ESLP 2011

Aug 25-27 2011

Center for Interdisciplinary Research,
Bielefeld University, Germany

Embodied and Situated Language Processing 2011
— August 25-27, 2011 —

The 4th Workshop on Embodied and Situated Language Processing

ZiF, Center for Interdisciplinary Research

Bielefeld University, Germany

Table of Contents

Acknowledgements 4

Organizing Committee 4

Reviewer Panel 4

Keynote Speakers 4

Workshop Dinner 4

Program Overview: Thursday, Aug 25 – Saturday Aug 27 5

List of Poster Presentations 7

Abstracts for Oral Presentations on Thursday Aug 25 8

Abstracts for Oral Presentations on Friday Aug 26 16

Abstracts for Oral Presentations on Saturday Aug 27 26

Abstracts for Poster Presentations 29

Acknowledgements

ESLP 2011 is supported by the **Center for interdisciplinary Research (Zif)**, the **Cognitive Interaction Technology Excellence Center**, and **CRC 673 "Alignment in Communication"**, all at Bielefeld University, Germany.

Organizing committee

ESLP 2011 was organized by Michele Burigo, Maria Nella Carminati, Helene Kreysa, & Pia Knoeferle (chair)

Reviewer Panel

Jens Apel, Ray Becker, Ben Bergen, Michele Burigo. Louise Connell, Seana Coulson, Matthew Crocker, Zachary Estes, Francesco Foroni, Stefan Frank, Elsi Kaiser, Yuki Kamide, Michael Kaschak, David Kemmerer, Johanna Kissler, Judith Koehne, Helene Kreysa, Anna Kuhlen, Dermot Lynott, Teenie Matlock, Kate Messenger, Lotte Meteyard, Andriy Myachykov, Rolf Zwaan, Shirley-Ann Rueschemeyer, Christoph Scheepers, Claudia Scorolli, Thomas Scott-Philipps, Michael Spivey, Maria Staudte, Larry Taylor, Marco Tettamanti, Roel Willems, Bodo Winter

Keynote speakers

Seana Coulson (UC San Diego, USA), Beatrice de Gelder (Tilburg University, Netherlands), Silvia Gennari (University of York, UK), Peter Hagoort (MPI for Psycholinguistics), Olaf Hauk (MRC Cognition and Brain Sciences Unit, UK), Marta Kutas (UC San Diego, USA), Michael Spivey (UC Merced, USA)

Workshop dinner

The workshop dinner will take place at Sparrenburg Castle, Bielefeld on Friday, August 26, 2011. There will be bus transport from the conference venue to the Sparrenburg. Further announcements will follow.

Universität Bielefeld
Zentrum für interdisziplinäre Forschung
Center for Interdisciplinary Research
ZiF

Embodied & Situated Language Processing 2011

AUGUST 25 - 27, 2011

Program Overview

Thursday, 25.08.2011

12:50 Opening remarks

13:00 Kutas: What electrophysiology has told us about incrementality, prediction, and semantic memory representations during sentence comprehension

14:00 Coffee break + snacks

14:30 Drenhaus, Staudte, & Crocker: The influence of referential gaze on scene-sentence integration: An ERP-study

15:00 Hald, Hocking, Marshall, Vernon, & Garnham: Modality switching and negation: ERP evidence for modality-specific simulations during negation processing

15:30 Coffee break

16:00 Spivey: The Embodiment of Language is the natural result of cascaded interactive dynamics

17:00 Nam, Bergmann, Waltinger, Kopp, Wachsmuth, & Zhang: Deciphering the Communicative Code in Speech and Gesture Dialogues by Autoencoding Hypernetworks

17:30 Burigo & Knoeferle: (Beyond) referential mechanisms in spatial language comprehension

Friday, 26.08.2011

9:00 Hauk: Spatio-temporal patterns of brain activation reflect embodied language processing

10:00 Coffee

10:30 Yee, Chrysikou, Hoffman, & Thompson-Schill: Playing patty-cake interferes with comprehending the names of manually experienced objects

11:00 Connell & Lynott: Beyond abstractness-concreteness and imageability ratings: A new measure of the perceptual experience of a concept that better predicts reading times

11:30 POSTER INTRO

12:00 LUNCH + POSTERS

13:30 Hagoort: Language processing: a disembodied perspective

14:30 Vandeberg & Zwaan: Syntactic Constraints on Embodied Language Processing

15:00 Franzmeier & Ferstl: Introducing combined TMS and eyetracking into neurolinguistic research

15:30 Coffee break

16:00 Anderson, Matlock, & Spivey: The future of grammatical aspect: The impact of aspect and tense on motor output

16:30 Aguirre Fernández de Lara, Soler Vilageliu, & Santiago de Torres: Do past and future potential events activate the mental time line?

17:00 Gennari: On the diverse nature of time in language comprehension

19:00 Workshop dinner (Downtown)

Saturday, 27.08.2011

9:00 Coulson: Synesthesia

10:00 Lachmair, de Filippis, Dudschig, de la Vega, & Kaup: When upwords meet down-sentences: New evidence for compatibility effects

10:30 Coffee break

11:00 Rummer & Schweppe: Emotion and speech: Evidence for the articulatory feedback hypothesis

11:30 de Gelder: Emotional Body

12:30 *Closing remarks*

Poster Presentations

1	Joao Queiroz, Pedro Ata	Cognition and semiotic artifacts – situatedness and embodiment in Peircean cognitive semiotics.
2	Helene Kreysa, Pia Knoeferle	Differences in comprehension task affect the processing of a speaker's gaze direction.
3	Claudia Scorolli, Pierre Olivier Jacquet, Ferdinand Binkofski, Roberto Nicoletti, Alessia Tessari, & Anna M. Borghi	Does catching a concept differ from catching a flower?
4	Olga Dragoy, Svetlana Malyutina, Maria Ivanova, Thomas Meindl, Aleksey Petrushevsky, & Evgeny Gutyrchik	Embodied representation of tools and abstract verbs: less action shows more sensorimotor ground
5	Ernesto Guerra & Pia Knoeferle	Visual context effects on abstract sentence comprehension: evidence from eye tracking
6	Uta Sassenberg, Ingo Helmich, & Hedda Lausberg	Hand movements as an indicator of implicit processing of emotions in alexithymia.
7	Louise Connell, Zhenguang Cai, & Judith Holler	Hitting the high notes: Gesture affects pitch discrimination via spatial grounding.
8	Sabine Weiss, Horst M. Müller, & Gert Rickheit	Large-scale synchronization during concrete and abstract language.
9	Thies Pfeiffer	Measuring speech-accompanied pointing gestures to identify the interaction between the precision of pointing and multimodal compensation strategies
10	Christopher Baumgärtner	Merging models of natural language parsing and visual attention.
11	Simone Alex-Ruf, Claudia Maienborn, & Rolf Ulrich	Past and future on the timeline: Grammatically and lexically induced temporal reference of sentences.
12	Dato Abashidze, Pia Knoeferle, Maria Nella Carminati, & Kai Essig	Recent real-world versus future events in the comprehension of referentially ambiguous sentences.
13	Raymond Becker, Pia Knoeferle, & Rolf Zwaan	The role of working memory in understanding temporal order statements.
14	Dermott Lynott, Louise Connell, & Felix Dreyer	Object size judgements are affected by tactile and proprioceptive stimulation

15	Seijin Yoo	Multiple speech codes for imitative learning from pseudowords to words
16	Byoung-Tak Zhang, Eun-Seok Lee, & Min-Oh Heo	Modeling situated language learning in early childhood via hypernetworks
17	Choong-Yeon Lee, Euon-Sol Kim, Joon-Shik Kim, & Byoung-Tak Zhang	Interaction of Language and Vision Memories in TV Drama Watching: An EEG Study

Abstracts for Oral Presentations

THURSDAY, AUGUST 25, 2011

Invited talk

What electrophysiology has told us about incrementality, prediction, and semantic representations during sentence comprehension

Marta Kutas
UC SAN DIEGO, USA

Event related brain potentials (ERPs) are an excellent measure of brain sensitivities during online sentence processing. They have provided insight into the nature of incremental interpretation of written sentences, which they (among other online measures) have revealed to be inherently predictive, at times less than fully incremental, and routinely subject to not only the contents but the functional organization (e.g., categories, events) of semantic memory. I will describe the logic and particulars of some of our electrophysiological studies that have led to these conclusions.

The influence of referential gaze on scene-sentence integration: an ERP study

Heiner Drenhaus, Maria Staudte, Matthew W. Crocker
SAARLAND UNIVERSITY, GERMANY

Eye-tracking studies have shown that cues such as *gaze* influence both listeners' visual attention and sentence understanding (Staudte *et al,* 2011). We present findings from an ERP study that investigated the time course of the underlying mechanisms involved in the integration of *referential* gaze and speech.

Using similar stimuli to Staudte *et al* (2011), 18 German participants saw videos of a virtual agent gazing at objects in a scene (1). After the video was presented, participants heard a computer-generated utterance describing the scene: The description was either congruent (2) or incongruent (3) with the linear order of the gaze cues shown in the video (actual stimuli were in German).

(1)

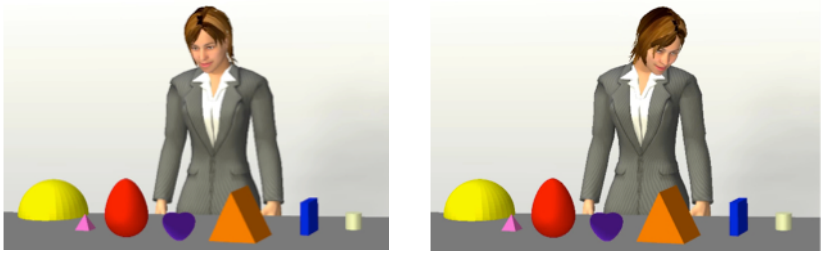

Figure 1. Agent gazes to the egg and the box.

(2) Congruent : [The egg] [is taller than] [the box], I believe.
(3) Incongruent : [The box] [is shorter than] [the egg], I believe.

After hearing the sentence participants had to determine its validity by pressing a button accordingly. Our analysis examined three time windows as indicated by squared brackets in the examples (NP1, predicate, and NP2).

On NP1, our analysis revealed (a) a globally distributed sustained negativity (350-800ms) for incongruent trials (e.g., *the box*) compared to congruent conditions (e.g., *the egg)*. For the predicate (*shorter/taller than*) we observed (b) a frontal positivity between 600-800ms. Finally, analysis of NP2 revealed (c) a centro parietal negativity (250-400ms) for incongruent trials (e.g., *the egg*) versus their congruent counterparts (e.g., *the box*). We provisionally interpret these findings as follows:

(a) Sustained negativity: This negativity likely indexes the maintenance of information in working memory (e.g., Donaldson & Rugg, 1999). The processer must maintain two different situation models in memory, namely the gazed-at *ball* and the mentioned *pyramid* as two possible referents (see also Coulson & Lovett, 2004).

(b) Frontal positivity (600-800ms): The frontal positivity may reflect retrieval of a specific mental representation in order to build a coherent mental model (e.g., Kaan & Schwaab, 2003). That is, the incongruent predicate forces the parser to update the mental representation.

(c) Centro-parietal negativity (250-400ms): This effect can be understood as an attempt of the parser to integrate the final incongruent NP (e.g. Kutas & Hillyard, 1984; van Berkum et al., 1999) into the recently updated mental model for the sentence.

More generally, our results suggest that external language information like gaze has an incremental influence on comprehension mechanisms. Specifically, we have shown that gaze alone can establish a referent, as suggested by increased memory load on the first incongruent NP. Additionally, we have shown an effect of referential gaze on building and retrieving a mental model (second chunk), and finally, on the integration of the second NP into the current mental representation (third chunk).

References:
Coulson & Lovett (2004). *Cognitive Brain Research, 19*, 275–288.
Donaldson & Rugg (1999). *Cognitive Brain Research, 8*, 1–16.
Kaan & Schwaab (2003). *JOCN*, *15*, 98–110.
Kutas & Hillyard (1984). *Nature*, *307*(5947),161–163.
Staudte et al (2011). *Proceedings of CogSic*.
van Berkum et al (1999). *JOCN*, *11*(6),657–671.

Modality switching and negation: ERP evidence for modality-specific simulations during negation processing

Lea Hald (1), Ian Hocking (1), Julie-Ann Marshall (1), David Vernon (1), Alan Garnham (2)
(1) CANTERBURY, CHRIST CHURCH UNIVERSITY; (2) UNIVERSITY OF SUSSEX, UK

The Perceptual Symbol Systems Theory (*1*) suggests that modality-specific simulations underlie representation of concepts. Supporting evidence comes from modality switch costs: Participants are slower to verify a property in one modality (e.g., auditory, BLENDER-loud) after verifying a property in a different modality (e.g., gustatory, CRANBERRIES-tart) compared to the same modality (e.g., LEAVES-rustling (*2*)). Modality switch costs have also been shown to lead to a modulation of the N400 effect in event related potentials (*3, 4*) using different tasks.

In a separate line of research, ERP studies have indicated that without a discourse context, negated sentences are more difficult to process. Negative sentences show larger N400s for true, semantically coherent sentences than for false, semantically incorrect sentences (*5*).

The goal of the current study was twofold. In a modality switch paradigm do negative sentences show a similar pattern in the ERPs as has been seen for affirmative sentences? Secondly, evidence suggests that comprehenders create a simulation of negative sentences (*6*). Can the processing of negation be aided by modality matching information?

Using a within-subjects design 160 experimental sentence pairs were either mismatched or matched on modality (drawn from (*2, 4*).

True-Modality match	*A giraffe is spotted.*	*Rice isn't black.*
True-Modality mismatch	*A light bulb is very hot.*	*Rice isn't black.*

Additionally, we explored veracity by making half of the experimental target sentences false ("Rice isn't white"). Participants judged whether each sentence was typically true or false. ERPs are aligned to the onset of the underlined word.

Our initial results indicate a modulation in the ERP based on modality and veracity. The effect of modality was as follows: For true statements ("Rice isn't black") a larger N400-like effect was seen in the modality mismatch condition compared to the match condition (Figure 1), but for false statements ("Rice isn't white") this effect was reversed (Figure 2).

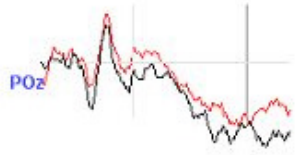

Figure 1. True sentences matched (black) versus mismatched (red), negative plotted up in all figures.

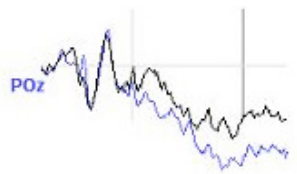

Figure 2. False sentences matched (black) versus mismatched (blue).

The effect of veracity: The modality mismatch condition elicited a large N400 for true sentences (Figure 3). However for the match condition no difference is seen between true and false sentences (Figure 4).

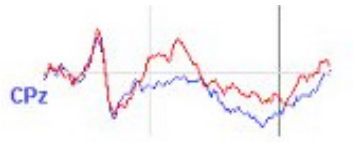

Figure 3. Mismatch condition, true (red) versus false (blue).

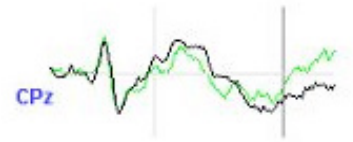

Figure 4. Match condition, true (black) versus false (green).

The evidence suggests that modality-specific simulations occur during language processing and that they may even aid the processing of negation.

References:

Barsalou, L.W. (1999). Perceptual symbol systems. Behav Brain Sci, 22, 577-609.

Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying properties from different modalities for concepts produces switching costs. PsychSci, 14, 119-124.

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click. Front Psychology, 2:10.

Hald, L., Marshall, J., Janssen, D., & Garnham, A. (2011). Switching modalities in a sentence verification task: ERP evidence for embodied language processing. Front Psychology, 2:45.

Fischler, I., Bloom, P., Childers, D., Roucos, S. & Perry, N. (1983). Brain potentials related to stages of sentence verification. Psychophysiology, 20, 400-409.

Kaup, B., Yaxley, R., Madden, C.J., Zwaan, R., & Lüdtke, J. (2007). Experiential simulations of negated text information. Q J Exp Psychol, 60, 976-990.

Invited talk

The embodiment of language is a natural result of interactive dynamics

Michael Spivey
UC MERCED, USA

Rather than a sequence of logical operations performed on discrete symbols, real-time cognition might be better described as continuously changing patterns of neuronal activity. The continuity in these dynamics indicates that, in between describable states of mind, much of our mental activity does not lend itself to the linguistic labels relied on by much of cognitive science. I will discuss eye-tracking and computer-mouse-tracking evidence for this temporal continuity in spoken word recognition, sentence processing, and question-answering. I will also provide geometric visualizations of mental activity depicted as a continuous trajectory through a state space of linguistic, perceptual, and motoric dimensions. In this theoretical framework, close visitations of labeled attractors may constitute word recognition events and object recognition events, but the majority of the mental trajectory traverses unlabeled regions of state space, resulting in multifarious mixtures of mental states. Finally, since these findings suggest that language is not informationally encapsulated from the sensorimotor context, this continuous interaction among linguistic, perceptual, and motor processes may be exactly what forces language to be embodied and situated.

Deciphering the communicative code in speech and gesture dialogues by autoencoding hypernetworks

J. Nam (1,2), Kirsten Bergmann (2), U. Waltinger (2), Stefan Kopp (2), Ipke Wachsmuth (2), & Byoung Zhang (1,2)
(1) SEOUL NATIONAL UNIVERSITY, KOREA; (2) BIELEFELD UNIVERSITY, GERMANY

What kinds of grammar or code are used in interactive communications with speech and gestures? How varied or invariant is this code among people in a language community? What types of communicative code facilitate the alignment of the speech and gesture for language understanding?

To study these and other related questions we develop computational techniques using coding theory and machine learning that decipher the communicative code in embodied multimodal interaction. We use data from the SaGA (Bielefeld Speech and Gesture Alignment) corpus which consists of 25 dyads of naturalistic, yet controlled, and systematically annotated speech and gesture use, engaged in a spatial communication task (Luecking, 2010). For the work we present here, a sub-corpus of 5 dyads is employed (473 noun phrases, 288 gestures) combining three kinds of information. First, gesture coding including gestural representation techniques (e.g., drawing, placing) and morphological gesture features (e.g., handshape). Second, a transcription of the spoken words and dialogue contextual information (information state, thematization, elemental actions of direction giving). And third, a coding of the gestures' referent objects and their spatio-geometrical properties (dimensionality, symmetries, etc.).

We formulate the gesture generation problem as an encoding problem and use the unsupervised, autoencoding technique, where the input vector x is transformed by some function f(.;W) to the output vector y which is the same as the input, i.e. y = f(x;W) = x. For transformation we use the hypernetwork graphical architecture. The hypernetwork is a hypergraph structure, where the edges are weighted and represent the subsets of the variables (variables). One advantage of the hypernetwork is that it can capture the compositional structures or code words (or construction grammar rules) in its hypergraph structure. We apply an expectation-maximization style of learning algorithm to build the best autoencoding hypernetwork for the observed gesture-speech dialogue data. Another advantage of the hypernetwork is its generativity, i.e. the hypernetwork model can generate the values of the unknown (unobserved) variables from those of the known (observed) variables by probabilistic inference. This feature is especially useful for artificial communicative agents since the learned hypernetwork can be used to synthesize the gestures for virtual avatars or humanoid robots.

References:

Bergmann, K., & Kopp, S. (2010). Modelling the production of co-verbal iconic gestures by learning Bayesian decision networks. Applied Artificial Intelligence 24(6):530–551.

Kopp, S., Bergmann, K. & Wachsmuth, I. (2008). Multimodal communication from multimodal thinking - towards an integrated model of speech and gesture production. Semantic Computing 2(1):115–136.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010). The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In LREC 2010 Workshop.

Zhang, B.-T. (2008). Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory. IEEE Comp. Intell. Mag., 3(3), 49-63.

Zwann, R. A. & Kaschak, M. P. (2008). Language in the brain, body, and world. Chap. 19, Cambridge Handbook of Situated Cognition.

(Beyond) referential mechanisms in spatial language comprehension

Michele Burigo, Pia Knoeferle
UNIVERSITY OF BIELEFELD, GERMANY

Everyday interaction often involves comprehending spatial language such as when trying to check the time and being told "the clock is *on* the table". It is well established that spatial language processing requires attention mechanisms (Carlson & Logan, 2005; Logan, 1994) but how precisely people deploy visual attention during real-time spatial language comprehension, is still unclear.

The Attention Vector Sum (AVS) model postulates that to comprehend sentence (1) people must shift their attention from the vase ('reference object') to the clock (located object, e.g., Carlson-Radvansky & Irwin, 1994; Carlson & Logan, 2005, Regier & Carlson, 2001). An alternative account from 'visual world' studies suggests people incrementally inspect objects as they are mentioned and thus for (1) inspect the clock followed by the vase (e.g., Tanenhaus et al., 1995). In sum, these two accounts predict opposing inspection orders although the visual world (but not the AVS) account specifies the time course of visual attention allocation.

Two eye-tracking studies examined gaze pattern to objects during spatial language comprehension, and evaluated their fit against predictions of the AVS (reference object -> located object) and visual world (located object -> reference object) accounts.
(1) "The clock is above the vase".

We recorded eye movements while people listened to spatial descriptions (e.g., (1)) and verified whether the sentence matched (vs. didn't match) the picture. We analyzed fixations and inspections (consecutive fixations to an object) for the matching picture-sentence pairs. Shortly after people heard "above" they fixated the vase more often than the clock, corroborating the visual world account. In contrast, analyses of inspections showed that people after hearing "above", and after one inspection to the vase, looked next to the clock on 70 percent of inspections The distribution of fixations during "the vase" confirms this view, in that 30 percent of fixations are directed at the clock (vs. 60 percent to the vase vs. 10 percent to a third unrelated distracter object).

In sum, gaze analyses after people had heard "above" revealed that (a) people anticipated the post-verbal object as predicted by the visual world account; but (b) after inspecting the vase, people next attended mostly to the clock as predicted by the Attention Vector Sum account. While the AVS model cannot accommodate findings (a), the visual world account alone cannot accommodate findings (b), suggesting we need aspects of both accounts to accommodate the linking of eye movements to comprehension and verification processes.

References:
Carlson, L. A. & Logan, G. D. (2005). *Attention and spatial language*. In L. Itti, G. Rees, & J. Tsotsos (Eds.), Neurobiology of attention (pp. 330-336). San Diego, CA: Elsevier.
Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33, 646-671.
Logan, G. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: HPP*, 1015-1036.
Regier, T. & Carlson, L. (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General*, 130, 273-298.
Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632- 1634.

FRIDAY, AUGUST 26, 2011

Invited talk

Spatio-temporal patterns of brain activation reflect embodied language processing

Olaf Hauk
MRC CAMBRIDGE, UK

Neuroscientific theories of embodied semantics make strong predictions about brain activation patterns evoked by words. If understanding a word referring to an object or action is grounded in perceptual or motor experience, then the corresponding visual or motor areas of the brain should be activated in the process. More specifically, different cortical motor areas should be activated depending on the effector of the action (e.g. hand, leg or mouth). I will start with a selective review of functional magnetic resonance imaging (fMRI) studies that have confirmed these expectations. For example, stimuli referring to different bodily effectors (e.g. arm-, leg- or face-related words such as “pick”, “kick” or “lick”) produced somatotopic activation patterns in motor cortex. This has meanwhile been replicated for action-words in different contexts, e.g. in idiomatic expressions (“he grasps the idea”). The findings have been extended to more specific categories of action-words, e.g. uni-manual action-words (“throw”) produced more strongly lateralized activation in motor cortex than bi-manual action-words (“clap”). Recent studies have reported different lateralisation patterns for uni-manual words in left and right handers, respectively. We could not replicate this finding, raising the question to what degree active (doing) or passive (experience) shape semantic representations during language acquisition.

fMRI methodology integrates brain activity over several seconds, and is therefore not able to distinguish between early retrieval and late post-translational processes. In order to address this problem, we showed that activation that is specific to action-words or object-words correlates negatively with word frequency, suggesting that this activation reflects retrieval of lexico-semantic information, rather than mental imagery. A more direct test, however, is to use electro- and magnetoencephalography (EEG, MEG) in order to measure brain activation with sufficient time resolution to distinguish early from late processes. In an ERP study, we found that visually presented arm-, leg- and face-related words could be distinguished as early as 210-230 ms after word onset, and that the putative generators of these signals showed a somatotopic pattern. This demonstrates that action-semantics affects early brain processes, but the evidence is still correlational: reading action words affects motor areas, but does activity in motor areas also affect action-word processing? In a transcranial magnetic stimulation (TMS) study, stimulation of hand and leg motor cortex at 150 ms facilitated arm- and leg-word processing, respectively, but only for the left hemisphere. In a combined EEG/MEG study, participants initiated trials by button press either with their finger or their foot. Words were presented while the button was still pressed, i.e. motor cortex was activated. We found effects of word-effector congruency (e.g. finger button press followed by arm-word is congruent, but followed by a leg-word incongruent) both in hand motor cortex and in posterior superior temporal gyrus, i.e. a classical perisylvian language area, around 150 ms. This demonstrates that motor cortex activation specifically affects action-word processing at early latencies.

We conclude that there is converging evidence from different metabolic and electrophysiological neuroimaging techniques supporting the idea of rapid access to embodied semantic representations for action-words in language comprehension.

Playing patty-cake interferes with comprehending the names of manually experienced objects

Eiling Yee (1), Evangelia Chrysikou (2), Esther Hoffman (3), Sharon Thompson-Schill (2)
(1) BASQUE CENTRE ON COGNITION, BRAIN AND LANGUAGE; UNIVERSITY OF PENNSYLVANIA, USA; (3) CORNELL UNIVERSITY, USA

How do we know the meaning of words? Sensorimotor-based theories of semantic representation claim that for words referring to actions and concrete objects, meaning is distributed over the neural substrates that are invoked when we perform/perceive/interact with those actions or objects. Thus, our experiences with a given object should determine which neural substrates are important to its representation. These theories therefore predict that 1) a concurrent secondary task that occupies a neural substrate that is an important part of a word's semantic representation will interfere with the ability to access that representation, and 2) the extent to which our experience with a given object recruits that same neural substrate should determine *how much* interference will be caused by that secondary task. Previous investigations of the first prediction have typically asked whether performing a concurrent action can interfere with the recognition of action verbs. Yet nouns, although less directly linked to action compared to action verbs, can also have associated actions. We test here whether concurrent manual action can interfere with noun comprehension, and importantly, whether the amount of manual experience with the referent category predicts the degree of interference.

Participants (*N*=72) heard words referring to objects and made concreteness judgments about the objects, while either simultaneously performing a patty-cake -like task on a table ("manual interference"), mentally rotating objects ("visual interference") or performing no concurrent task ("no interference"). The words referred to objects which varied (according to obtained ratings) in the extent to which they are interacted with manually (e.g., tiger = low manual interaction, pencil = high manual interaction). We found that performing a concurrent task increased errors for all words. Critically, however, when performing the patty-cake task, the increase in errors was greatest for words referring to objects rated as high in manual interaction. (In contrast, the concurrent visual task did not disproportionately increase errors for manual words.) Further, the amount of interference was predicted by the manual experience ratings, with greater manual experience associated with greater interference.

These findings demonstrate that engaging brain regions underlying manual interaction (with an incompatible manual task) interferes with comprehending the names of that are manually experienced. Hence, these regions are *part of* (rather than peripheral to) the representation of frequently manipulated objects. These data also highlight the role of experience-based knowledge in language comprehension: modality-specific experience determines the neural and representational substrates through which meanings are encoded and words comprehended.

Beyond abstractness-concreteness and imageability ratings: A new measure of the perceptual experience of a concept that better predicts reading times

Louise Connell & Dermott Lynott
UNIVERSITY OF MANCHESTER, UK

The abstract-concrete distinction has been a cornerstone of language and cognition research for decades, with several sets of published abstractness-concreteness and imageability norms in widespread empirical use. These ratings have been assumed to represent the extent of sensory experience of a given word or concept, and have played a key role in developing our understanding of mental representation, conceptual structure and semantic deficits. In the present study, we find that this assumption is unjustified.

We collected modality-specific ratings of experiential strength for over 600 words, reflecting how strongly a concept can be experienced through each of five perceptual modalities: auditory, gustatory, haptic, olfactory or visual (e.g., Lynott & Connell, 2009), and correlated them with ratings from abstractness-concreteness and imageability norms. Results showed that existing abstractness-concreteness and imageability ratings fail to reliably capture the extent of perceptual experience associated with a concept. In particular, abstractness-concreteness ratings neglected the sense of taste and misconstrued the sense of hearing: at the abstract end of the scale, sound-related experience was negatively correlated (i.e., more auditory meant more abstract). Imageability ratings were strongly biased towards visual experience (and its closely-related sense of touch) while neglecting sound and taste.

We next examined the predictive ability of our modality-specific ratings by reducing them to a single variable of maximum modality strength (i.e., a word's highest rating across all five modalities). Words that refer to concepts with a strong perceptual basis are generally read faster than words with little perceptual basis to their referents (i.e., the “concreteness” effect: James, 1975). Taking latency data for 593 words from the e-Lexicon database (Balota et al., 2007), we ran regression analyses (from a basic model of word length in letters, number of syllables, and contextual diversity) to compare the predictive ability of three alternative measures of word “concreteness”. Results showed that maximum modality strength outperformed both abstractness- concreteness and imageability ratings in explaining variance of lexical decision times. Furthermore, while maximum modality strength could also predict naming times, abstractness-concreteness and imageability could not.

These novel findings suggest that abstractness-concreteness and imageability ratings do not, as is often assumed, reflect the perceptibility of a concept. We recommend that researchers needing to ascertain the degree to which a word or concept is (un)related to sensory experience should collect separate ratings for each perceptual modality, and select the maximum rating of experiential strength if a single variable is desired. Failure to do so means that words will erroneously be classed as “abstract” or “low imagery” when, in fact, they have a strong sensory component that participants neglected to consider.

References:

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

James, C. T. (1975). The role of semantic information in lexical decisions. J*ournal of Experimental Psychology, 104*, 130–136.

Lynott, D. & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*, 558-564.

Invited talk

Language processing: a disembodied perspective

Peter Hagoort
MPI FOR PSYCHOLINGUISTICS, THE NETHERLANDS

Abstract tba

Syntactic Constraints on Embodied Language Processing

Lisa Vandeberg, Rolf Zwaan
ERASMUS UNIVERSITY ROTTERDAM, THE NETHERLANDS

Research has demonstrated that visual experience affects language processing (e.g., Wassenburg and Zwaan, 2010). Here we ask more specifically which linguistic constructions are impacted. Dutch participants were presented with a picture of an object in a certain shape or orientation, after which they used key presses to read a sentence about that object phrase by phrase. Both a prepositional phrase (PP1: on the location of the object) and the verb phrase (VP: on the action that was performed to the object) contained disambiguating information that either matched or mismatched with the object's visual properties. The order of presentation of these critical phrases was systematically varied in agreement with Dutch syntax.

We found significant match-mismatch effects on the PP when it was preceded by a VP (on PP1 in structure 1: e.g., The visitor has/the umbrella/laid down/on the counter /in the hall (S-O-VP-PP1-PP2) and on PP2 in structure 2: The visitor has/the umbrella/on the counter/laid down/ in the hall (S-O-PP1-VP-PP2)). This demonstrates that visual experience influences reading in certain regions of the sentence, but not in others. Follow-up experiments are currently conducted to pinpoint under what syntactic circumstances these effects manifest and diminish, providing possible constraints on theories of embodied cognition.

Wassenburg, S.I.; Zwaan, R.A. (2010). Readers routinely represent implied object rotation: The role of visual experience. *Quarterly Journal of Experimental Psychology*, 63, 1665-1670.

Introducing combined TMS and eyetracking into neurolinguistic research

Imke Franzmeier, Evelyn C. Ferstl
UNIVERSITY OF FREIBURG, GERMANY

Objectives: In this study single-pulse transcranial magnetic stimulation (TMS) is linked to eyetracking for the first time to explore semantic processing in a sentence context. Single pulse TMS is usually applied at a fixed time point requiring an inflexible stimulus presentation. Linking eyetracking to TMS, the stimulation pulse can be applied synchronously with the fixation of a specific word, allowing a natural reading behaviour and the investigation of contextual processing.
To test this methodological combination, a paradigm similar to Pulvermüller et al. (2005) was chosen. They associated the left motor cortex to action word processing using lexical decisions on single words. They provided information about lexical access but not about contextual processing. Yet, motor representations for actions strongly rely on subsequent objects: Movements for "*opening a bottle*" are different to "*opening a door*". Its representation is likely to differ. Due to the contextually influenced action representation, the question arises if and how the motor cortex is involved in contextual processing. This study investigates the change of eye movement patterns when the motor cortex is stimulated during action word processing.

Methods: 120 sentences with hand-actions were presented to 18 right-handed, English native speakers, who were asked to silently read the sentences and answer comprehension questions, to ensure attention. The sentences included transitive action verbs followed by objects from various categories (e.g. *Sue* ***decorated*** *the* ***cake*** *with sugar icing*.). Verbs and nouns were matched for length, frequency.

The stimulation was triggered online, synchronously with the first fixation on the action verb (*decorated*) or the subsequent object (*cake*). Single-pulse TMS (90% of the motor threshold) was applied over the left motor cortex (LMC) or the right temporal lobe (control). Eye movement behaviour was measures for (1) the whole sentence, (2) the targets (stimulated word) and (3) the non-targets (either verb, followed by a stimulated noun, or noun subsequent to a stimulated verb).

Results: LMC stimulation of both action verbs and nouns increased the reading times for the whole sentence by 200 ms compared to the control side. Participants fixated verbs longer than nouns. Compared to control stimulation, participants moved back to previous parts more often after LMC stimulation on an action verb (cf. Pulvermüller et al.). Verb stimulation increased dwell times for verbs and subsequent nouns. LMC stimulation increased fixation counts and regressions out of targets and non-targets.

Conclusion: LMC stimulation increased reading times not only for stimulated action words but also on a sentence level. The stimulation effect on the sentence level was caused by semantic processing beyond lexical access. As the LMC stimulation affected both verbs and nouns it is proposed that stimulation disrupted the construction of action representations.
This study successfully showed that the motor cortex is not only involved in lexical processing but also in the contextual integration of action words.

References:
Pulvermüller, F., Hauk, O., Nikulin, V., & Ilmoniemi, R. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797.

The future of grammatical aspect: The impact of aspect and tense on motor output

Sarah Anderson (1), Teenie Matlock (2), Michael Spivey (2)
(1) UNIVERSITY OF CINCINNATI, USA; (2) UNIVERSITY OF CALIFORNIA, MERCED, USA

How does grammatical aspect influence event understanding? Linguists have studied differences between perfective (e.g., *walked*) and imperfective (e.g., *walked*) aspect, but little is known about how these forms are processed or what their role is in event understanding. Some psycholinguistic work has found that people take an internal perspective when reading about events described with imperfective aspect and an external perspective with events described with perfective aspect (Madden & Zwaan, 2003). Other work has found differences in amount of action conceptualized (Matlock, in press) or motor output produced (Bergen & Wheeler, 2010). Although such work suggests that aspect is grounded in perceptual and motor information, details about online processing are unclear.

In the current work, we use mouse-tracking (Spivey, Grosjean, & Knoblich, 2005) to explore motor output in response to motion descriptions that varied on both tense and aspect. Previously, we found that aspect differentially interacts with contextual information (Anderson, Matlock, & Spivey, 2010). Here, we extend our methodology to investigate the interaction of tense and aspect. In our task, participants heard motion descriptions with past or future tense and with perfective or imperfective aspect, and moved a character into a scene to make it match a spoken sentence (Figure 1).

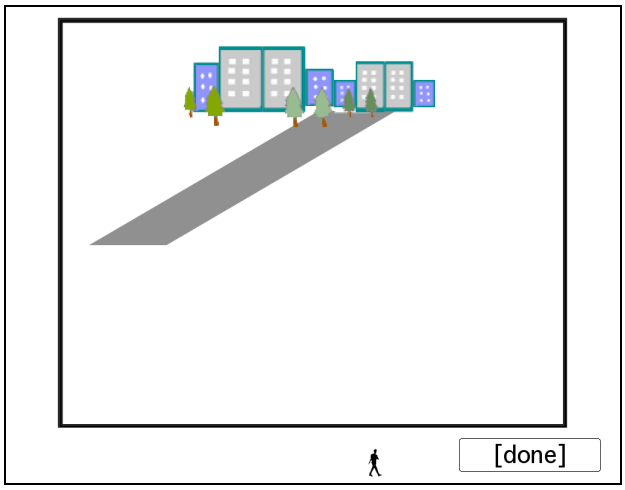

Figure 1: Visual scene corresponding to, “David walked/was walking/ will walk/ will be walking to campus.”

Mouse-tracking is robust for measuring various response/motor dynamics indices, including three used here. First, in analyzing drop locations, the character was placed farther from destinations (e.g., campus) with future than past tense, indicating increased attention to the location of a future action. Second, this increased attention is also reflected in spatial differences. The x-coordinates reveal a main effect of tense, indicating that processing future tense sentences entails increased attention to the location of that future event. The y-coordinates show main effects of both tense and aspect, suggesting that they additively result in differences in movement trajectories. Finally, we examined movement durations, finding a main effect of aspect. Therefore, perfective sentences induce shorter movement durations than do imperfective sentences, regardless of tense. These data provide new insights on processing aspect and tense in event understanding and on how perceptual simulation occurs at the level of grammar.

References:

Anderson, S.E., Matlock, T., & Spivey, M.J. (2010). On-line interactions of context and grammatical aspect. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1198-1204), Mahwah, NJ: Lawrence Erlbaum Associates.

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language, 112*, 150-158

Madden, C.J. & Zwann, R.A. (2003). How does verb aspect constrain event representations? *Memory & Cognition, 31*, 663-672.

Matlock, T. (in press). The conceptual motivation of aspect. In K. Panther and G. Radden (Eds.), *Motivation in Grammar and the Lexicon.* Amsterdam/Philadelphia: John Benjamins.

Spivey, M.J., Grosjean, M. & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*, 10393-10398.

Do past and future potential events activate the mental time line?

Roberto Aguirre Fernández de Lara (1), Olga Soler Vilageliu (1), & Julio Santiago de Torres (2)
(1) AUTONOMOUS UNIVERSITY OF BARCELONA, SPAIN (2) UNIVERSITY OF GRANADA, SPAIN

In the current priming study, we presented Spanish sentences with verbs both in a realis (indicative) mood and non-realis (past subjunctive and future conditional) moods. Past subjunctive verbs describe an event that might have occurred in the past. Future conditional verbs describe events that might happen in the future if certain conditions are met.

Spanish participants were asked to judge the tense of a given situation (past versus future), independently of whether it was an actual event or a potential one, by pressing either a left or a right key. The assignment of keys to past and future could be either congruent (left-past, right-future) or incongruent (left-future, right-past). Verbal stimuli were sentences without temporal adverbs and had a conversational implicature at the beginning (e.g. "*Por eso él marchaba*.").

Research on Time Metaphor has demostrated that when people think about actual events, they tend to activate a left-right mental time line (Santiago et al, 2010; Fuhrman & Boroditsky, 2010). The same occurs when the events are presented by means of sentences or isolated verbs in a realis (indicative) mood (Santiago et al, 2007; Ouellet et al, 2010; Ulrich & Maienborn, 2010).

Will potential events also activate this time line? So far, there is no evidence available on this question. Note that future indicative verbs also describe events that have not occurred yet, but verbal mood implicates that they will occur. Will events of a lower or null probability of actual occurrence also be located on the left-right time line?

Results showed that the tense of indicative verbs (past vs. future) interacted with response hand, producing faster and more accurate responses in congruent than incongruent conditions, thereby replicating prior findings. In contrast, no congruency effect was observed for sentences with verbs in a norealis mood. Thus, the mental time line seems to be only activated by sentences describing actual events, and not by sentences describing potential events. Further research is needed to explore whether including temporal adverbials in potential events would activate the time line metaphor.

References:

Fuhrman, O. & Boroditsky, L. (2010). Cross-Cultural Differences in Mental Representations of Time: Evidence from an Implicit Non-Linguistic Task. Cognitive Science. DOI: 10.1111/j.1551-6709.2010.01105.x

Quellet, M., Santiago, J., Funes, M. J., & Lupiáñez, J. (2010). Thinking about the future moves attention to the right. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 17–24

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14,* 512–516

Santiago, J., Roman, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research, 74,* 59–70.

Ulrich, R., & Maienborn, C. Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition* (2010), doi:10.1016/j.cognition.2010.08.001

Invited talk

On the diverse nature of time in language comprehension

Silvia Gennari
UNIVERSITY OF YORK, UK

We conceive of events as taking time. Every day we read about events in the world and effortlessly understand how long these events last. In reading “Somali pirates held two hostages”, we understand that hostages were held for some period of time during which other likely events happened, e.g., a ransom request. How do we understand the duration of this event, if we never took part in such affairs? What aspects of the actual event are retained in our mental representation of it? This issue is central to human cognition, since it underlies a fundamental aspect of our mental life, namely, our experience of the passage of time.

The talk will examine the mental processes by which we understand the duration of events. It is argued that understanding the duration of events is determined by our knowledge and experience of which events are associated with which others. Specifically, it is argued that the diversity of these associated events (how they differ from one another) determines our understanding of event duration. What is called time in the context of understanding an event is not external clock time, but rather, the diversity of the events that our experience tells us are associated with that event. The talk will review evidence for this proposal and will discuss the challenges that it poses for theories of embodied simulations.

SATURDAY, AUGUST 27, 2011

Invited talk

Synesthesia

Seana Coulson
UC SAN DIEGO, USA

Abstract tba

When up-words meet down-sentences: New evidence for compatibility-effects

Martin Lachmair, Monica De Filippis, Carolin Dudschig, Irmgard De la Vega, Barbara Kaup
UNIVERSITY OF TÜBINGEN, GERMANY

In a series of experiments participants processed words referring to entities with a typical location in the upper or lower part of the visual field (e.g., up-word: 'roof' vs. down-word: 'root'). Participants responded with compatible or incompatible key presses (i.e., with an up- or down response to an up- or down word). We manipulated (a) whether the experimental task required word reading or not (lexical decision vs. responding to the font color), (b) whether responses involved an upwards/downwards movement or were stationary, and (c) whether stimuli were nouns (see above) or verbs (e.g., 'rise' vs. 'fall'). We observed significant compatibility effects with faster response times in compatible than incompatible trials in all conditions.

The embodied cognition framework provides a plausible explanation. According to this framework, interacting with the world leaves experiential traces in the brain which are later re-activated when people process words referring to the respective entities. If words appear in sentences, the activated traces are presumably combined to yield representations consistent with the meaning of the sentence (Zwaan & Madden 2005). Our data suggest that during word processing people automatically re-activate spatial aspects of their experiences with the objects and events referred to. However, what happens when words are embedded in sentences that either support or cancel the word-based information? Two experiments were conducted to investigate this issue.

Participants read sentences ending in an up- or down-noun (e.g. 'The soldier looks at the waving flag'), and judged their sensibility by responding with an upwards or downwards movement. Sentence length and frequency of content words was controlled. Care was taken that the verbs were not associated with an up- or down-movement when presented in isolation (rating study). A second rating study assured that the up- or down-nature of the target words was supported by the sentence context in this experiment (i.e., a sentence with an up- vs. down target word described an up- vs. down-situation, respectively). The results of this experiment replicated our previous effect: Response times were faster in compatible than in incompatible trails. As all sentence contexts were supportive, it is impossible to decide whether the compatibility effect is due to the target words or due to the content of the sentences as a whole.

In a second experiment, we included sentences that reversed the location information induced by the target words (e.g., 'The sad soldier presents the widow with the flag.'). Interestingly, response times were affected by the congruency between target word and sentence context: Supporting sentences were processed significantly faster than reversing sentences, suggesting that sentence context did not prevent word-based information from being activated.

Taken together our studies show that location information gets activated automatically when words are being processed, both in isolation as well as when embedded into sentences. Future studies need to clarify how the respective word- and sentence-based processes interlock during language comprehension.

References:

Zwaan, R.A., & Madden, C.J. (2005). Embodied Sentence Comprehension. In: D. Pecher & R.A. Zwaan (Eds.), *The Grounding of Cognition* Cambridge, UK: CUP.

Emotion and speech: Evidence for the articulatory feedback hypothesis

Ralf Rummer, Judith Schweppe
UNIVERSITY OF ERFURT, GERMANY

It has been debated for a long time whether the relation is arbitrary or not. Researchers arguing for a non-arbitrary relation assume that vowels evoke positive or negative feelings as a function of their sounds. In the literature on sound symbolism it is suggested that /i/ (as in peace) is associated with positive words and /o/ is associated with negative words. We resurrect this debate from the perspective of embodied emotion research. Our articulatory feedback hypothesis postulates a direct link between face muscle contraction during speech production and self-perception of an individual's emotional state. Our first study was based on an experimental paradigm introduced by Strack, Martin, and Stepper (1988). The authors instructed their participants to rate the funniness of cartoons while holding a pen between their teeth – which activates the zygomaticus major muscle (ZMM, the muscle responsible for smiling) – or while holding a pen between their lips – which blocks the ZMM. The cartoons were rated as funnier when the ZMM was contracted.

Interestingly, the ZMM is also contracted during the articulation of vowel /i/ and blocked during the articulation of the vowel /o/ (Hardcastle, 1976). In order to reveal this relation between vowel production and the emotional state, we used an articulatory suppression paradigm. In Experiment 1, participants had to produce /i/ or /o/ once per second when rating the funniness of the cartoons. As predicted, the cartoons were rated as funnier when the secondary task was to articulate /i/ than when it was to articulate /o/. Hearing /i/ or /o/ while looking at the cartoons did not influence the funniness ratings.

Experiment 2 aimed at demonstrating that the articulatory feedback mechanism has a direct impact on the emotional impact of words. In the first condition of this study, participants had to articulate and rate the pleasantness of Consonant-Vowel-Consonant-Vowel-pseudowords. The vowel were either /i/ and /e/ (i&e-pseudowords; e.g., "gife") or /o/ and /ü/ (o&ü-pseudowords; e.g., "gofü"). I&e-pseudowords were rated as more pleasant than o&ü-pseudowords. In a control condition, participants listened to these pseudowords (i.e., the critical face muscles were not contracted). In contrast to the articulation condition, no pleasantness differences between the i&e-pseudowords and the o&ü-pseudowords were observed here. In sum, these findings provide clear evidence for the articulatory feedback hypothesis.

References:

Strack, F., Martin, L.L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54*, 768–777.

Invited talk

Emotional Body

*Beat*rice *de Gelder*
TILBURG UNIVERSITY, THE NETHERLANDS

Abstract tba

Friday, AUGUST 26, 2011

Abstracts for Poster Presentations

Cognition and semiotic artifacts – situatedness and embodiment in Peircean cognitive semiotics.

João Queiroz, Pedro Atã
FEDERAL UNIVERSITY OF JUIZ DE FORA, BRAZIL

According to Peirce's semiotic theory of mind, cognitive process fundamentally depend on semiotic process, in a sense that diverges radically from computationalism. For Peirce, thinking involves the process of sign action. Against any form of internalism, Peirce can be considered a precursor of situated mind and distributed cognition thesis. But differently from the anti-cartesianism defended by some embodied-situated cognitive science, which is predominantly anti-representacionalist, as recently explored in a Merleau-Pontyan (Dreyfus 2002), Heidegerian (Wheeler 2005), or a Gibsonian (Chemero 2009) trend, for Peirce, mind is semiosis in a dialogical -- hence communicational -- materially embodied form, and cognition is the development of available semiotic material artifacts in which it is embodied as a power to produce interpretants. It takes the form of development of semiotic artifacts, such as writing tools, instruments of observation, notational systems, languages, and so forth, as stressed by Skagestad (2004) with respect to the concept of intelligence augmentation. Our approach here centers on the consideration of relevant properties and aspects of Peirce's concept of mind as semiosis.

References:

Chemero, A. (2009). Radical Embodied Cognitive Science. MIT Press.

Dreyfus, H. L. (2002). Intelligence Without Representation-Merleau-Ponty's Critique of Mental Representation: The Relevance of Phenomenology to Scientific Explanation. *Phenomenology and the Cognitive Sciences. Special issue on Dreyfus and Representationalism* 1(4), 367–383.

Skagestad, P. (2004). Peirce's semeiotic model of the mind. In: C. Misak (Ed). *The Cambridge Companion to Peirce*. Cambridge University Press. (pp. 241-256)

Wheeler, M. (2005). Reconstructing the Cognitive World -- The Next Step. The MIT Press.

Differences in comprehension task affect the processing of a speaker's gaze direction.

Helene Kreysa, Pia Knoeferle
BIELEFELD UNIVERSITY, GERMANY

Research using the 'visual world' paradigm has shown that listeners can rapidly integrate the unfolding speech content with information in visual context (Tanenhaus et al., 1995). Since speakers robustly gaze at objects before mentioning them (Griffin & Bock, 2000), one such contextual cue may be the shifting focus of the speaker's gaze, as suggested by recent studies on referential disambiguation (e.g., (Hanna & Brennan, 2008). A series of eyetracking experiments served to identify the specific factors influencing how speaker gaze is used to anticipate upcoming referents, and to determine the extent to which this generalizes across subtly different comprehension tasks and/or levels of syntactic processing difficulty.

In three studies, a video-taped speaker referred to depicted characters, using either a subject-initial or a non-canonical object-initial German sentence. She shifted gaze once from the pre-verbal to the post-verbal referent, thus potentially allowing listeners to anticipate which character would be mentioned post-verbally. We compared participants' fixations of the NP2 referent during sentence comprehension depending on the presence or absence of the speaker, as well as post-sentence verification times to a task-specific response template. In Experiment 1, the task was to determine whether a schematic depiction of the sentence content correctly highlighted the characters referred to linguistically. The response templates in Experiments 2 and 3 required correct identification of the patient or the thematic role relations of the sentence, respectively.

The availability of speaker gaze had no effect on response times, while sentence structure affected them only in the task of verifying thematic roles (Experiment 3, with earlier responses for subject- than object-initial sentences). However, the continuous recording of eye movements during comprehension revealed that listeners did anticipate the post-verbal referent when they saw the speaker fixating this character before mentioning it, compared to when no speaker was visible. This effect was reliable in all three experiments (ps < .001). At the same time, considerable task differences were found in the eye movement patterns with regard to the effect of sentence structure and its modulation by speaker gaze: In the critical time period between the speaker's gaze shift and the onset of the NP2, the patient verification task (Experiment 2) led to more anticipatory fixations of the NP2 referent during SVO than OVS sentences, while this pattern flipped for reference assignment and thematic role verification (Experiments 1 and 3, respectively). We will discuss these results in detail, concluding that participants' task can affect how anticipatory eye movements relate to sentence structure during spoken language comprehension. To accommodate these findings, any comprehensive processing account of situated language comprehension must include task effects on the allocation of visual attention.

References:

Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 632–634.

Griffin & Bock (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Hanna & Brennan (2008). Speaker's eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57,* 596–615.

Does catching a concept differ from catching a flower?

Claudia Scorolli (1), Pierre Olivier Jacquet (1), Ferdinand Binkofski (2), Roberto Nicoletti (1), Alessia Tessari (2), Anna M. Borghi (2)
(1) UNIVERSITY OF BOLOGNA, ITALIY; (2) UNIVERSITY OF AACHEN, GERMANY

Embodied accounts have focused on language grounded in bodily experiences but have neglected that language also plays a role in shaping experience. Recently scholars (Borghi & Cimatti, 2009; Dove, 2010; Barsalou, et al, 2008) have tried to integrate linguistic and modal approaches. Specifically Borghi & Cimatti (2009) proposed that words are forms of actions; therefore words do not simply involve a form of superficial processing and cannot be reduced to mere signals of something. The WAT (Words As Tools) proposal has implications for the neural representation of abstract word meanings: because they are mostly acquired through language, they should rely on the social experience of language more than concrete ones.

The aim of the present study is to test WAT scrutinizing the role of left primary motor cortex during abstract and concrete sentence processing. We combined action-related and non-action-related verbs with nouns of graspable and non-graspable objects. While participants performed a sentence sensibility task, single-TMS pulses were delivered over the hand area 250 ms after verb or noun presentation in each of four combinations of abstract and concrete verbs and nouns.

Analysis of motor evoked potentials (MEPs) showed greater peak-to-peak amplitude for Concrete than for Abstract Verbs during the verb processing. As to verb-noun integration, analysis of MEPs after TMS pulse during noun presentation revealed larger MEPs in sentences containing Abstract rather than Concrete Verbs. Finally, analysis of response times showed that compatible combinations were processed faster than mixed ones; moreover in combinations containing concrete verbs, participants were faster when the pulse was delivered on the first than on the second word. These results extend previous findings showing precocious activation of hand related areas after concrete verbs processing. The delayed activation of the same areas by abstract verbs is consistent with WAT, particularly with the emphasis placed on the role of different acquisition mechanisms (Borghi et al, 2011). The processing of abstract verbs could precociously engage mouth related motor areas, that later affect contiguous areas (hand areas). Mouth areas could be crucial for abstract word processing as the acquisition of a word like “democracy” implies the presence of somebody explaining the word meaning, using language. Abstract words still involve a bodily experience but the contribution of the social/linguistic dimension, so crucial for their acquisition, is reflected in their representation.

References:

Barsalou, L.W., Santos, A., Simmons, W.K., & Wilson, C.D. (2008). Language and simulation in conceptual processing. In M. De Vega, A.M. Glenberg, & A.C. Graesser (Eds.). *Symbols, embodiment, and meaning* (pp. 245–284). Oxford, U.K.: Oxford University Press.

Borghi, A.M., & Cimatti, F. (2009). Words as tools and the problem of abstract words meanings. In N. Taatgen & H. van Rijn (eds.). Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp. 2304-2309). Amsterdam: Cognitive Science Society.

Borghi, A.M., Flumini, A., Cimatti, F., Marocco, D. & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. *Frontiers in Cognition*, 2:15.

Dove, G. (2010). An additional heterogeneity hypothesis. *Behavioral and Brain Science, 33*, 209-10.

Embodied representation of tool, hand and abstract verbs: less action shows more sensorimotor grounding

Olga Dragoy (1,2), Svetlana Malyutina (3), Maria Ivanova (4,5), Thomas Meindl (6), Aleksey Petrushevsky (4), Evgeny Gutyrchik (7)

(1) UNIVERSITY OF GRONINGEN; (2) MOSCOW RESEARCH INSTIUTE OF PSYCHIATRY, RUSSIA; (3) LOMONOSOV MOSCOW STATE UNIVERSITY, RUSSIA; (4) CENTER OF SPEECH PATHOLOGY AND NEUROREHABILITATION, RUSSIA; (5) HIGHER SCHOOL OF ECONOMICS, RUSSIA; (6) INSITUTE OF CLINICAL RADIOLOGY MUNICH, GERMANY; (7) MUNICH UNIVERSITY, GERMANY

Understanding action-related linguistic material has been shown to be mediated by sensorimotor cortex (e.g. Hauk, Johnstrude, & Pulvermüller, 2004). Most studies report additional brain activation related to processing specific action semantics (body part/tool action verbs) in contrast to a more general semantics or non-linguistic baseline. However, the opposite pattern, with more activation for less specific hand actions as opposed to tool actions, was found too (Yang & Shu, 2011). The present study focused on the relationship between the amount of sensorimotor activation and specificity of action verb semantics.

In a blocked fMRI experiment, German verbs denoting motor actions performed by hand with a tool (tool verbs, e.g. *to cut*) were contrasted to verbs denoting motor actions performed by hand with no tool (hand verbs, e.g. *to milk*), and both categories of motor verbs – to abstract verbs (e.g. *to hate*). The semantic category of the verb and deep semantic processing were insured by the task. A verb (e.g. *to cut*) and two nouns (e.g. *bread* and *blood*) were presented in a triangular array; participants were required to choose the appropriate object for the verb by pressing a button. Groups of verbs and pairs of nouns were balanced on frequency and length; appropriateness of objects was tested in a preliminary questionnaire. 17 native German speakers (mean age 33 y.o.) participated in the study. Imaging was performed on a 1.5-T Siemens Avanto scanner.

Along with vast extra activation in frontal and temporal cortex (including classical language-related left inferior frontal and superior temporal gyri), abstract verbs as compared to motor verbs elicited stronger effects in the supplementary motor area (BA 6) bilaterally. The same held true for hand verbs contrasted to tool verbs. Moreover, hand verbs were associated with more activation in the right postcentral gyrus (BA 2/3) than tool verbs. No extra activation was found for the reversed contrasts. Interestingly, the more cerebral activation accompanied processing verbs of a particular kind, the longer reaction times were. Participants were significantly slower when choosing the appropriate object for abstract verbs than for hand verbs, and for hand verbs it took longer than for tool verbs. A follow-up questionnaire completed by 50 German speakers revealed the imageability hierarchy *abstract<hand<tool*, with abstract verbs being the least and tool verbs the most imageable.

The findings suggest that imageability of the verb might influence both the time of its processing and the amount of cerebral activation it is related to. In this respect sensorimotor mediation does not reflect neuroanatomical distribution of specific aspects of verb semantics, but the overall effort an action verb requires to be processed.

References:

Hauk, O., Johnstrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307.

Yang, J., & Shu, H. (2011). Embodied representation of tool-use action verbs and hand action verbs: Evidence from a tone judgment task. *Neuroscience Letters, 493*, 112-115.

Visual context effects on abstract sentence comprehension: evidence from eye tracking

Ernesto E. Guerra & Pia Knoeferle
BIELEFELD UNIVERSITY, GERMANY

Experimental research on language embodiment provides evidence that perceptual processes contribute to the comprehension of concrete language (e.g., Zwaan & Kaschak, 2008). Furthermore, embodied theories of language argue these perceptual processes are also implied in abstract interpretation (Gallese & Lakoff, 2005). Results from similarity judgment tasks suggest abstract concepts (e.g., *similarity*) are linked to experiential ones (e.g., *distance*, Casasanto, 2008). Similarity-distance mapping implies `non-referential' concept-world relationships (e.g., linking "battle and war" to `close' versus "peace and war" to `far' in space). While referential visual context can rapidly inform comprehension, it's unclear whether this holds for non-referential concept-world relationships.

Two eye-tracking reading studies examined a) whether spatial distance affects incremental semantic interpretation of abstract sentences and b) whether non-referential visual context can modulate real-time comprehension. Experiment trials had three parts: (i) Participants (n=32) inspected a picture (cards far versus close to one another). (ii) They read a sentence implying similarity (1) versus dissimilarity (2) and judged its veracity. (iii) They verified whether a picture matched (vs. mismatched) the cards from part (i).

Experiment 1 used a visual context linked to a sentence (two playing cards displayed the first two sentential nouns) whereas Experiment 2 used unrelated, empty, cards. To account for immediate effects of card distance on sentence comprehension requires more than just a referential mechanism. In contrast, later or no effects can delineate effects of non-linguistic perceptual processes in comprehension.

(1) *Kampf(NP1) und(coord.) Krieg(NP2) sind(VP1) freilich(ADV) entsprechend(ADJ), das verriet(VP2) der Anthropologe(NP3).*

'Battle and war are surely analogical, so suggested the anthropologist'.

(2) *Frieden(NP1) und(coord.) Krieg(NP2) sind(VP1) bestimmt(ADV) verschieden(ADJ)*...

'Peace and war are certainly different…'

Card distance affected first-pass reading times. However, the effect patterns were different when the cards displayed words (Experiment 1) versus were blank (Experiment 2). In Experiment 1, we observed a reliable *facilitation effect* in first-pass (ADJ; VP2) and total (NP3) times, both shorter for sentences implying similarity after seeing cards close together (vs. far apart), and vice versa for sentences implying dissimilarity. However, total times at NP1 revealed an *interference effect* with longer reading for sentences implying similarity after seeing cards close together (vs. far apart), and vice versa for sentences implying dissimilarity. In contrast, we observed only *interference effects* in Experiment 2 (first-pass: NP2; ADJ and regression path: NP2; total times: ADJ).

The rapid (first-pass) and extended time course with which word (Experiment 1) and card (Experiment 2) distance differentially affected semantic interpretation for sentences ((1) versus (2)) implicates more than just a referential mechanism, and suggests that relating spatial distance to abstract content is instantaneous and part and parcel of ongoing semantic interpretation.

References:

Zwaan, R., & Kaschak, M. (2008) Language in the brain, body, and world. In Robbins & Aydede (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge University Press.

Gallese, V., & Lakoff, G. (2005) The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22, 455–479.

Casasanto, D. (2008) Similarity and proximity: When does close in space mean close in mind. *M&C*, 36, 1047-1056.

Hand movements as an indicator of implicit processing of emotions in alexithymia.

Uta Sassenberg (1), Ingo Helmich (2), Hedda Lausberg (2)
(1) HUMBOLDT UNIVERSITY BERLIN, GERMANY; (2) GERMAN SPORT UNIVERSITY COLOGNE

Alexithymia ("without words for emotions", Sifneos, 1973) is a common personality trait shared by about 8 % of the population. The deficiency in identifying and describing emotions can contribute to medical, psychosomatic, and psychological disorders. Alexithymia is often accompanied by a lack of understanding other people's feelings, less vivid imagination, and difficulty distinguishing between feelings and the bodily sensations of emotional arousal. Lane and Schwartz (1987) postulate in their model of emotional awareness a cognitive developmental process of the ability to perceive emotions on five levels. On the first two levels, emotional processing is implicit and bound to the perception of bodily sensations and action tendencies. On the three higher levels, emotions are processed explicitly and can be increasingly differentiated. As a measure to judge a person's level of maturity, Lane and colleagues developed the Levels of Emotional Awareness Scales (LEAS). However, it is methodologically problematic that assignment of a person to the first two implicit levels is based on explicit, verbal accounts. Therefore, the aim of this study is to test implicit processing of emotions directly by non-verbal hand movement behaviour.

Fifteen right-handed alexithymic men and a matched control group were video-taped while relating fictional emotional scenarios of the LEAS and - as a control situation - the Hamburg-Wechsler Intelligence Test. Hand movement behaviour was coded with the NEUROGES-ELAN System (Lausberg & Slöetjes, 2009) by two raters blind to whether participants belonged to the alexithymia or control group and to the hypotheses of the study.

Alexithymic participants made significantly more communicative gestures than non-alexithymic participants during the LEAS test. There was no such difference between groups during the intelligence test. Overall, alexithymic participants showed more position shifts than non-alexithymic participants.

The findings indicate that, when processing emotional topics, alexithymic participants might show an increased conceptualisation on the more implicit gestural level than non- alexithymic participants. Similarly, gestures reveal implicit knowledge in cognitive tasks (Broaders, et al., 2007) and might be a product of simulated actions as a form of embodied representation (Hostetter & Alibali, 2008). Further, position shifts might reflect pre- conceptual processes that indicate stress. Thus, these results of individual differences concerning the ability to identify and express emotions are very important to the understanding of the role of the embodiment of emotions.

References:

Broaders, S., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S (2007). Making children gesture reveals implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136* (4), 539-550.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495–514.

Lane, R. D. & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive developmental theory and its application to psychopathology. *American Journal of Psychiatry, 144*, 133–143.

Lausberg, H. & Slöetjes, H. (2009). Coding gestural behavior with the NEUROGES– ELAN system. *Behavior Research Methods, 41* (3), 841-849.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22*, 255–262.

Hitting the high notes: Gesture affects pitch discrimination via spatial grounding.

Louise Connell (1), Zhenguang Cai (1), Judith Holler (2)
UNIVERSITY OF MANCHESTER, UK; (2) MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, THE NETHERLANDS

The nature of the connection between musical and spatial processing is controversial. While pitch is often described in spatial terms such as “high” or “low”, it is unclear whether pitch and space are merely associated dimensions or whether the representation of pitch is actually grounded in space.

In the present study, we used a basic psychophysical task of pitch discrimination, where participants must judge whether a target vocal note is the same as (or different from) a preceding cue note. Importantly, target trials were presented as video clips where a singer gestured upward, downward, or not at all, while singing that target note, thus providing an alternative, concurrent source of spatial information.

Results showed that pitch discrimination was significantly affected by gesture. Signal detection analysis found that gesture created a criterion shift: the spatial movement of gesture biased participants towards the belief they had perceived a pitch movement (i.e., that the target note was different to the cue), despite an underlying tendency to assume all notes were the same. Furthermore, analysis of error trajectory showed that participants were sensitive to the *direction* of spatial information in gesture, where downward gestures made pitch appear lower in frequency, and upward gestures made pitch appear higher in frequency.

Together, these findings provide evidence that musical pitch is spatially grounded. When people hear a musical note, its pitch is not just represented in the auditory modality, but comprises both an auditory and spatial representation of the note's frequency. Critically, this representation is not a static association, where high-pitch notes are mapped to a high spatial location, and low-pitch notes to a low spatial location (e.g., Rusconi, Kwan, Giordano, Umiltà & Butterworth, 2006). Rather, a dynamic, grounded representation of pitch in space, where pitch is represented not only in terms of spatial position but also movement and direction, is consistent with our results. As theories of embodied cognition hold that mental representations are grounded in experience (e.g., Barsalou, 1999; Lakoff & Johnson, 1999), upward / downward bodily movement (e.g., in larynx and torso) when vocalising notes that are higher / lower in pitch that the voice's normal fundamental frequency provides a possible basis for the experiential grounding of pitch in the vertical spatial axis.

References:

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Lakoff, G., & Johnson, M. (1999). Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought. New York: Basic Books.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2005). Spatial representation of pitch height: The SMARC effect. *Cognition, 20*, 1–17.

Large-scale synchronization during concrete and abstract language

Sabine Weiss, Horst M. Müller, & Gert Rickheit
BIELEFELD UNIVERSITY, GERMANY

Since more than 25 years the relevance of the "Embodied Cognition" approach for models in Cognitive Linguistics has been discussed. In general, it has been postulated that neuronal substrates underlying senso-motoric processes are not only engaged during perception and action but also during language comprehension (e.g., Barsalou, 2008; Glenberg, 2010). More specifically, it has been proposed that not only the processing of concrete but also abstract language partly base on the mental simulation of physical entities and actions.

A still unsolved issue is the way the brain integrates these cognitive and linguistic processes occurring more or less simultaneously during language comprehension. One hypothesis raised in the neurosciences providing a possible solution to this "binding-problem" is the Binding-by-Synchrony-Theory (BBS). This theory states that, under certain circumstances, neurons with similar feature properties synchronize their discharges. Neuronal synchrony has not only been shown for adjacent but also between neuronal assemblies of distant brain regions, e.g., somatosensory, motor, visual and parietal association cortices (for a recent review Ulhaas et al., 2009). This large-scale synchronization seems particularly important with respect to distributed neuronal assemblies, which have to be integrated during language processing (Weiss & Müller, 2003 for review).

As an example for the application of the BBS-theory on language comprehension, EEG data were recorded whilst participants performed a semantic judgment task with concrete and abstract congruent or incongruent sentences. In order to investigate whether concrete or abstract language elicits similar or distinct neuronal synchronization patterns, EEG coherence was analyzed in various frequency bands using a bivariate AR-model with time-varying parameters.

Results indicate that different cognitive operations are reflected in different EEG frequency bands. The theta frequency band (4-7 Hz) reflected common and different synchronization networks related to working memory processes and memory-related lexico-semantic retrieval during processing both sentence types. In contrast, the beta1 band (13-18 Hz) showed prominent differences between both sentence types whereby concrete sentences were associated with higher coherence implicating a more widespread range of mental simulation processes. The gamma band (35-40 Hz) reflected the congruency of the sentences and indicated the more difficult integration of incongruent final nouns into the sentence context.

These findings demonstrate that the degree of concreteness of a sentence correlates with the height of EEG coherence during its processing. Moreover, understanding abstract language shares common as well as distinct features with the comprehension of concrete language. Mental simulation during concrete and abstract language also supports the notion that different cognitive operations during sentence processing are associated with multiple brain oscillations.

References:

Barsalou, L.W. (2008). Grounded Cognition. *Ann Rev Psychol* 59, 617-645.

Glenberg, A.M. (2010). Embodiment as a unifying perspective for psychology. Wiley Interdisciplinary Reviews: *Cognitive Science* 1, 586-596.

Uhlhaas, P. J., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolic, D. and Singer,W. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front Integr Neurosci* 3, 17.

Weiss, S. and Müller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain Lang* 85, 325-343.

Measuring speech-accompanied pointing gestures to identify the interaction between the precision of pointing and multimodal compensation strategies

Thies Pfeiffer
BIELEFELD UNIVERSITY, GERMANY

Speech-accompanied pointing gestures are an important part of situative communication. When referring to objects in the environment, an utterance can achieve its meaning only in the partnership of speech and gesture (Kendon, 2000). However, a comprehensive model describing this partnership is still to be found. The presented work aims at substantiating such a model by focusing on the interaction between manual pointing gestures and speech.
In multimodal language production, content has to be distributed between the available modalities. The way to investigate the model behind this distribution followed here is first, the analysis of the shortcomings in precision within manual pointing and second, an analysis of the strategies found in speech used to compensate the decrease in precision.

The analysis is based on a study on human-human interactions with 46 participants. To achieve a high accuracy in measuring the precision of manual pointing, state-of-the-art motion tracking technology was used to record the movements of the participants' hands and index fingers. Based on the collected data, a data-driven model for the pointing-ray was created. This so-called vector-extrapolation model describes how the direction of a manual pointing gesture is constructed and how the referent object is derived from this gesture. The predictions of the model were compared with results obtained from human interlocutors.

Overall, the model could describe only 9.7% of the correct identifications given by the human interlocutor. Thus, either the vector-extrapolation model, which implements the common view (Haviland, 2000; Kita, 2003) and a naïve approach to modeling pointing, is inadequate for describing manual pointing, or, manual pointing is highly imprecise.

Further analysis showed that in the study, the participants' referential expressions were found to be less successful (only 86% correct identifications), i.e. the human interlocutors had more problems, when they were asked to use only gestures. The results suggest that manual pointing is indeed imprecise, but not to the extent predicted by the vector-extrapolation model.

The analysis of the interaction between speech and gesture production revealed a bi-directional interaction. When participants were allowed to use speech and gesture, the numbers of words used to refer to objects increased with distance, thus in speech production the speaker compensated for the loss of precision in manual pointing, which decreases with distance, by increasing the content provided in speech. On the other hand, when participants were not allowed to use speech, they showed different strategies to increase the precision of their manual pointing, e.g. by leaning forward (61%), raising their hands high above the pointing domain (48%) or doing both (30%).

References:

Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), Language and gesture (pp.~13). Cambridge: Cambridge University Press.

Kendon, A. (2000). Language and gesture: Unity or duality. In McNeill, D. (Ed): Language and gesture, pp. 47–63. Cambridge University Press.

Kita, S. (2003). Pointing: Where language, culture, and cognition meet. Mahwah, New Jersey: L. Erlbaum Associates.

Merging Models of Natural Language Parsing and Visual attention

Christopher Baumgärtner
UNIVERSITY OF HAMBURG, GERMANY

We introduce a computational model for the integration of visual information into a system for nat- ural language processing. Merging knowledge from both sources will result in different analyses of sentences, depending on the available context knowledge about visual surroundings. As a rich context will contain information that is helpful for language processing as well as information that is irrelevant, processing all available knowledge is problematic for two reasons.

First of all, the described percepts might contain visual entities that do not apply to a given sentence. Integrating them nonetheless will result in misleading processing steps of the natural language sys- tem. Even worse, different contextual units might contain information that is contradictory, resulting in situations where the language processing system is unable to process a sentence in a way con- forming to the external knowledge. Furthermore, using every single piece of information might exhaust resources of the system processing them. This holds true on a natural processor, like the human cognitive system, as well as on an artificial entity, like a computer. Restricting processing to those parts of the context that are relevant for a given language input is saving those resources, making them available for other tasks.

To avoid the problems of irrelevant or misleading information and consumption of resources, the system limits processing of context to those parts that seem to be relevant for language processing. This choice of a subset of available information units is dependent on the input sentence and its properties. As these properties change due to the integration of context, the interaction between lan- guage and vision becomes a bi-directional process where language processing influences the choice of visual information to be integrated, which in turn influences language processing.

Our system is inspired by attentional processes found in humans. Human visual attention can be divided into bottom-up attention which depends exclusively on features of sensory input like move- ment, color, brightness or shape of objects (e.g. a flash-light in a dark room that draws attention) and top-down attention which is imposed by cognitive processes not directly linked to visual features (Egeth & Yanti, 1997). Language input can have an effect on top-down attention by influencing humans into a certain goal directed behavior after hearing words and phrases referring to objects in the visual field (e.g. looking at keys after hearing the sentence "Where are the car keys?") (Eberhard et al., 1995). The system contains a model of these attentional processes in order to make the above-mentioned choice of contextual entities that apply to a given sentence.

References:

Kathleen M. Eberhard, Michael J. Spivey-Knowlton, Julie C.Sedivy, Michael K. Tanenhaus. 1995. Eye Movements as a Window into Real-Time Spoken Language Comprehension in Natural Contexts. Journal of Psycholinguistic Research, 24:409–436.

Howard E. Egeth, Steven Yantis. 1997. Visual attention: Control, representation, and time course. Annual Review of Psychology, 48: 269-297.

Past and future on the timeline: Grammatically and lexically induced temporal reference of sentences

Simone Alex-Ruf, Claudia Maienborn, Rolf Ulrich
UNIVERSITY OF TÜBINGEN, GERMANY

Recent studies in the field of Grounded Cognition, e.g. Santiago et al. (2007), suggest that there exists a mental timeline running from left to right which we use to locate different time concepts on it. It has been shown that reaction time is shorter when words which refer to the past (e.g. *yesterday*) have to be judged according to their temporal reference with a lefthand keypress and words which refer to the future (e.g. *tomorrow*) with a righthand keypress in contrast to the reversed assignment (past – right, future – left).

We examine if this space-time congruency effect emerges when subjects have to judge whole sentences. Our leading questions concern the function of the timeline in language processing: Is it necessary to understand the temporal information of a sentence, is it just an epiphenomenon in the construction of meaning or are the above findings only due to a memory effect? In the latter case there is no automatic activation of the timeline; only the congruent assignment (past – left, future – right) leads to shorter reaction times, but not directly the temporal reference of the sentences.

We conducted two basic studies where subjects had to judge sentences according to their temporal reference (past or future). In the first experiment this temporal reference was given by grammar, i.e. the sentences were written in the past or the future tense. In the second one the sentences were written in the present tense, but had a lexically induced reference to the future or to the past. E.g. the following sentence refers to the future because of the verb *sich freuen* (to look forward):

Manfred freut sich auf die Theatervorstellung.
Manfred looks forward to the stage performance.

The event which the direct object of the verb refers to is – in relation to the event described by the whole sentence – a future event. The following sentence refers to the past because its verb *erinnern* (to remember) locates the event expressed by its direct object in the past:

Der Zeuge erinnert sich an den Pistolenschuss.
The witness remembers the pistol-shot.

In both studies reaction time was shorter when a sentence which refers to the past had to be judged with the left hand and a sentence which refers to the future had to be judged with the right hand (compared to the reversed assignment).

To test if these findings are due to a memory effect, we conducted two similar experiments with a different task for the subjects: Now they had to judge the content of the sentences (Does it make sense or not?). The results didn't show any congruency effect. Is there indeed no automatic activation of the timeline during sentence processing? Or do we process the temporal information of a sentence only in a shallow manner if we don't have to pay attention to it (like it is the case here)?

References:
Santiago, J., Lupiàñez, J., Pérez, E. & Funes, M. J. (2007): Time (also) flies from left to right. *Psychonomic Bulletin & Review* 14(3): 512-516.

Recent real-world versus future events in the comprehension of referentially ambiguous sentences

Dato Abashidze, Pia Knoeferle, Maria Nella Carminati, & Kai Essig
BIELEFELD UNIVERSITY, GERMANY

Visual world experiments have demonstrated that people prefer to rely upon recently depicted (vs. possible future) clipart events during spoken comprehension. Knoeferle & Crocker, (2007 Experiment 3, Knoeferle & Crocker, 2007) found that, when the verb in NP1-VERB- ADV-NP2 sentences was referentially ambiguous between a recently performed action and an equally plausible future action, participants more frequently inspected the recently acted upon object (vs. an equally plausible object that hadn't yet been acted upon) while hearing the verb. Once verb tense and the post-verbal adverb clarified whether the sentence referred to a past (vs. future) event, participants showed a (non-reliable) tendency to gaze more often at the recently acted-upon object.

It's unclear whether these findings generalize to real-world actions. Knoeferle and Crocker depicted only one event, prior to sentence comprehension, which became the "past" event during subsequent comprehension. Never depicting the future event may have created a bias towards relying on recent events (see Chambers & San Juan, 2008 for evidence against a preferred reliance on perceptual scene aspects).

Two eye-tracking studies addressed these issues. Experiment 1 used the design from the study by Knoferle and Crocker, and examined whether their findings replicate for real-world actions (see Fig. 1 and Sentences 1a. / 1b. in German). The design of Experiment 2 was identical but now the experimenter performed an action before (e.g., sugaring pancakes) and after (e.g., sugaring strawberries) each critical sentence. Each participant saw the ‘future’ and the ‘past’ action of the sentence 50 / 50 in the experiment.

Experiment 1 replicated the findings by Knoeferle & Crocker (2007). The recently acted upon object (the pancakes) was inspected more often than the equally plausible target of a future action (the strawberries) during and shortly after the verb. A tendency towards effects of verb tense and temporal adverbs did not reach significance (see Knoeferle & Crocker, 2007). These findings were confirmed in Experiment 2 (i.e., there was a reliable difference in inspecting the past over the future object during and shortly after the verb and non- reliable effects of tense).

These findings suggest (a) recent real-world actions can rapidly influence comprehension, and (b) people prefer to direct their attention towards a recent action target (vs. an object that will be acted upon in the near future), even when past and future actions occurred with equal frequency. A simple frequency-based account of visual context effects on language comprehension cannot accommodate these findings.

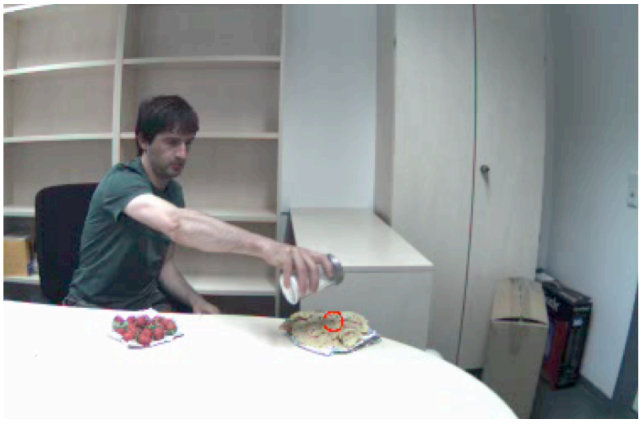

Fig. 1: Example video snapshot: The experimenter sugars the pancakes.

Example sentence set:
1a. *Der Versuchsleiter zuckerte gerade die Pfannkuchen.*
'The experimenter sugared recently the pancakes.'

1b. *Der Versuchsleiter zuckert sogleich die Erdbeeren.*
'The experimenter sugars soon the strawberries.

References:
Chambers, C. & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition, 108,* 26-50.
Knoeferle, P., & Crocker, M.W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543.

The role of working memory in understanding temporal order statements

Raymond Becker (1), Pia Knoeferle (1), & Rolf Zwaan (2)
(1) BIELEFELD UNIVERSITY, GERMANY; (2) ERASMUS UNIVERSITY, ROTTERDAM

Research on the comprehension of temporal order statements such as biclausal *Bevor/Nachdem* ('Before'/'After') sentences is sparse and sometimes conflicting. Solidifying its base is important for examining how both event structure and event ordering relate cognition (e.g., with non-linguistic stimuli, Zacks, et al., 2007; Raisig et al. 2010) to language comprehension. Analyses of event-related brain potentials (ERPs) suggest event order cues (e.g., *bevor* vs. *nachdem*) are processed immediately, although *Bevor* sentences cue greater working memory demands to which HWM (but not LWM) readers are immediately sensitive (Münte et al., 1998). We revisited Münte et al.'s findings using eye tracking allowing sentence re-reading (unlike ERPs with rapid serial presentation). Longer reading time is interpreted as indexing greater working memory demands.

First-pass time analyses revealed temporal cue effects. As in Münte et al., temporal cue effects appeared shortly afger HWM (but not LWM) participants read *Bevor* versus *Nachdem*. Surprisingly, both HWM and LWM readers showed longer re-reading (second pass) post-conjunction in *Bevor* versus *Nachdem* sentences. Thus, on the *second* (but not first) sentence reading, both HWM and LWM readers were influenced by increased working memory demands, suggesting these two groups differ only initially in event order comprehension.

References:

Münte, T.F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395,* 71-73.

Zacks, J.M., Speer, N.K., Swallow, K. M., Braver, T. S., & Reynolds, J.R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin, 133,* 273-293.

Raisig, S., Welke, Tl, Hagendorf, H., & van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology, 76,* 1-8.

Object size judgements are affected by tactile and proprioceptive stimulation

Dermott Lynott (1), Louise Connell (1), & Felix Dreyer (2)
(1) UNIVERSITY OF MANCHESTER, UK; (2) UNIVERSITY OF LEIPZIG, GERMANY

Which is bigger: a grapefruit or an apple? A mansion or a bungalow? People have little difficulty in making such conceptual size comparisons. It has long been assumed that during language comprehension people represent size properties of objects with a strong visual component, distinct from propositional representations (e.g., size(huge); Paivio, 1975).

However, there is more to size than visual appearance. Additional information about object size comes from physical interactions; our arms, hands and fingers provide touch and proprioceptive feedback when interacting with an object. Embodied cognition research argues that such body- specific information plays an important role in conceptually representing objects because cognition is grounded in the same neural systems that govern perception and action (e.g., Barsalou, 1999). A strong interpretation of such theories would assume that making conceptual size judgements requires that past experiences across various modalities (visual, motor, tactile, proprioceptive, etc.) are re-enacted (or simulated) during comprehension, and the resulting object simulations then compared. Simulating non-visual information, however, requires being able to interact physically with the relevant objects. We investigated whether tactile and proprioceptive representations are independently involved in conceptual evaluations of size, and whether such representations are limited to objects of a physically manipulable size.

In Experiment 1, participants saw pairs of object names on-screen and were asked to name aloud the object that was bigger/smaller in real life (bigger/smaller judgements were counterbalanced across participants). For each trial, both objects were either of small size (i.e., can be held in one hand, such as GRAPE/EGG) or large size (i.e., greater than arms' width, such as ELEPHANT/WHALE). Critically, participants made these judgements while experiencing concurrent perceptual stimulation of the sense of touch: in one block, cushions under the hands vibrated (providing a tactile percept), while in another block, cushions under the feet vibrated (providing a control of equivalent sensory distraction). Results showed that, compared to the control condition, tactile stimulation to the hands led to faster size judgements for small objects, but made no difference to the judgement times for large objects. This difference occurred regardless of whether participants made "bigger" or "smaller" judgements for the object pairs.

Experiment 2 was identical but with concurrent perceptual stimulation of the sense of proprioception. Participants made size judgements while holding an inflatable beachball (30cm diameter) between their hands (providing isometric proprioceptive feedback: stable muscular tension during passive holding, without change in the length of muscle fibres) or between their lower legs (providing a control of equivalent sensory distraction). Results mirrored those of Experiment 1, with proprioceptive stimulation of the hands leading to faster judgements for small objects only, with no effect on judgements of large objects.

These novel findings validate embodied assumptions that both tactile and proprioceptive information are involved in conceptual size judgements. Furthermore, such body-specific information only affects judgements about objects that are small enough to be manipulated by the hands.

References:
Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-609.
Paivio, A. (1975). Perceptual comparison through the mind's eye. *Memory & Cognition, 3*, 635-647.

Multiple speech codes for imitative learning from pseudowords to words

Sejin Yoo
SEOUL NATIONAL UNIVERSITY, KOREA

In the view of sensorimotor integration, speech production may modulate the auditory system during speech perception, and at the same time, it may be shaped by feedback control of perceived speech (Hickok et al., 2011). The motor-centric view of speech processing also suggests a significant interaction between both ends (Galantucci et al., 2006). In this vein, we hypothesized that speech codes might be differentially managed by the specific content of auditory sounds. To test the hypothesis in on-line speech processing, we contrasted verbal repetitions of melodic tones, pseudowords, and words, each of which is tantamount to sound imitation, phonology repetition, and word recitation at different linguistic levels, respectively. We asked subjects to repeat immediately what they heard in MR scanner and investigated BOLD signal changes by different stimuli.

From fMRI data, the audition-articulation interface common to all stimuli was identified at bilateral Sylvian fissures and superior temporal sulci. We also found some regions in each lobar area selectively activated by specific stimulus only, implying differential recruitment of the acoustic, phonological, and semantic representations in the linkages with articulation. The superior frontal and parietal regions were bilaterally involved in the sound imitation of melodic tones. Interestingly, contrasting the stimuli revealed activities unique to word repetition at temporoparietal regions (left posterior middle temporal and bilateral inferior parietal areas) and those unique to pseudoword at frontal regions (left inferior frontal gyrus and anterior parts of right inferior and middle frontal gyri). It is notable that the frontal-versus-temporoparietal dissociation between pseudowords and words was reproduced in another experiment in which phonologically identical stimuli were repeated either as pseudowords or as words by subjects' perception.

These findings indicate that speech processing is multi-faceted and involves specialized speech-code subsystems. While the auditory-motor loop around the Sylvian fissure implements the basic acoustic-articulatory linkage, distinct brain networks subserve additional requirements entailed by phonological and semantic processes. This notion is consistent in part with the dual-stream model that distinguished articulation-based and acoustic-phonetic codes in speech processing (Hickok and Poeppel, 2004). In contrast to the acoustic-phonetic codes reserved for words, the articulation-based codes for pseudowords are likely to be maintained by the analysis-by-synthesis. The distinct speech-code subsystems may underlie imitative learning of vocabulary (Iacoboni, 2005) and participate in different modes of auditory communication which range from simple imitation of acoustic signals to more complex verbal exchange using phonological and semantic representations.

References

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). "The motor theory of speech perception reviewed." *Psychonomic bulletin & review* **13**(3): 361-377.

Hickok, G., Houde, J., and Rong, F. (2011). "Sensorimotor integration in speech processing: computational basis and neural organization." *Neuron* **69**(3): 407-422.

Hickok, G. and Poeppel, D. (2004). "Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language." *Cognition* **92**(1-2): 67-99.

Iacoboni, M. (2005). "Neural mechanisms of imitation." *Current opinion in neurobiology* **15**(6): 632-637.

Modeling situated language learning in early childhood via hypernetworks

Byoung-Tak Zhang, Eun-Seok Lee, and Min-Oh Heo
SEOUL NATIONAL UNIVERSITY, SEOUL KOREA

Human language is grounded, i.e. embodied and situated in the environment (Barsalou, 2008; Zwaan & Kaschak, 2008), and grounded language models rely on multimodal sensory data, such as gaze and gestures (Knoeferle & Crocker, 2006; Spivey, 2007; Yu et al. 2008). From the computational point of view, language grounding requires a flexible modeling tool that can deal with high-dimensional data. Here we explore a machine learning-based computational modeling method for situated language learning. The crux of the method is the hypernetwork model (Zhang, 2008), a probabilistic graphical model that learns multimodal associative memory incrementally from multisensory data.

We report on two sets of situated-language experiments simulating language acquisition in early childhood. We collected 10 series of cartoon videos for children of ages 3 to 7 and use them as a surrogate for large-scale multimodal experiment data. The first experiment uses dialogue sentences only, while the second experiment uses the visual scenes aligned to the sentences. The hypernetwork language model is learned serially on the videos of increasing order of age. We analyzed the concept maps (Griffiths et al., 2007) constructed by the probabilistic hypernetwork model to investigate the evolution of the concepts (or "visually-grounded" concepts in the second set of experiments). Using the generative property of the hypernetworks we generated the sentences and "mental" images from the learned model and analyzed them to study the evolution of grounded linguistic concepts as learning experience proceeds.

We found that the complexity of the sentences grows and more complex concepts emerge from simple ones as the learner observes diversified situations, as expected in language acquisition of infants and children (Bowerman & Levinson, 2001). Grammatical sentences were generated even though learning used neither lexical nor part-of-speech information. The *post hoc* analysis of the structure of the generated sentences showed the emergence of grammar rule-like patterns in the language model. Similar evolution of the multimodal visuo-linguistic concept maps was observed in the visually-grounded language learning experiments, though with a higher level of uncertainty due to the high noise ratio in the image data.

References:

Barsalou, L. W. (2008). Grounded cognition, Ann. Rev. Psych., 59, 617-645.
Griffiths, T., Steyvers, M., & Tenenbaum, J., (2007) Topics in semantic representation, Psych. Rev., 114(2), 211–244.
Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye-tracking, Cog. Sci., 30, 481-529.
Spivey, J. M. (2007). The Continuity of Mind, Oxford Univ. Press.
Yu, C, Smith, L. B., & Pereira, A. F. (2008). Grounding word learning in multimodal sensorimotor interaction, CogSci-2008, pp. 1017-1022.
Zhang, B.-T. (2008). Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory. IEEE Comp. Intell. Mag., 3(3), 49-63.
Zwaan, R. A. & Kaschak, M. P. (2008). Language in the brain, body, and world. Chap. 19, Cambridge Handbook of Situated Cognition.

Interaction of language and vision memories in TV drama watching: An EEG study

Choong-Yeon Lee , Eun, Sol Kim, Joon-Shik Kim, & Byoung-Tak Zhang
SEOUL NATIONAL UNIVERSITY, KOREA

Two memory retrieval sessions involving language-vision multimodal cues and their presentation order were studied using the multimodal memory game (MMG) platform. One session relied on script-to-scene cues and the second on scene-to-script cues. In this setting, the participant views an episode of TV drama and chooses a scene or script mostly associated with the presented one. During the game, we measure the brain activity of the players using electroencephalogram (EEG).

The MMG result shows that the retrieval performance is better in scene-to-script sessions but requires more response time than in script-to-scene sessions. This phenomenon can be explained based on the property of an image implicating diverse possibilities of interpretation while a text gives a more direct meaning. In other words, a language-dominant retrieval cue is faster but its retrieval accuracy might be lower than a vision-based cue because language has a narrower scope for memory searching. Besides, EEG analysis shows that there is an interaction between the frontal lobe and the occipital lobe while participants are playing the MMG multimodal memory game. The activation of the occipital lobe is higher in scene-to-script than script-to-scene sessions. These experimental results suggest that the brain system evokes visual imagery through the interactivity between brain areas of memory retrieval and visual processing.

The behavioural and EEG findings on the cross-modal interaction of language and vision in drama watching settings may shed light on embodied and situated cognition in language processing and its connection to mental imagery.

References:

Barsalou, L. W. (2008). Grounded Cognition, *Ann. Rev. Psy.*, 59, 617–645.
Barthes, R. (1977). Image-Music-Text, Fontana Press, London.
Buckner, R. L., Raichle, M. E., Miezin, F. M. & Petersen, S. E. (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures, *J. Neurosci.*, 16(19):6219–6235.
Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research, *JVLVB*, 11(6):671–684.
Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S. J. & Dolan, R. J. (1996). Brain activity during memory retrieval: The influence of imagery and semantic cueing, *Brain*, 119:1587–1596.
Hubel, D. H. & Wiesel, T. N. (2005). Brain and visual perception: The story of a 25-year collaboration, Oxford University Press.
Nyberg, L., Habib, R., McIntosh, A. R. & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval, *PNAS*, 97(20):11120–11124.
Ratcliff, R. & McKoon, G. (1988). A retrieval theory of priming in memory, *Psych. Rev.*, 95:385–408.
Tulving, E., Kapur S., Craik F. I., Moscovitch M., & Houle S. (1994). Hemispheric encoding / retrieval asymmetry in episodic memory: Position emission tomography findings, *PNAS*, 91:2016–2020.