized by high forward and low backward probabilities (forward-going probabilities, dubbed "hilo" in the legend), versus low forward and high backward probabilities backward-going (dubbed "lohi" in the legend).

Forward-going items are consistent with Korean word sequences ("school to"), while backward-going items are consistent with English word sequences (e.g., "to school"). Results are for 7- and 13-month-old infants in English-exposed and Korean-exposed infants. Error bars indicate standard error.



Reference:

[1] Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, 126(2), 268-284.

[2] Thiessen, E., Onnis, L., Hong, S., Lee, K.S. (2018). Early developing syntactic knowledge influences sequential statistical learning in infancy. Accepted at *Journal of Experimental Child Psychology*.

Saturday Poster.88

ENHANCED IMPLICIT LEARNING IN BILINGUALS

Luca Onnis, Nanyang Technological University

Bilingual experience has been shown to correlate with higher executive functions. Here we asked whether degree of bilingual experience correlated with another core human ability, the implicit extraction of complex patterns of statistical regularities. All languages exhibit statistical structure, but in language-specific ways, and thus bilinguals must learn and juggle different sets of statistical regularities all the time. A recent study [1] showed that degree of bilingualism predicted learning scores in a dual-grammar task, in which participants were exposed to two artificial languages simultaneously. As the artificial grammar was instantiated with multimodal stimuli composed of visual objects coupled with pseudowords, that study could not assess whether the statistical learning (SL) advantage is modality-specific or modality-general. Finding higher auditory SL would be compatible with a near transfer account – on the assumption that the main medium of communication in non-signing speakers is auditory. Alternatively, finding both higher auditory and visual SL would be compatible with a far transfer account. We also wanted to establish whether individual differences in artificial grammar learning could be better explained by other cognitive abilities such as executive functions, rather than the language dominance measure. In two experiments with young adults (n=45 in each experiment), we found that degree of bilingualism assessed via the Bilingual Language Profile (BLP) modulated auditory statistical learning (ASL) but not visual statistical learning (VSL) (mixed-effects models were run in R). In addition, inhibitory control did not contribute any significant variance, suggesting that there may be a genuine heightened ability in learning among bilinguals independent of executive control. Future research should establish the direction of causality of the correlation we found. Bilingual experience may impart higher learning abilities or individual differences may support more balanced bilingualism in the long run.

Reference:

[1] Onnis, L., Chun, W. E., & Lou-Magnuson, M. (2018). Improved statistical learning abilities in adult bilinguals. *Bilingualism: Language and Cognition*, 21(2), 427-433.

Index

A

abad, 24, 250
abashidze, 16, 27, 149, 302
abdel rahman, 11–12, 25, 79, 261
abeille, 25
ackerman, 22, 221
adani, 14, 18, 169
adelt, 14
aguilar, 16, 21, 139, 157, 214, 224
alario, 4, 13, 100, 133, 250, 282
alexiadou, 28, 292, 317
allen, 4, 8, 14, 33, 104, 151, 309
almeida, 4
ambridge, 9, 45, 317
andrews, 27, 300
andringa, 302
apurva, 23, 25, 240–241, 272
arai, 26, 280
arciuli, 8, 14, 108
arehalli, 11, 64
arnold, 67, 152, 187, 208
arnon, 19, 108, 177, 187, 244
arruda de souza, 26
artisini, 21, 205
augurzky, 21, 212

avetisyan, 24, 63, 254
avrutin, 11, 24, 62, 248

B

baayen, 19, 47, 177, 267
bader, 4, 12, 25–26, 84, 199, 240, 266, 285
baguley, 12, 76
balatsou, 26, 282
baltaretu, 27, 301
bartelt, 20, 200
barzy, 15, 124
beck, 13, 102
beese, 18, 161
benjamin, 8, 34
benz, 12, 71
bergmann, 21, 209
besserman, 27, 294
bhatia, 23, 240
bhattasali, 26, 290
bidgood, 21, 219
bilbiie, 22, 231
bisson, 21, 211
black, 15, 119, 124, 170, 196, 202, 249,
278
blagoveschenskii, 15, 18, 129, 160

Index

blagoveshchensky, 19, 174, 190
boila, 20, 203
bordag, 13, 87
bosker, 4, 10–11, 13–14, 58–59, 65, 101
bott, 21, 208, 215, 229, 295
boyce, 10, 22, 228
brand, 8, 31, 91, 180
branigan, 9–10, 12, 16, 20, 25–26, 33, 37,
40, 48, 61, 81, 113, 117, 142,
194–196, 235, 247, 269, 271, 283,
291, 296
breheny, 16, 21, 71, 213
brehm, 230
brian, 27–28, 300, 321
brocher, 13, 91, 166, 294
broersma, 24, 173, 251, 256
brough, 25, 269
brown, 4, 9, 39, 77, 118, 136, 265, 296,
323
brybaert, 4, 10, 20, 85, 173, 198, 200,
204, 307
burattin, 25–26, 273, 284
burchert, 14
burchill, 24, 29, 255, 323
bushong, 12, 15, 24, 29, 77, 134, 255, 323

C

caffarra, 20
cai, 4, 29, 330
calignano, 24, 249
campanelli, 8, 15–16, 21, 122, 138, 206
caplan, 14, 106
carlson, 4, 13, 86, 88, 157, 209
castillo, 20, 196
cecchetto, 25–26, 130, 273, 284
cera, 28, 317
chambers, 20, 27, 51, 149, 201–202, 301
chamorro, 20, 194, 315
chang, 4, 8, 33–35, 74, 113, 117, 194,
220, 256, 322
cholin, 24, 250
chondrogianni, 20, 195
choo, 16, 143
chow, 4, 21, 207
chrabaszcz, 24, 258
christiansen, 8, 14, 35, 70, 108, 145, 183,
262–263

christianson, 4, 22, 78, 232, 247, 280
chung, 15, 125, 135
chuprasova, 19, 185
ciaccio, 22, 227
colcher, 25, 265
contemori, 17, 152
cook, 24, 258
coulson, 10, 50
crocker, 3–4, 8, 10, 16, 18, 30, 39, 57,
136, 170
culbertson, 4, 19, 26, 187, 287
cummins, 4, 20, 71, 191, 295

D

daelemans, 12, 85, 177
damian, 10, 20, 198, 268, 330
de beer, 18, 164
de bot, 26, 279
de koster, 22, 229
de la fuente, 22, 231
de pedis, 8, 13, 98
de-dios-flores, 22
degani, 21, 218
degen, 16, 144, 213, 295
dekydtspotter, 23, 236
dell, 8, 34, 103, 106, 250, 268, 291
dijkhuis, 10, 20, 198
dillon, 27, 44, 193, 225, 281, 300, 310
divjak, 19, 155, 182, 331
dokudan, 28, 319
dolscheid, 20, 197
donati, 26, 284
donnelly, 17, 153
dragoy, 14, 118, 281
drenhaus, 8, 16, 135–136, 225, 241
durliek, 24, 28, 251, 315
durrant, 21, 219
duverger, 25, 274

E

eekhof, 20, 204
eik, 18–19, 168, 176
elizaveta, 19, 174, 190
engelhardt, 4, 21, 210, 257
engelmann, 9, 44–45
ernestus, 14
esaulova, 20, 197

Index

- eskew, jr., 13, 96
evers-vermeul, 12, 72–73
- F**
- faber, 20, 204
fabre, 11, 26, 28, 290, 318
faghiri, 22, 231
fano, 10, 54
farhy, 22, 222
federmeier, 10, 34, 38, 49, 57, 75, 112
felser, 4, 15, 87, 127, 130, 236
ferguson, 4, 15, 124
fernandez, 4, 14, 104, 257
ferreira, 4, 9, 19, 29, 36, 64, 76, 78, 90,
106, 183, 247, 280, 285, 291, 317, 328
fieder, 14, 109
filik, 4, 17, 151, 167
filipovic durdevic, 9
filippova, 19, 174, 190
flecken, 27, 291
foltz, 4, 11, 60
foppolo, 13, 25–26, 100, 273, 275, 284
fosker, 24, 252
frank, 11, 14, 24, 30, 62, 140, 144, 240,
248, 256, 259, 262–263, 313
franz, 25, 159, 266
frau, 21, 205, 230
frederiksen, 19, 175
frenck-mestre, 4, 16, 143
freywald, 19, 180
friederici, 18, 137, 161–162, 169, 232, 270
friedrich, 14, 115, 209
frisson, 10, 53
fritzsche, 16, 20, 141, 203
frost, 8, 21, 25, 31, 62, 133, 219, 248,
262–263
fuchs, 13, 95
fuellenbach, 11, 66
fuhrhop, 14, 114
fujita, 12, 69
fukumura, 13, 100, 152, 208
- G**
- galkina, 9, 41
gambi, 11, 14, 25, 66, 112, 133, 178, 269
gao, 20, 27, 204, 302
garcia, 8, 12, 15, 74, 231
garcia-castro, 15
gardner, 20, 23, 181, 195, 239
garrido rodriguez, 39
gaskell, 18, 75, 171
gelman, 15, 44, 123
generalova, 14, 105
gennari, 18, 35, 81, 171, 176, 286
gerwien, 8, 22, 27, 37, 220, 291
ghio, 16, 143
giannopoulou, 24
gilbert, 23, 236, 299
ginzburg, 28, 231, 312
giora, 17, 137, 151
gnedykh, 15, 18, 129, 160
goldberg, 87, 298
gor, 24, 185, 258, 304
grainger, 11, 24, 253
granlund, 9, 45
grant, 4, 10, 51, 129, 160, 174, 190, 304
grillo, 16, 21, 28, 139, 214, 317
grossman, 19, 187
gruber, 16, 147
guerra, 4, 11, 25, 260
gunter, 16, 50, 137, 270
- H**
- haendler, 14
hagoort, 136, 270
hale, 11, 26, 28, 30, 138, 206, 228, 241,
290, 313, 318
hall, 15, 132, 264, 277, 307
hameau, 14, 109
hanne, 18, 164
hanulikova, 4
harbusch, 15, 119
harris, 28, 281, 294, 310
hartshorne, 27, 175, 298–299
haser, 26, 277
hashimoto, 12, 69
hatzidaki, 25, 270
heinat, 9, 18, 167
heller, 4, 10, 51
helo, 11, 25, 260
hemforth, 4, 8–9, 15, 19, 21, 120, 181,
217, 274
hendriks, 16, 22, 229
hendrix, 19, 47, 177, 186, 188

Index

- herve, 13
hesse, 12, 71
hinterwimmer, 18, 25, 165–166, 264
hintz, 10, 16, 20, 148, 198
hintzen, 29, 327
hirose, 16, 145
hobbelink, 27, 293
hong, 82, 247, 332
hopkins, 16, 64, 92, 142, 226
huang, 9, 23, 29, 74, 107, 203, 223, 237, 291, 330
huettig, 9, 39, 54, 98–99, 133, 210
husain, 17, 23, 25, 240–241, 272
husband, 4, 11, 66, 169, 254
husky, 13, 28, 96, 320
huttenlauch, 18, 164
hwang, 24, 135, 247
- I**
- iao, 8, 32
isbilen, 8, 14, 108
ito, 4, 11, 15, 28, 66, 133, 320
ivanova, 17, 26, 152, 194, 288
iverson, 23, 236
- J**
- jachmann, 8, 16, 136
jacobs, 67, 134
jaeger, 12, 15, 24, 29, 33, 44, 46, 77, 106, 113, 134, 144, 187, 255, 271, 304, 313, 323–326, 329
jahn, 15, 122
jennings, 27, 298–299
jescheniak, 25, 268
jessop, 8, 35
jin, 21, 207, 220
johns, 15, 64, 92, 122, 226
jones, 4, 8, 32
jongman, 8, 10, 13, 16, 20, 98–99, 148, 198
jorschick, 24, 250
justice, 8, 32
- K**
- kaan, 15, 54, 121, 149, 163, 230, 281
kaiser, 15, 23, 27, 84, 102, 131, 224, 243, 294
kamide, 4, 10, 51, 55, 143–144, 214, 271
kantola, 14, 117
kaps, 28, 310
kaufeld, 10, 58
kempen, 15, 119
kentner, 25, 164, 266
kerz, 29, 327
keshev, 20, 192–193
kessler, 14, 115, 140
kholodova, 8, 33
khudyakova, 14, 118
kidd, 8, 14, 17, 108, 153
kim, 4, 11, 16, 20, 24, 27, 57, 66, 97, 143, 149, 178, 194, 227, 247, 275, 302, 307
king, 28, 37, 163, 176, 321
kireev, 27, 304
kirsanov, 19, 174, 190
kizilkaya, 18, 159
kleinman, 10, 49
klepousniotou, 12, 75
klingsvall, 9, 18, 167
knoferle, 1, 3–4, 11, 21–22, 26, 209, 233, 239, 278
knowles, 10, 53
koenig, 4, 11, 28, 294, 316
koh, 16, 143
kolak, 9, 45
konieczny, 4, 21, 26, 217, 272, 274, 277, 292
konina, 12, 70
konopka, 8, 21, 36, 87, 216
konrad, 26, 284
konuk, 13, 97
korotkov, 27, 304
kostromina, 15, 18, 129, 160
kreiner, 21, 218
kreysa, 4, 21, 136, 209
krott, 10, 53
kuhlen, 12, 79, 94
kuijpers, 20, 204
kukona, 10, 21, 55, 211
kunikoshi, 28, 320
kurmakaeva, 15, 18, 129, 160
kurtz, 25, 268
kurumada, 23, 239
kush, 15, 18–19, 122, 130, 168, 176

Index

L

lago, 15, 22, 24, 130, 230, 254
lai, 4, 8, 32, 57, 220, 262
laka, 15, 25, 128, 270
lampe, 14, 109
landi, 8, 16, 138
lassotta, 14
lau, 18, 90, 163, 199, 230, 233
laurinavichyute, 14, 18, 116, 118, 172
lawn, 28, 310
lawrence, 23, 238–239
lawyer, 25, 120, 151, 157, 257, 267
lee, 12, 23, 82, 86, 135, 243, 332
lemke, 22, 28, 223, 313–314
lengeris, 21, 211
lentz, 11, 63
leone, 15, 121
lev-ari, 12, 83
levinson, 9, 39, 215, 293
levy, 10, 22, 30, 64, 123, 128, 145–146,
156, 168, 206, 213, 224, 228, 241,
255, 259–260, 272, 280, 292, 313–314,
319
lewendon, 11, 60
lewis, 8, 13, 15, 21, 44, 98, 122, 172, 199,
206, 215, 225, 236, 241, 272, 292, 308
li, 11, 22, 24, 28–29, 73, 220, 247, 318,
329
lieven, 9, 45
lindenberger, 18, 161
lindsay, 4, 10, 16, 48, 142
linzen, 13, 22, 47, 92, 226
liu, 15, 27, 121, 156, 292
logacev, 28, 319
lopukhina, 14, 116
lorenz, 11, 25, 261
lorusso, 25, 270
lou-magnuson, 333
lu, 14, 18–19, 107, 146, 179
luh, 11, 28, 318

M

maciejewski, 12, 75
maddeaux, 10, 51
madsen ii, 24, 257
magomedova, 19, 185
maier, 28, 309

mak, 12, 20, 72–73, 111, 176, 204
manawamma, 26, 279
mancini, 20
marantz, 27, 47, 305
marecka, 24, 251–252
martin, 4, 10–12, 14, 16, 47, 54, 58, 66,
76, 79, 81, 112, 149, 224, 264, 270
martohardjono, 24, 257
marton, 21, 206
mascelloni, 16, 147
maslowski, 11, 65
mauner, 16, 145, 294
mauranen, 12, 70
mayberry, 19, 175
mazza, 24, 249
mcdonough, 16, 27, 149, 302
mckenzie, 13, 90
mcqueen, 10, 16, 20, 60, 148, 198
medimorec, 19, 182
medvedev, 27, 304
meltzer-asscher, 16, 20, 140, 192–193
meng, 13, 26, 74, 87, 285
menon, 25, 265
mertzen, 9, 15, 44, 123
meuser, 25, 264
meyer, 10–11, 14, 16, 18, 20, 36, 65, 79,
136, 148, 161, 198, 216
milin, 9, 19, 47, 182
miller, 23, 27, 108, 236, 303
minematsu, 28, 320
mkrytychian, 15, 18, 129, 160
mohd taha, 21, 210
molimpakis, 26, 276
momsen, 10, 50
monaghan, 4, 8, 21, 25, 31, 62, 219, 248,
262–263
mooshammer, 19, 184
moradlou, 28, 312
morand, 27, 296
morgan, 9, 19, 23, 29, 137, 183, 224, 239,
313, 328
mosca, 26, 279
mueller, 16, 26, 99, 147, 161, 277
muenz, 21
mundhenk, 14, 114
murali, 11, 62
musola, 21, 205
myachykov, 8–9, 14, 17, 29, 41, 110, 155,

Index

331
myers, 11, 67

N

naumann, 10
ness, 16, 61, 140, 285
neumann, 29, 327
nguyen, 22, 178, 233, 320
nicenboim, 11, 25, 260
nickels, 14, 79, 109, 268
nikiforova, 19, 174, 190
nishikawa, 28, 320
nixon, 8, 11–12, 63, 80
norcliffe, 9, 39, 328
nunnemann, 21, 209

O

obrig, 18, 162
oganian, 15, 126
onnis, 332–333
opacki, 24, 251
opitz, 13, 87
oppenheim, 26, 268, 282
ostrand, 11, 61
ots, 21, 216
otwinowska-kasztelanica, 18, 24
ovadia, 20, 192

P

paape, 13, 89
paillereau, 24, 246
palmer, 27, 160, 298
paolazzi, 28, 317
park, 15, 125, 135, 270
paspali, 27, 306
patil, 18, 25, 165–166, 264
paulus, 24, 249
pavani, 21, 205
payne, 161
pearlmutter, 4, 13, 96
penke, 20, 197
peter, 8, 19, 21, 33, 73, 76, 120, 177, 186,
188, 208, 219, 266, 291
phillips, 24, 28, 136, 151, 168, 192–193,
199, 225, 230, 236, 257, 267, 281, 322
pickering, 11–12, 14, 33–34, 36, 39–40,

53, 61, 66, 81, 112–113, 117, 127,
133, 142, 178–179, 194, 215, 235,
247, 271, 274, 283, 291, 296, 309

pine, 9, 45, 74
pinto, 28, 311
pokhoday, 8, 17, 29, 155, 331
polter, 13, 87
poppels, 10, 22, 64, 228
portele, 12, 22, 84, 235
potter, 13, 86, 253, 309
pozniak, 8, 15, 120
prasad, 221
prokopenya, 22, 234

R

rabagliati, 4, 9–10, 14, 25–26, 40, 48, 112,
269, 283
rabovsky, 9, 46
radulescu, 11, 24, 62, 248
ralph-nearman, 17, 151
ramscar, 23, 47, 80, 177, 186, 222, 244,
303
rasenberg, 9, 38
ravenschlag, 10
rawlins, 28, 308
regel, 11, 16, 25, 137, 261
reich, 22, 28, 89, 91, 223, 313–314
reinisch, 10, 13, 58–60, 101
remiszewski, 28, 315
riemenschneider, 19, 184
roberts, 15–16, 26, 127, 139, 166, 281, 327
rodd, 14, 75, 103
roeser, 8, 12, 32, 74, 76
roland, 4, 16, 145, 176
rommers, 8–9, 34, 38, 112, 115
ronai, 27, 295
ros, 15, 128
rowland, 8, 21, 33, 74, 219
ruigendijk, 14, 114

S

saenko, 22, 234
saldana, 26, 287
sanders, 12, 72–73
sandra, 4, 12–13, 18, 85, 100, 164
santestaban, 4, 25
santi, 16, 26, 28, 139, 276, 317–318

Index

- saryazdi, 20, 201–202
scanzi, 26, 275
scarinci, 20
schaadt, 18, 162
schauffler, 21, 212
scheepers, 4, 8–9, 14, 17, 29, 33, 39, 41, 104, 117, 168, 247, 331
scheibel, 26, 286
scherbakova, 19
schlechtweg, 13, 94
schliewe, 4, 22, 233
schlueter, 20, 191
schriefers, 25, 98, 176, 184, 268
schroeder, 10, 20, 198, 200
schroeter, 20, 200
schulte, 19, 180
schumacher, 4, 10, 53, 97, 212
schuster, 16, 144
schwartz, 24, 257, 268
schwarz, 4, 26, 157, 275
seidel, 13, 93
sekerina, 14, 118
sekicki, 10, 52
shin, 24, 247
shtyrov, 8–9, 14–15, 17–19, 29, 41, 110, 129, 155, 160, 174, 190, 331
siew, 23, 238
sikos, 4, 10, 18, 57, 170
silva, 18, 173
sjerps, 10, 13, 59, 101
skarnitzl, 24, 246
slevc, 90
slioussar, 14, 19, 27, 105, 185, 304, 306
smith, 19–20, 31, 74, 142, 187, 196, 259
smithson, 8, 14, 108
smolka, 21, 218
snedeker, 9, 23, 74, 175, 237, 271, 275
snell, 11, 24, 253
solstad, 21, 208
song, 12, 15, 82, 135
sorace, 20, 152, 191, 315
spalek, 15, 19, 126, 184
spenader, 22, 229
spinelli, 23, 242
spit, 78
sprenger, 28, 307
st. pierre, 11, 28, 316
stadie, 14
staub, 4, 26–27, 127, 145, 176, 210, 214, 254, 284, 300
staudte, 4, 8, 10, 16, 52, 136
stefan, 8, 18, 24–25, 36, 165–166, 256, 259, 264
stoltmann, 13, 95
suckow, 20, 97, 199
suffill, 12, 81
sun, 16, 19, 21, 177, 213, 261
swanson, 23, 236
szewczyk, 17, 24, 28, 98, 154, 251–252, 315
- T**
- tal, 13, 16, 19, 22, 92, 140, 187, 226
talmina, 28, 308
tan, 204
teixeira, 26, 289
ten bosch, 14
theakston, 9, 45
thiberge, 9, 19, 181
thierry, 11, 26, 60, 79, 207, 282
thiessen, 332
thompson, 12, 14, 24, 82, 107, 118, 151, 259, 282, 318
tian, 21, 28, 213, 312
tobar, 9, 26, 40, 283
torrance, 12, 76
tourtouri, 18, 170
trofimovich, 16, 27, 149, 302
trotter, 25, 262–263
tsiola, 22, 232
tsoukala, 24, 256
tulkens, 12, 85
turco, 16, 139
tutunjian, 19, 179
- U**
- ueda, 12, 69
uygun, 27, 297
- V**
- vaidya, 17, 156
van bergen, 38
van den bosch, 24, 177, 256
van den hoven, 20, 204

Index

van dyke, 8–9, 15–16, 21, 44, 122, 138,
206, 321
van gompel, 13–14, 84, 90, 100, 113, 117,
122, 168, 208, 221
van hout, 27, 293
van rij, 11, 28, 63, 307, 318
van schijndel, 13, 22, 92, 226
van t' hoff, 10, 20
vanek, 26, 281
vasishth, 4, 9, 13, 15, 24, 44, 89, 120, 123,
156, 163, 199, 206, 225, 230,
240–241, 254, 272, 292
vergilova, 10, 57
verhagen, 26, 281
vespignani, 16, 21, 24, 102, 147, 205, 249
vieira, 26, 289
vihman, 45
villata, 13, 100, 192, 306
vogelzang, 14, 114
von der malsburg, 4, 10, 18, 22, 29, 172,
228, 230, 328
von heusinger, 13, 17, 91, 97, 157
vorwerg, 27, 296
vujovic, 23, 27, 244, 303

W

waegemaekers, 12, 82
wakeford, 14, 117
walther, 17, 154
wang, 18, 22, 29, 81, 129, 160, 171, 205,
220, 227, 252, 330
wartenburger, 18, 164
watson, 4, 11, 67, 86, 139
weber, 13, 69, 71, 102
weeber, 17, 157
wei, 12, 72–73
welby, 23, 242
wen, 11, 24, 28, 253, 318
winkle-bergner, 18, 161
wiechmann, 29, 327
wijnen, 11, 24, 62, 248
wiklund, 19, 179
willems, 20, 70, 204
williams, 15, 124, 176, 213, 294
wittenberg, 4, 11, 22, 29, 64, 224, 328
wodniecka, 17, 24, 28, 154, 251–252, 315
wolna, 17, 28, 154, 315

wolpert, 20
wonnacott, 23, 27, 244, 303
wray, 27, 305
wu, 8, 37, 50, 60, 207

X

xiang, 4, 27, 107, 135, 225, 292, 295, 328
xie, 29, 324–325, 329
xiong, 23, 236

Y

yadav, 17, 25
yamauchi, 28, 320
yan, 29, 46, 326
yoshida, 14–15, 26, 107, 132, 276
yuen, 21, 210

Z

zamparelli, 16, 147
zawiszewski, 15, 25, 128, 270
zehr, 26, 275
zheng, 28, 74, 246, 312
zhou, 13, 96
zuckerman, 28, 311
zwitserlood, 11, 25, 261
zygis, 13, 95