

Wissenschaftliches Arbeiten in der germanistischen Linguistik (Technische Übung)

5220008 BA Germanistische Linguistik, Modul 4: Text und Diskurs 1
2 SP, Die 16-18, DOR 26, 207

Prof. Manfred Krifka

Institut für deutsche Sprache und Linguistik
Humboldt-Universität zu Berlin

krifka@rz.hu-berlin.de

<http://amor.rz.hu-berlin.de/~h2816i3x/lehrstuhl>



Einführung und Überblick

Koordinaten

- Sprechstunde Mi 13-15, bitte bei Frau Klein anmelden, und nach Vereinbarung
- E-Mail: krifka@rz.hu-berlin.de, bitte “Wissenschaftliches Arbeiten” in Subjekt-Zeile
- Büro: Dorotheenstr. 24, 3.303
- Sekretärin: Anina Klein Büro: 3.306, Tel. 030-20939639, E-Mail: anina.klein@cms.hu-berlin.de
- Moodle-Seite des Kurses:
<https://moodle.hu-berlin.de/course/view.php?id=73324> Schlüssel: **Panini**
- **Nicht zu verwechseln mit:**
Moodle-Seite “Technische Hilfsmittel für die Sprachwissenschaft”
<https://moodle.hu-berlin.de/course/view.php?id=50448>
- Bitte bei der Anmeldung zum Moodle-System ein Foto verwenden
– und zwar eines, das Sie zeigt, und zwar Ihr Gesicht (von vorne)!

Leistungsnachweis

- Fünf Moodle-Tests, die innerhalb einer Woche bearbeitet werden müssen;
Beurteilung: Sehr gut – Bestanden – Nicht bestanden;
- Wird ein Test als Nicht Bestanden bewertet, gibt es am Ende einen Zusatztest,
der als Bestanden bewertet werden muss.
- Teilnahme als Versuchsperson an Experimenten (3 Teilnahmepunkte),
am Institut für deutsche sprache und Linguistik oder am Zentrum für Allgemeine Sprachwissenschaft

Kursüberblick

24.10.	Einführung, Wissenschaft und Wissenschaftskommunikation. Literatur-Recherche: Bibliotheken, Online-Recherche, Bibliographieprogramme	
31.10.	Textverarbeitung I	
07.11.	Textverarbeitung II	Erster Test
14.11.	Seminararbeiten, Referate, Abschlussarbeit	
21.11.	Wissenschaftliche Argumentation am Beispiel eines linguistischen Artikels	Zweiter Test
28.11.	Linguistische Korpora: Einführung und Überblick	
05.12.	Entfällt	
12.12.	Arbeit mit linguistischen Korpora (Transkription (ELAN), Phonetik (PRAAT) – entfällt diesmal)	Dritter Test
09.01.	Empirische Untersuchungen: Befragungen, Experimente	
16.01.	Beschreibende Statistik: Darstellung von Resultaten	Vierter Test
23.01.	Schließende Statistik: Testverfahren I	
30.01.	Schließende Statistik: Testverfahren II	
06.02.	Schließende Statistik: Testverfahren III	Fünfter Test
13.02.	Wissenschaftliche Argumentation, Beispiel: psycholinguistischer Artikel	

Ziele der technischen Übungen:

Die inhaltlichen Kursen (z.B. Grundlagen der Linguistik) sollen in die Sprachwissenschaften einführen.

Die Übungen zum wissenschaftlichen Arbeiten soll diese Kurse begleiten durch

- Hinweise zum eigenständigen Umgang mit wissenschaftlichen Literatur
- Anweisungen zum Schreiben von Seminar- und Abschlussarbeiten, von Handouts und Präsentationen.
- Vorstellung von technischen Werkzeugen und Einführung in den Umgang mit diesen.
- Bereitstellung der theoretischen Grundlagen für eigenständige empirische Untersuchungen, deren Darstellung und statistische Auswertung.

Hierzu wurde die stationäre Moodle-Seite

„Technische Hilfsmittel für die Sprachwissenschaft“ eingerichtet.

Wissenschaft und Wissenschaftskommunikation

Wissenschaftliche Methode

Hier:

Keine Einführung in die Wissenschaftsphilosophie oder Wissenschaftssoziologie, vgl. hierfür z.B. Wolfgang Balzer, *Die Wissenschaft und ihre Methoden*, (2002).

Dennoch:

Eine Darstellung der wissenschaftlichen Methode.

Wir beginnen nicht mit den alten Griechen, sondern mit **Abu Ali al-Hasan ibn al-Haitham** besser bekannt als **Alhacen** (965-1039).

zahlreiche wissenschaftliche Werke, z.B. zur Optik, zum Aufbau des menschlichen Auges.

Daneben gilt er als der Begründer der wissenschaftlichen Methode.

Als abendländischer Begründer gilt Roger Bacon (1214-1292).

Leseempfehlung: Artikel „Scientific Method“ in der englischsprachigen Wikipedia.



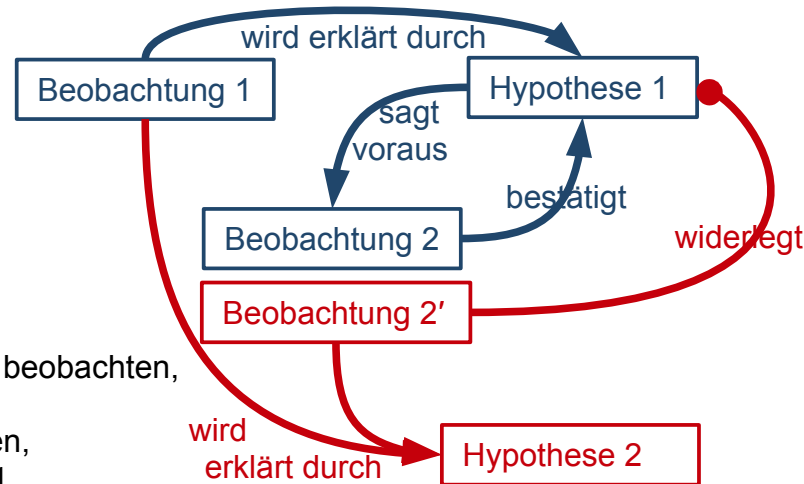
Grundidee der wissenschaftlichen Methode:

Zyklus von Beobachtung, Hypothese, Beobachtung

- Wir beobachten ein Phänomen B1.
- Wir versuchen B1 mit einer Hypothese H1 zu erklären.
D.h., wir halten den folgenden Schluss für plausibel: Wenn H1 gilt, dann tritt B1 auf.
- Wir versuchen, aus H1 weitere Folgerungen abzuleiten.
D.h., wir versuchen, hypothetische Beobachtungen B2 zu finden, für die gilt:
Wenn H1 gilt, dann sollte auch B2 gelten.
Wir machen also eine Voraussage.
- Wir versuchen, B2 zu beobachten
z.B. durch Experimente,
welche die Bedingungen herstellen,
unter denen nach H1 die Beobachtung B2
eintreten sollten.

Wenn wir tatsächlich B2 beobachten,
können wir dies als eine Bestätigung
der Hypothese H1 auffassen.

Wenn wir stattdessen ein anderes Phänomen B2' beobachten,
das nach H1 nicht zu erwarten wäre,
müssen wir H1 zu einer Hypothese H2 modifizieren,
aus der sowohl die ursprüngliche Beobachtung B1
als auch die Beobachtung B2' folgt.



Beobachtung vs. Regel

Beobachtungen und Hypothesen sind dabei wesensmäßig verschieden:

- Beobachtungen beziehen sich auf konkrete Ereignisse;
- Hypothesen stellen allgemeine Regeln dar, die vorhersagen, dass bestimmte Ereignisse unter bestimmten Bedingungen (mit einer spezifizierten Wahrscheinlichkeit) eintreten.

Beispiel: Voraussage einer Sonnenfinsternis nach Hypothesen über die Bewegung von Himmelskörpern, erstmals bei der Sonnenfinsternis von 585 v. Chr. durch Thales von Milet.

Intersubjektive Überprüfbarkeit und Offenheit

ist wesentlich für die Wissenschaft:

Die Beobachtungen müssen für andere Beobachter nachvollziehbar sein.¹

Bedingung hierfür:

Die wissenschaftliche Diskussion ist öffentlich,
sie wird möglichst herrschaftsfrei geführt:

- Wissenschaftler können fremde Beobachtungen und Hypothesen überprüfen
- Wissenschaftler können auf fremden Beobachtungen und Hypothesen aufbauen und die Wissenschaft weiterentwickeln.

Die öffentliche und freie Diskussion gehört zum Wesen der Wissenschaft selbst.

¹ Nicht zufällig lautet der Titel eines wissenschaftlichen Satiremagazins
The Journal of Irreproducible Results. <http://www.jir.com/>

Natur- vs. Geisteswissenschaften

Die wissenschaftliche Methode charakterisiert das Ideal der **Naturwissenschaften**, die das Ziel haben, beobachtete Phänomene nach allgemeinen Regeln zu erklären.

Trifft es auch auf die **Geisteswissenschaften** zu?

Wilhelm Dilthey (1883): *Einleitung in die Geisteswissenschaften*,

Ziel der Geisteswissenschaften:

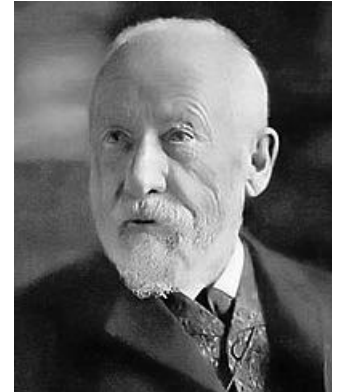
- **Verstehen** von **konkreten** individuellen geistigen Leistungen
- Verstehen von kulturellen Beziehungen und Entwicklungen
- Methode: Hermeneutik;
Annäherung durch Verstehen von handelnden und erfahrenden Subjekten und ihren kulturellen Hervorbringungen
- Experimentelle Überprüfung typischerweise nicht möglich.

Ziel der Naturwissenschaften:

- **Erklären** von **allgemeinen** naturgesetzlichen Kausalitäten und Zusammenhängen
- Methode: Hypothesenbildung und experimentelle Überprüfung
- Experimentelle Überprüfung typischerweise möglich, aber nicht immer (z.B. Astronomie).

Sprachwissenschaft zwischen Geistes- und Naturwissenschaften:

- Sprache als geistige Hervorbringung des Menschen, historisch situiert
- Aber auch: ein Teil der Biologie des Menschen, experimentellen Methoden oft zugänglich.



Der Gegenstandsbereich der Linguistik

Sprachwissenschaftler wenden sich ihrem Untersuchungsobjekt auf unterschiedliche Weise zu.

E-Language vs. I-Language

Noam Chomsky (1986, *Knowledge of Language*)

➤ **E-Language** (für: „extern“):

Die Erscheinungsformen des sprachlichen Verhaltens von Menschen, zum Beispiel schriftliche Texte, Ton- und Videoaufnahmen von sprechenden Menschen in konkreten Situationen

➤ **I-Language** (für: „intern“):

Die der E-Language zugrundeliegenden mentalen Fähigkeiten von Sprechern und Hörern.

Chomsky argumentiert, dass I-Language der Gegenstand der linguistischen Forschung sein soll; er hält die E-Language für zu beschränkt, von zufälligen Faktoren abhängig, um als Forschungsgegenstand von tieferem Interesse sein zu können.

Die Linguistik ist für ihn damit ein Teilgebiet der Psychologie (und weiter, der Biologie).

Dieser Auffassung stehen allerdings andere entgegen:

- E-Language kann ein wichtiges Instrument sein, um der I-Language auf die Spur zu kommen (z.B. Korpora von alten Sprachen als Evidenz für die I-Language ihrer Sprecher)
- E-Language kann ein unabhängig relevantes Studienobjekt sein, z.B. Einfluss von Kultur und Geschichte, von sozialen Strukturen auf die Sprache.
- Bedingungen, denen die E-Language unterliegt, können auf die I-Language zurückwirken.



Die drei Ebenen der Abstraktion

Innerhalb der Forschungsrichtung, die sich mit der I-Language befasst, gibt es eine weitere sehr einflussreiche Differenzierung, die auf David Marr zurückgeht (Marr & Poggio 1977), einem Neurowissenschaftlern, der diese für das Sehen entwickelt hat.

Wenn wir eine komplexe Fähigkeit wie die I-Sprache beim Menschen verstehen wollen, dann können wir unterscheiden zwischen:

- Eine **abstrakte Beschreibung** der Fähigkeiten eines kognitiven Systems, Beispiel: die Beschreibung der Grammatik einer Sprache.
- Die Angabe von präzisen **Algorithmen**, welche die in der abstrakten Beschreibung erfassten Strukturen erzeugen.
- Die Untersuchung, wie ein solcher Algorithmus im Gehirn des Menschen **realisiert** wird

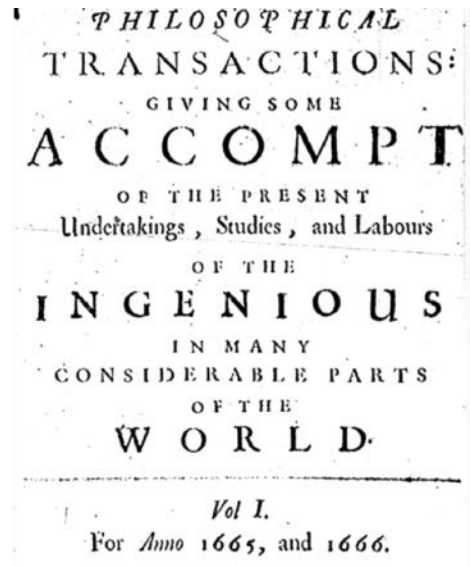
Das erste Ziel entspricht der **deskriptiven Grammatik**, das zweite der **generativen Grammatik**, und das dritte der **Psycholinguistik** und **Neurolinguistik**.



Geschichte der Wissenschaftskommunikation

Die wissenschaftliche Kommunikation hat sich erst nach und nach zu einer öffentlichen Kommunikation entwickelt.

- Frühzeit, z.B. die Schule der Pythagoräer, 6. Jhd. v. Chr.:
Geheimhaltung von Wissen innerhalb einer verschworenen Gemeinschaft.
- Öffnung der Wissenschaften in den Universitäten
- Private Kommunikation von Wissenschaftlern in Briefen
- Bildung wissenschaftlicher Akademien
(Leopoldina 1652, Royal Society 1660)
- Veröffentlichungen wissenschaftlichen Zeitschriften
(1665: *Journal de Scavants*, *Transactions of the Royal Society*).
- Veröffentlichungen in Büchern:
Monographien, Lehrbücher, Enzyklopädien



Sprachwissenschaften

Sprachwissenschaftliche Gesellschaften

Es gibt zahlreiche sprachwissenschaftliche Gesellschaften; hier einige wichtige:

- Deutsche Gesellschaft für Sprachwissenschaft (DGfS),
<http://www.dgfs.de/>
Jährliche Jahrestagungen, Sommerschulen
- Der Deutsche Germanistenverband (Hochschulgermanistik),
<http://www.germanistenverband.de/hochschule/>
- Societas Linguistica Europaea,
<http://www.societaslinguistica.eu/>
- Linguistic Society of America (LSA),
<http://www.lsadc.org/>

Die Jahrestagung 2016 der DGfS findet in Saarbrücken statt,

Webseite: <http://dgfs2017.uni-saarland.de/wordpress/>

Fragen:

- Was ist das Generalthema?
- Welche Arbeitsgruppen sind geplant?

Sprachwissenschaftliche Zeitschriften

Deutschsprachig:

- Zeitschrift für Sprachwissenschaft (ZS), das Organ der DGfS:
<http://www.degruyter.com/view/j/zfsw>
- Linguistische Berichte (LB): http://www.buske.de/index.php?cPath=4_13&content=recherche
- Zeitschrift für germanistische Linguistik (ZGL):
<http://www.degruyter.com/view/j/zfgl>
- Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB):
<http://www.mediaevum.de/zeitschriften/pbb4.php>,
<http://www.reference-global.com/loi/bgs1>

Aufgabe: Finden Sie heraus, welche Artikel in der jeweils letzten Nummer veröffentlicht sind.



Englischsprachig:

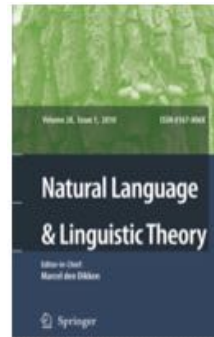
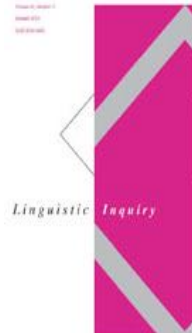
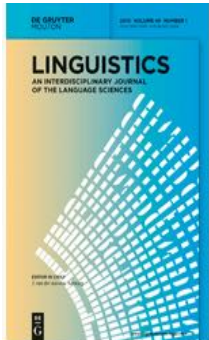
- Language (Lg), das Organ der LSA: <http://www.lsadc.org/info/pubs-language.cfm>
- Linguistics: <http://www.degruyter.com/view/j/ling>
- Linguistic Inquiry (LI): <http://www.mitpressjournals.org/loi/ling>
- Natural Language and Linguistic Theory (NLLT): <http://www.springer.com/linguistics/journal/11049>

In eigener Sache:

- Theoretical Linguistics (TL): <http://www.degruyter.com/view/j/thli>

Daneben gibt es Dutzende von Zeitschriften mit engerem Fokus.

- Finden Sie heraus, welche Artikel in der jeweils letzten Nummer veröffentlicht sind.
- Wenn Sie sich über das HU-Netz einloggen: Laden Sie die PDF-Datei eines Artikels, dessen Titel Ihr Interesse weckt, auf Ihren eigenen Computer und versuchen Sie, diesen zu lesen.



Lexika

Es gibt eine Reihe von linguistischen Lexika, die vor allem für das Nachschlagen von spezifischen Informationen geeignet sind:

- Hadumod Bußmann, *Lexikon der Sprachwissenschaft*. 4. Aufl.. Kröner, Stuttgart 2008.
<http://www.kroener-verlag.de/details/product/lexikon-der-sprachwissenschaft/>
- Helmut Glück, *Metzler Lexikon Sprache*, Aufl. 2016.
<http://www.springer.com/de/book/9783476026415>

Online-Lexika:

- Glottopedia, im Wiki-Format: <http://www.glottopedia.org/>

Neues Projekt:

- Wörterbücher der Sprach- und Kommunikationswissenschaft, Verlag Walter de Gruyter,
<http://www.degruyter.com/view/serial/16647>



Handbücher

Bei Handbüchern handelt es sich um Werke mit umfangreicheren Artikeln zu bestimmten Spezialgebieten der Sprachwissenschaft.

Sie erlauben es, sich einen Überblick über die Forschungssituation auf einem Gebiet zu verschaffen.

Beispiel: Die Reihe *Handbücher der Sprach- und Kommunikationswissenschaft*, <http://www.reference-global.com/action/showBookSeries?seriesCode=hsk>

Behandelte Themen:

Dialektologie, Sprachgeschichte, Soziolinguistik, Computerlinguistik, Korpuslinguistik, Semantik, Sprachphilosophie, Linguistic Disorders and Pathologies, Syntax, Schrift und Schriftlichkeit
Namenforschung, Kontaktlinguistik, Semiotik, Fachsprachen, Text- und Gesprächslinguistik, Morphologie, Deutsch als Fremdsprache, Sprachtypologie, Lexikologie, Psycholinguistik, Übersetzen, Quantitative Linguistik, Phraseologie.

Über die Webseite der Zentralbibliothek der HU sind die Inhalte der HSK für HU-Angehörige frei zugänglich, z.B.

<https://www.degruyter.com/view/product/176368?rskey=8qiQnZ&result=2>



Wissenschaftliche Monographien und Sammelwerke

Monographien zu bestimmten Themen geschrieben; in der Regel länger als Zeitschriftenartikel.

Sammelwerke versammeln verschiedene Artikel, oft zu einem Thema oder einem Themenkomplex, und oft auch von verschiedenen Autoren.

Tagungs-Sammelbände (Proceedings) sind Sammelwerken ähnlich; sie versammeln Artikel, die auf der Basis von Konferenzbeiträgen verfasst wurden (s.u.).

Wissenschaftliche Qualifikationsarbeiten

Nicht selten finden sich große wissenschaftliche Fortschritte gerade in linguistischen Qualifikationsarbeiten, als da sind

- Dissertationen zum Erlangen des Doktorgrades
- Master-, Magister-, Diplom-, Bachelor- und Zulassungsarbeiten (“Theses”)
- Seminararbeiten (über die noch zu reden sein wird).

In Deutschland müssen Dissertationen veröffentlicht werden.

Dies ist in vielen anderen Ländern nicht der Fall, wo es genügt, sie in Bibliotheken zu deponieren.

Dissertationen der USA sind ferner seit 1938 per Mikrofilm verfügbar (siehe UMI, <http://dissexpress.umi.com/dxweb/search.html>).

Viele Dissertationen sind heute über das Internet von Webseiten abrufbar.

Zentral geführte Webseiten streben eine Kontinuität der Präsentation an.

Publikationsserver HU: <http://edoc.hu-berlin.de/browsing/dissertationen/>

Bibliographien

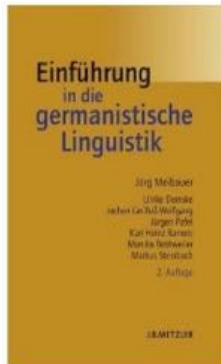
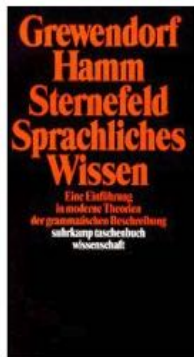
Diese Werke sammeln und ordnen Literaturhinweise zu bestimmten Themen oder für das Fachgebiet allgemein.

Besonders hilfreich sind hierbei kommentierte Bibliographien.

Die große Bibliographie ist die *Bibliographie linguistique*, in jährlichen Jahresbänden und übers Internet verfügbar: <http://bibliographies.brillonline.com/browse/linguistic-bibliography>

Lehrbücher

Wie der Name sagt, sind Lehrbücher für die Lehre bestimmt und in der Regel nicht der Ort, an dem man neue wissenschaftliche Erkenntnisse findet. Mit diesen Werken sind Sie im Verlauf Ihres Studiums am meisten konfrontiert.



Populärwissenschaftliche Sachbücher

Populärwissenschaftliche Sachbücher sind an einen allgemeinen Leserkreis gerichtet – aber durchaus auch für Studenten des Fachgebiets empfehlenswert!

Der populärste Autor ist Steven Pinker mit Werken wie

- *The Language Instinct / Der Sprachinstinkt*, 1994 (Deutsch: Droemer-Knaur, 1998)
- *Words and Rules, the ingredients of language / Wörter und Regeln, die Natur der Sprache*, (Deutsch: Spektrum Verlag, 2000).
- *The Stuff of Thoughts*, 2008.

Ferner zu erwähnen:

- David Crystal, *The Cambridge Encyclopedia of Language / Die Cambridge-Enzyklopädie der Sprache*, 1993.
- Hans-Joachim Störig, *Abenteuer Sprache – Ein Streifzug durch die Sprachen der Erde*, dtv 2002.
- Ilse Achilles und Gerda Pighin, *Vernäht und zugefixt. Von Versprechern, Flüchen, Dialekten und Co.* Mannheim: Duden-Verlag, 2008.



Konferenzen, Sommerschulen, Diskussionsforen

Es gibt zahlreiche Konferenzen – einige davon in regelmäßigem Turns stattfindend, andere zu bestimmten Gelegenheiten. Beispiele:

- Jahrestagung der DGfS (24.-26. Februar 2016 in Konstanz)
- Jahrestagung des Instituts für Deutsche Sprache (8.-10. März 2015 in Mannheim)
- Annual Meeting, Linguistic Society of America (7.-10. Januar. 2016, Washington D.C.)
- International Congress of Linguists (CIL), alle 5 Jahre

Über aktuelle Veranstaltungen in Berlin und Potsdam informieren die folgenden Webseiten:

- <http://hpsg.fu-berlin.de/calendar/month.php>
- <http://www.zas.gwz-berlin.de/>

Neben Konferenzen gibt es verschiedene Sommerschulen zu erwähnen:

- Die Sommerschule der LSA (alle 2 Jahre),
- Sommerschule der DGfS (alle 2 Jahre), 2013 an der HU Berlin
- Sommerschule der GLOW (“Generative Linguistics in the Old World”
- European Summer School of Language, Logic and Information ESSLLI, August 2016, Bozen

Sehen Sie sich die Webseiten der jüngsten Sommerschulen an.

Wann und wo gibt es die nächsten Sommerschulen, und zu welchen Themen?

Das wichtigste Internet-Diskussionsforum ist die *Linguist List*. Hier gibt es Diskussionen zu speziellen Themen, Umfragen, Buchbesprechungen, Hinweise auf Konferenzen, Stellenangebote, Software etc.: <http://www.linguistlist.org/issues/> Machen Sie sich mit der Linguist-List vertraut. Lesen Sie einige Diskussionsbeiträge.

Literatur-Recherche

Relevante Bibliotheken in Berlin

Sehen Sie sich das Angebot der Bibliotheken selbst an!

Die Liste erhebt keinen Anspruch auf Vollständigkeit. Vielmehr werden Bibliotheken aufgeführt, die sich bei der Recherche sprachwissenschaftlicher Literatur als hilfreich erwiesen haben.

Zentralbibliothek der Humboldt-Universität (ZUB) Jakob-und-Wilhelm-Grimm-Zentrum

Neu zugänglich Oktober 2009!

Homepage: <http://www.ub.hu-berlin.de/>

Online-Katalog: http://opac.hu-berlin.de:80/F/?func=file&file_name=find-b

Anmeldung/Ausleihe:

Für Studenten der HU funktioniert dies über den Studierendenausweis
(= Bibliotheksausweis durch Barcode).

Genauere Informationen hierzu auf der Homepage und an den Bibliotheksschaltern.



Portal der Zentralbibliothek

Anmelden

Benutzernr.:

Passwort:

[Passwort vergessen?](#) **Login**

[Suchen und Kataloge](#) [Standorte](#) [Fächer](#) [Öffnungszeiten](#) [Kontakt](#) 



- Literatur suchen
- Recherchieren lernen
- Bibliothek benutzen
- Über uns

Primus  [erweiterte Suche](#)

Ich bin hier: Startseite

Universitätsbibliothek der Humboldt-Universität zu Berlin

-  **UB Blog**
-  **Online Tutorials**
-  **Humboldt-UB App**
-  **eBooks on Demand**
-  **Ich möchte ...**

UB Blog

08.10.2015

**Ein herzliches Willkommen an alle unsere
Erstsemester und Campusneulinge in Mitte und
Adlershof!**

Wir erleichtern Ihnen die Orientierung in Ihrem neuen Umfeld mit einem umfassenden Führungs- und Schulungsangebot an unseren verschiedenen Standorten. Für Erstnutzer der UB empfiehlt sich unsere [Rechercheeinführung](#). [Recherchieren lernen](#) mit Hilfe

Aktuelle Meldungen



**28.9.-30.10.15: Verkürzte
Öffnungszeiten der
Ausleihtheken im Grimm-
Zentrum: 9-18 Uhr** [Mehr »](#)





UB-Jahresbericht 2014 online [Mehr »](#)

Angebote der Zentralbibliothek

Die Zentralbibliothek der HU hält zahlreiche Angebote bereit.

Bitte jetzt anschauen:

- Die UB kennenlernen
- Das Internet nutzen
- BLL (Linguistik)
- MLA (Sprache / Literatur)
- BDSL (Germanistik)

 Online-Kurs Informationskompetenz	 LOTSE Video Tutorials
Mithilfe des E-Learning-Angebotes lernen sie den Umgang mit den Diensten und Recherchemöglichkeiten ihrer Universitätsbibliothek.	LOTSE und die UB präsentieren in den Video-Tutorials eine unterhaltsame allgemeine Einführung in wissenschaftlichen Recherchestrategien.
<ul style="list-style-type: none">■ Die UB kennenlernen<ul style="list-style-type: none">■ Bibliothekssystem■ Benutzerkonto■ Recherche vorbereiten<ul style="list-style-type: none">■ Informationsbedarf formulieren■ Informationsmittel wählen■ Suchstrategie entwickeln■ Recherche durchführen<ul style="list-style-type: none">■ Primus■ Online-Katalog■ KOBV / Fernleihe■ ZDB / Fernleihe■ EZB■ DBIS■ MLA■ Online-Contents■ BLL■ BDSL (Germanistik)■ Information verarbeiten<ul style="list-style-type: none">■ Zitieren■ Transliterieren (Slawistik)■ Literaturverwaltung mit Citavi	<ul style="list-style-type: none">■ Das Internet nutzen<ul style="list-style-type: none">■ Einstiege in das Internet - Quiz■ Deep Web - Quiz■ Suchstrategien<ul style="list-style-type: none">■ Vom Thema zu den Suchbegriffen - Quiz■ Von den Suchbegriffen zur Literatur - Quiz■ Datenbanken / Fachbibliographien<ul style="list-style-type: none">■ Warum sind Datenbanken und Fachbibliographien nützlich? - Quiz■ Suche in Datenbanken: Wesentliche Schritte■ Extras und Bearbeitung von Trefferlisten■ Richtig Zitieren und Plagiate vermeiden<ul style="list-style-type: none">■ Plagiate - Quiz■ Zitierregeln - Quiz <p>Die Skripte zu den Tutorials liegen auch im PDF-Format zum Nachlesen vor.</p>

Zweigbibliothek Germanistik der HU (ZwB Germanistik)

Anschrift: Dorotheenstr. 24 (DOR 24)

Online-Katalog: Über Online-Katalog der ZUB

Die ZwB Germanistik besteht aus den Teilbibliotheken (TB) Skandinavistik und Historisch-Vergleichende Sprachwissenschaft.

Zweigbibliothek Fremdsprachliche Philologien der HU (ZwB Fremdsprachl. Philo.)

Dorotheenstr. 65 (DOR 65) und Unter den Linden 6 (UL 6 → Hauptgebäude).

Online-Katalog: Über Online-Katalog der ZUB

Anmeldung/Ausleihe: Vgl. ZwB Germanistik

Weitere Informationen: TB verteilen sich auf DOR 65 und UL 6 → siehe Homepage

Bibliothek des Zentrums für Allgemeine Sprachwissenschaft (ZAS)

Schützenstrasse 18, 2. Stock.

Homepage: <http://zas.gwz-berlin.de>

Online-Katalog: <http://opac.zas.gwz-berlin.de/index.php>

Es handelt sich um die Präsenzbibliothek eines linguistischen Forschungsinstituts. Zugang in der Regel 9 – 15 Uhr, linguistische Spezialliteratur, Grammatiken, Zeitschriften usw. Ca. 70,000 Bände.

Philologische Bibliothek der Freien Universität zu Berlin

Habelschwerdter Allee 45, Öffnungszeiten 9 – 22 Uhr, Sa/So 10 – 17 Uhr

<http://www.ub.fu-berlin.de/fbb/phil-geist/philobib.html>

Staatsbibliothek Berlin (Stabi)

Anschrift: Stabi Ost: Unter den Linden 8 (UL 8) → U6 Friedrichstr.

Stabi West: Potsdamer Str. 33 → S+U Potsdamer Platz

Homepage: <http://staatsbibliothek-berlin.de/>

Online-Katalog: <http://www.stabikat.de/>

Anmeldung/Ausleihe: Beantragung eines Nutzersausweises ist notwendig. Ausweis mit der Gültigkeit von einem Monat kostet 10 Euro und von einem Jahr 25 Euro. Ausweise sind bei Stabi Ost und West erhältlich.

Weitere Informationen: Auch wenn die Nutzung der Stabi nur gegen Gebühr möglich ist, so ist doch ein Großteil der Literatur (vor allem Zeitschriften) ausleihbar und kann mit nach Hause genommen werden. Außerdem deckt die Stabi erfahrungsgemäß ein großes Spektrum an Fachliteratur ab und führt Publikationen, die u.U. in den anderen Bibliotheken nicht zu finden sind.

Online-Angebote

Ähnlich wie bei den Berliner Bibliotheken lohnt sich auch hier das Stöbern auf den einzelnen Homepages, um einen Einblick in die zur Verfügung gestellten Leistungen zu bekommen.

Datenbank-Informationssystem der ZUB (DBIS) der HU

Über DBIS ist es möglich zu erfahren, welche Datenbanken es zu einem gesuchten Interessengebiet gibt und ob man auf die zugreifen kann.

Homepage: <http://rzblx10.uni-regensburg.de/dbinfo/fachliste.php?>

Elektronische Zeitschriftenbibliothek der HU (EZB)

Online-Zugang zu Zeitschriften

Homepage: <http://rzblx1.uni-regensburg.de/ezeit/fl.phtml?bibid=UBHUB>

Nicht auf alle angeführten Zeitschriften kann man zugreifen (die verfügbaren Zeitschriften werden durch das Ampelsymbol „Grün“ angezeigt). Von einem HU-Rechner (Bibliothek) kann man auf mehr Zeitschriften zugreifen. Am PC zuhause loggt man sich am besten über die HU ein, über die Webseite: https://ssl.cms.hu-berlin.de/dana-na/auth/url_default/welcome.cgi

Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)

Internet-Portal der Berliner und Brandenburgischen Bibliotheken, das die Suche nach Literatur in allen KOBV-Bibliotheken gleichzeitig ermöglicht. Der KOBV bietet einen Überblick über die regionalen Bibliotheken und Informationen zu regionalen und überregionalen Bibliotheksangeboten.

Homepage: <http://www.kobv.de/>

Zeitschriftendatenbank (ZDB)

Bietet die Möglichkeit der schnellen und effektiven Recherche in den Zeitschriftenbeständen deutscher Bibliotheken.

Homepage: <http://www.zeitschriftendatenbank.de/>

Deutsche Nationalbibliothek

Die Deutsche Nationalbibliothek (Frankfurt und Leipzig) sammelt alle öffentlichen deutschen Publikationen. Dies gilt auch für Dissertationen.

Homepage: <http://www.d-nb.de>

Karlsruher Virtueller Katalog (KVK)

Ein Meta-Katalog, bei dem die eingegebene Suchanfragen an mehrere Bibliothekskataloge gleichzeitig weitergeleitet wird. Es wird dann eine Trefferliste mit der Referenz auf den jeweiligen Katalog angezeigt.

Homepage: <http://www.ubka.uni-karlsruhe.de/kvk.html>

Suchmaschinen

Heute werden allgemeine und spezialisierte Suchmaschinen für die Literatur-Recherche immer wichtiger. Die bekannteste Suchmaschine ist Google.

Bei der Eingabe von Suchbegriffen gilt:

- Suchmaschinen suchen nach Wörtern, nicht nach Bestandteilen von Wörtern.
- Mehrere Suchbegriffe sind durch “und” verbunden, d.h. Webseiten müssen alle Suchbegriffe enthalten. Die Suche nach “Jacob Grimm” gibt ihnen alle Webseiten, die “Jacob” und “Grimm” enthalten.
- Sie können nach Wortketten suchen, wenn sie diese in Anführungszeichen setzen, z.b. “Jacob Grimm”
- Es wird Ihnen gesagt, wie viele Treffer die Suchmaschine gemacht hat. Die Treffer werden in einer Reihenfolge aufgelistet, die vor allem widerspiegelt, wie viele Links auf die betreffenden Webseiten verweisen.

Es gibt spezielle Webseiten, die für wissenschaftliche Recherche von Bedeutung sind:

- Google Scholar
- Microsoft Academic Search
- Google Print / Google Buchsuche
- Google Bilder
- Zur feineren Abstimmung der Suche: http://www.google.de/advanced_search?hl=en

Google Scholar

Eingeschränkte Suche auf wissenschaftliche Webseiten.

Funktion eines Zitationsindizes:

- Wie oft wurde ein Beitrag rezipiert? Rückschlüsse auf Rezeption und Bedeutung.
- Wo wurde ein Beitrag rezipiert? Hinweise auf neuere Forschungsliteratur zum Thema.

Homepage: <http://scholar.google.com/>

Beispiele von Suchen:

Google animacy "word order"

Scholar Ungefähr 5.850 Ergebnisse (0,05 Sek.)

Beliebige Zeit
Seit 2012
Seit 2011
Seit 2008
Zeitraum wählen...

Nach Relevanz sortieren
Nach Datum sortieren

Web-Suche
Seiten auf Deutsch


Patente einschließen
 Zitate einschließen

Tip: Suchen Sie nur nach Ergebnissen auf **Deutsch**. Sie können Ihre Sprache in den **Scholar-Einstellungen** festlegen.

[Contributions of **animacy** to grammatical function assignment and **word order** during production](#) [\[PDF\]](#) von posugf.com.br
Volltext
HP Branigan, MJ Pickering, M Tanaka - *Lingua*, 2008 - Elsevier
We examine how the relationship between **animacy** and syntactic structure might be explained in terms of an influence of **animacy** on the psychological processes that underlie the construction of syntactic structure during language production. In this account, ...
Zitiert durch: 56 Ähnliche Artikel Alle 6 Versionen

[A corpus study into **word order** variation in German subordinate clauses: **Animacy** affects linearization independently of grammatical function assignment](#) [\[PDF\]](#) von mpg.de
G Kempen, K Harbusch - *TRENDS IN LINGUISTICS STUDIES ...*, 2004 - books.google.com
The grammar of German does not impose hard constraints on the linear order of Subject (SB), Indirect Object (IO) and Direct Object (DO) in finite complement or adverbial clauses (for an overview of the linguistic literature see Miller 1999). All six possible orders are ...
Zitiert durch: 36 Ähnliche Artikel Alle 4 Versionen

Suche nach Personen, mit Zitationsindex



Hubert Truckenbrodt

Centre for General Linguistics (ZAS) Berlin
[Linguistic interfaces](#)
Bestätigte E-Mail-Adresse bei zas.gwz-berlin.de - [Startseite](#)

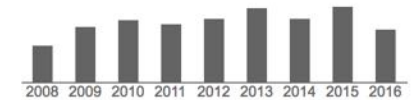
[Folgen](#)

Titel	1–20	Zitiert von	Jahr
On the relation between syntactic phrases and phonological phrases	H Truckenbrodt Linguistic inquiry 30 (2), 219-255	693	1999
Phonological phrases--their relation to syntax, focus, and prominence	H Truckenbrodt Massachusetts Institute of Technology	626 *	1995

Google Scholar

[Eigenes Profil erstellen](#)

Zitationsindexe	Alle	Seit 2011
Zitate	2645	1338
h-index	18	17
i10-index	28	25



Alternative: Microsoft Academic Search

<https://academic.microsoft.com/>

Fachbibliographien

sammeln wissenschaftliche Literatur zu bestimmten Fächern.

Webseite zu Fachbibliographien allgemein: Datenbank-Informationssystem DBIS, <http://rzblx10.uni-regensburg.de/dbinfo/fachliste.php?lett=> am besten über die Webseite der HU-Zentralbibliothek.

Besonders relevant in der DBIS-Übersicht:

- Germanistik, Niederlandistik, Skandinavistik
- Allgemeine und vergleichende Sprach- und Literaturwissenschaft

Wichtige Fachbibliographien:

- Linguistic Bibliography Online, <http://lb.brillonline.nl/>
- Bibliography of Linguistic Literature, <http://www.blldb-online.de/>
- Modern Language Association (MLA) International Bibliography, <http://www.mla.org/bibliography>
- Germanistik im Netz, <http://www.germanistik-im-netz.de/>
- Bibliographie der Deutschen Sprach- und Literaturwissenschaft, <http://www.bdsl-online.de/>

Web-Browser

Web-Browser

Es gibt verschiedene Web-Browser
(wie Internet Explorer, Firefox, Safari).

Empfohlen: **Firefox**, ein “offener” Browser.

<http://www.mozilla.com/en-US/firefox/upgrade.html>

Wichtig: Legen Sie ein gut organisiertes Verzeichnis von “Bookmarks” an,
das spart Zeit und Energie (buchstäblich, weil es die Suchanfragen etwa bei Google reduziert).
Ein wichtiges plug-in für die wissenschaftliche Arbeit werden wir später kennenlernen (Zotero).



PDF-Programme

Es gibt zwei wichtige Formate für textbasierte Web-Inhalte:

- Das html-Format, ein strukturiertes Format für Webseiten.
- Das pdf-Format, für schriftliche Dokumente allgemein.

Das html-Format versteht jeder Web-Browser;

Sie sollten allerdings sicherstellen, dass sie die jeweils neueste Version haben.

Das pdf-Format verstehen verschiedene Programme zum Lesen von Dokumenten, allen voran der Adobe Acrobat Reader.

PDF-Leser gibt es in zwei Versionen:

- Der kostenlosen einfachen Version, zum Lesen und auch zum Annotieren.
- Der Standard-Version, mit der man pdf-Dateien bearbeiten und erzeugen kann.

Die Annotationsmöglichkeiten sind wichtig, wenn man am Texte am Bildschirm liest und dort Hervorhebungen, Randbemerkungen usw. machen will.

Es gibt auch kostenlose Programme, die Annotationsmöglichkeiten bieten:

- Für Windows: PDFXVer
<http://www.download25.com/install/pdf-xchange-pdf-viewer.html>
- Für Mac OSX: Skim,
<http://sourceforge.net/projects/skim-app/>

Beispiel: Skim

contribution to studies of linguistically-expressed meaning.

Chapman portrays Grice as a "skillful heretic", but he emerges also as a talented mediator. He showed that neither the formalists nor the informalists had it quite right, but neither had it completely wrong either. In 'Logic and conversation', he sets out the idea that linguistic communication is rooted in conventions about meaning, but that speaker meaning often extends far beyond this. Chapman tracks the emergence of this subtle balance, which begins (like 'Logic and conversation' itself) with the relationship between the classical logical connectives and their usual natural language glosses and culminates in a full-blown theory of meaning.

Chapman recounts (p. 186) the story of the accidentally dual publication of 'Logic and conversation'. Davidson and Harman 1975 is reported to be Grice's preferred home for the work. Its appearance in Cole and Morgan 1975 might trace back to a long evening at a Texas bar. I confess to some dismay that Grice favored the Davidson and Harman 1975 publication. There, 'Logic and conversation' is

Note		Page
G	A	3
li	A	3
e	○	3
T	✎	3
lo	A	3
n	A	3
	→	3

Scannen und Optical Character Recognition

Es lohnt sich, Artikel nicht zu kopieren, sondern zu scannen.

Vorteil: Sie haben den Artikel auf dem Computer und können ihn in ihre Literaturdatenbank integrieren.

Diese Artikel liegen in einer Bitmap-Datei vor, d.h. es werden die einzelnen Bildpunkte kodiert. Dies gilt auch dann, wenn es sich um eine pdf-Datei handelt.

Dies betrifft auch ältere Artikel, die man z.B. von Internet-Repositorien herunterladen kann und die eingescannt wurden.

Nachteil: Bitmap-Dateien können nicht nach Text durchsucht und meist auch nicht annotiert werden.

Abhilfe: Sie unterziehen die Datei einer optischen Zeichenerkennung (OCR).

Dies ist z.B. mit Adobe Professional möglich (DOCUMENT OCR TEXT RECOGNITION)

Die Zeichenerkennung ist relativ gut, aber nicht perfekt.

Es kommt vor allem auf die Qualität der Vorlage an.

Bibliographie-Programme

Früher sammelte man bibliographische Angaben in Zettelkästen.



Heute gibt es verschiedene Bibliographieprogramme zur Literaturverwaltung:

- Endnote (kostenpflichtig), <http://www.endnote.com/>
(das am weitesten verbreitete Programm, relativ teuer)
- Citavi (eingeschränkte freie Version), <http://www.citavi.com/>;
Campus-Lizenz der HU, vgl.
http://info.ub.hu-berlin.de/onlinekurs/ik_modul_d/02_01_literaturverwaltungsprog_01.html
- Zotero (freie Version, siehe eigene Darstellung) <http://www.zotero.org/>
- Das freie Textprogramm LibreOffice, das wir später ausführlich behandeln, besitzt ein eingebautes Modul für eine bibliographische Datenbank.
- Bookends (für Macintosh, kostenpflichtig), <http://www.sonnysoftware.com/>

Hier stelle ich Bookends vor, das ich selbst verwende.

Literaturverarbeitung mit Zotero

Was ist Zotero?

Zotero ist eine Erweiterung des Firefox-Browsers, mit der man Online- und Offline-Quellen sammeln, verwalten und zitieren kann.

Als Freeware (die Entwicklung wurde von Stiftungen unterstützt) kann man sich Zotero herunterladen unter: www.zotero.org.

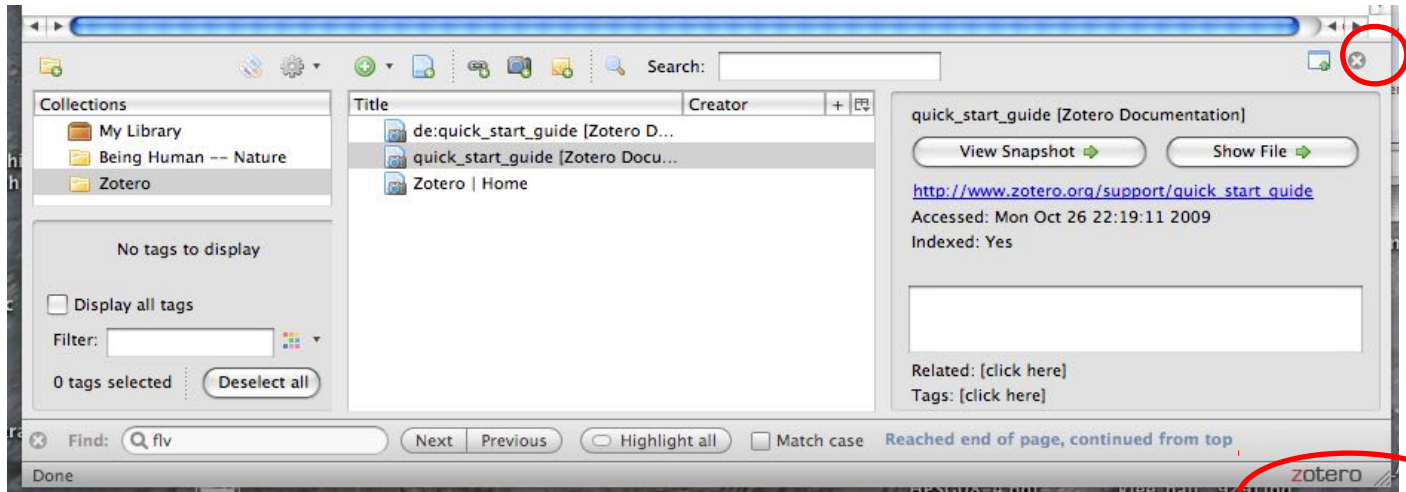
Deutsche Beschreibung: http://www.zotero.org/support/de/quick_start_guide

Was kann man mit Zotero machen?

- Sammeln und Archivieren von Information, vom Internet und anderen Quellen
- Organisieren von Informationen in Ordnern
- Kommentare und Beschreibungen von Informationen
- Festhalten von bibliographischen Informationen
- Erzeugen von Literatur- und Quellenlisten.

Einstieg in Zotero

Wenn Zotero installiert ist, kann man es vom Firefox-Webbrowser öffnen. Schließen des Fensters wie angegeben.



Das Zotero-Fenster hat drei Spalten:

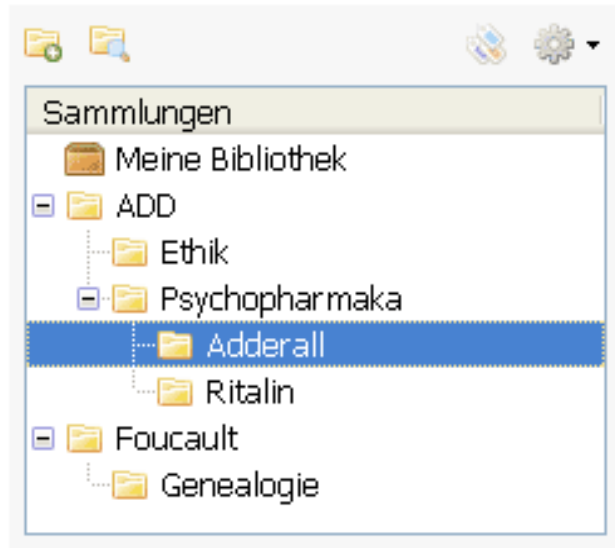
Links: Meine Bibliothek
Verzeichnis Sammlungen

Mitte: Einträge der
ausgewählten Sammlung

Rechts: Informationen
über ausgewählten Eintrag




Die drei Spalten von Zotero

Linke Spalte

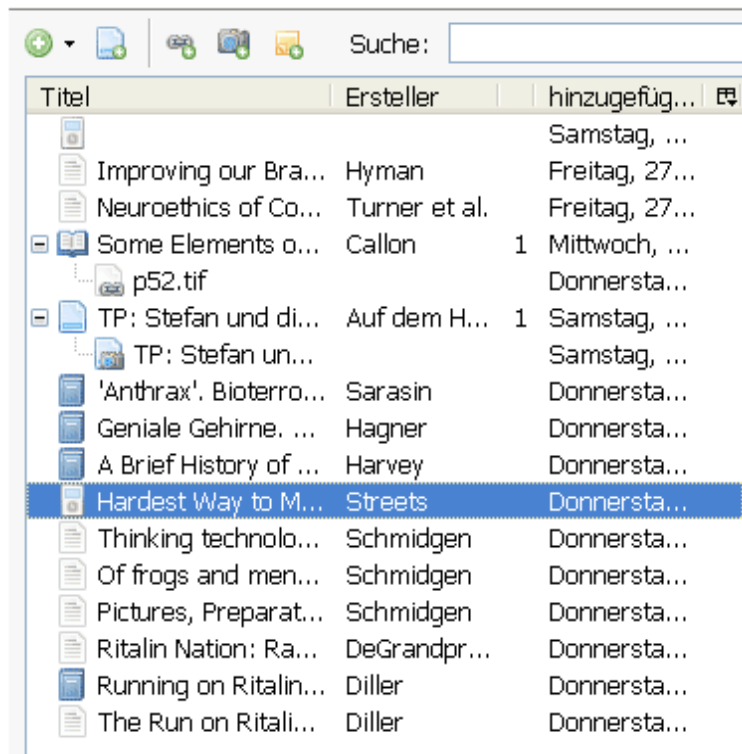


Die linke Spalte dient zur Gliederung der Sammlung in Verzeichnisse und Unterverzeichnisse.

SYMBOLE:

-  **neue Sammlung hinzufügen**
-  **Tags organisieren**
-  **Sammlung importieren/exportieren**

Mittlere Spalte

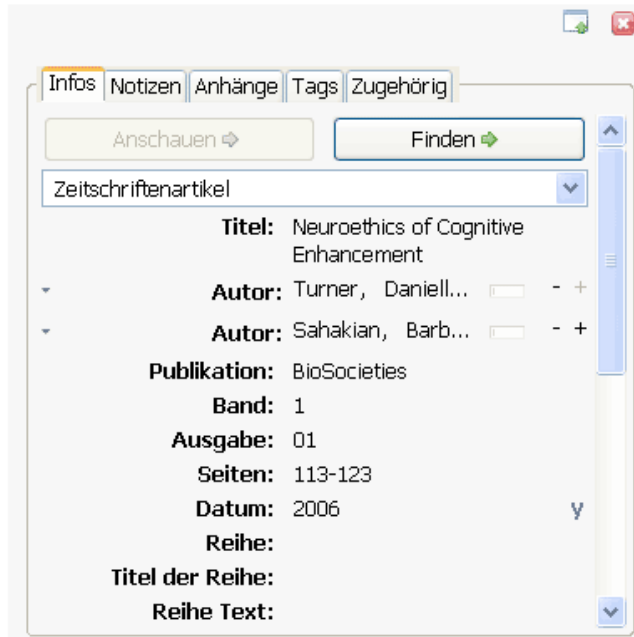


Titel	Ersteller	hinzugefügt...
		Samstag, ...
Improving our Bra...	Hyman	Freitag, 27...
Neuroethics of Co...	Turner et al.	Freitag, 27...
Some Elements o...	Callon	1 Mittwoch, ...
p52.tif		Donnersta...
TP: Stefan und di...	Auf dem H...	1 Samstag, ...
TP: Stefan un...		Samstag, ...
'Anthrax'. Bioterro...	Sarasin	Donnersta...
Geniale Gehirne. ...	Hagner	Donnersta...
A Brief History of ...	Harvey	Donnersta...
Hardest Way to M...	Streets	Donnersta...
Thinking technol...	Schmidgen	Donnersta...
Of frogs and men...	Schmidgen	Donnersta...
Pictures, Preparat...	Schmidgen	Donnersta...
Ritalin Nation: Ra...	DeGrandpr...	Donnersta...
Running on Ritalin...	Diller	Donnersta...
The Run on Ritali...	Diller	Donnersta...

-  **manuell Eintrag hinzufügen**
-  **Website in Sammlung speichern**
-  **Link zu Website (~Lesezeichen)**
-  **Schnappschuss von Website**
-  **einzelne Notiz hinzufügen**

Die mittlere Spalte enthält die Einträge des angewählten Ordners.

Rechte Spalte



Die rechte Spalte enthält bibliographische Informationen zu dem ausgewählten Eintrag.

SYMBOLE:

 **Vollbildmodus**

 **Zotero schließen**

Einige Eigenschaften von Zotero:

- Drag-and-drop, zum Beispiel von Bildern, Tondateien, Filmdateien.
- Bibliographische Datenbank: "New Item".
- Bei Webseiten mit bibliographischer Information: Durch Anklicken des Buch-Ikons oder des Artikel-Ikons in der URL-Leiste des Browsers kann man die bibliographische Information unmittelbar erzeugen.

- Buch (z.B. im Bibliothekskatalog)



- Bei Listen von bibliographischen Informationen auf einer Webseite kann man die ganze Liste oder Elemente davon übernehmen.

- Gruppe von Einträgen (z.B. Google Scholar)




- Bibliographische Referenzen kann man durch Drag-and-Drop in das Textverarbeitungsprogramm kopieren.
- Durch Ctrl+Klick einer Referenz oder von Referenzen kann man einen bibliographischen Eintrag nach ausgewählten Style-Sheets erzeugen.
- Für bestimmte Textverarbeitungsprogramme (MS Word) gibt es Plug-Ins zur Erzeugung von Bibliographien.

Literaturverarbeitung mit Zotero

Hier sollen die einzelnen Schritte systematisch dargestellt werden.

Angenommen, Sie suchen Literatur zu trennbaren Präfixen mit Google Scholar:



Scholar Ungefähr 1.190 Ergebnisse (0,06 Sek.)

Beliebige Zeit
Seit 2012
Seit 2011
Seit 2008
Zeitraum wählen...

Nach Relevanz sortieren
Nach Datum sortieren

Web-Suche
Seiten auf Deutsch

Patente einschließen
 Zitate einschließen

Alert erstellen

[\[PDF\] Von OV zu VO: ein Vergleich zwischen Zimbrisch und Plodarisch](#) [PDF] von '...
G Grewendorf, C Poletto - *Diversitas Linguarum*, 2005 - 157.138.7.88
... Page 18. 18 **trennbarer** Verpräfixe. Während **trennbare Präfixe** in VO-Sprachen in der Regel dem Partizip folgen (cf. ... **trennbarer** Verpräfixe zu **Trennbare Präfixe** könne hier nur nach dem Partizip auftreten: [Zimbrisch von Roana] (48) a. Häüte dar Gianni is gont-vort. ...
Zitiert durch: 15 [Ähnliche Artikel](#) [Alle 9 Versionen](#) [Mehr](#) ▾

[\[BUCH\] Handbuch der deutschen Grammatik](#)
E Hentschel, H Weydt - 2003 - [books.google.com](#)
... sich nur durch eine Einheit (hier: /k/ und /g/). Im Gegensatz dazu sind [k] und [k] sogenannte Allophone (von grie- chisch allos ‚anders beschaffen als‘ + phone ‚Lautq), dh unterschied- allo- dient in der sprachwissenschaftlichen Terminologie als **Präfix**, das Varianten einer ...
Zitiert durch: 365 [Ähnliche Artikel](#) [Alle 3 Versionen](#) [Mehr](#) ▾

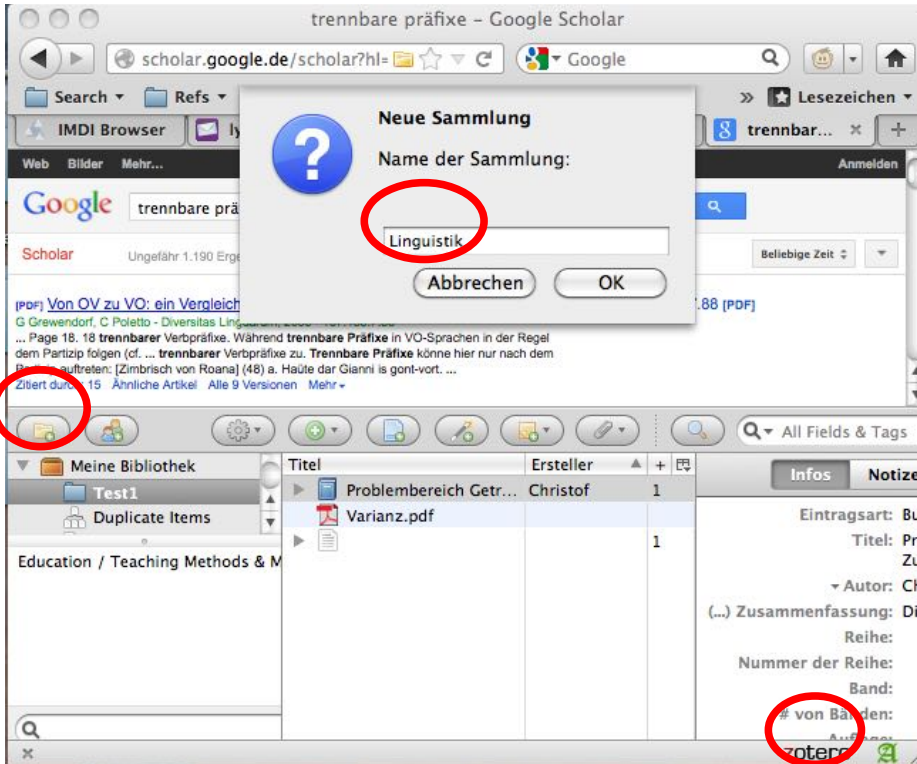
[\[ZITATION\] Zur Modifikation der Verben durch Präfigierung im Deutschen und Litauischen](#)
E Račienė - *Kalbotyra*, 1998
Zitiert durch: 2 [Ähnliche Artikel](#) [Alle 2 Versionen](#)

[\[HTML\] 'Wir bitten Sie das nicht misszugeneralisieren'. Sprachverhalten in grammatischen Zweifelsfällen am Beispiel trennbarer und nicht-trennbarer Verben](#) [HTML] von '...
T Becker, C Peschel - *Linguistik online*, 2003 - [linguistik-online.de](#)
... So geben Helbig/Buscha für die **trennbare** Variante von durch- die Bedeutung 'Resultat einer ... finden sich drei Typen von Infinitiven: 1. eindeutige Formen, die nur **trennbar** oder untrennbar ... Ziel dieser Aufgabe war zu überprüfen, ob mit den **trennbaren** wie untrennbaren **Präfixen** ...
Zitiert durch: 4 [Ähnliche Artikel](#) [Alle 14 Versionen](#) [Mehr](#) ▾

[\[BUCH\] Lexikalische Argumente und Adjunkte: zum semantischen Beitrag von verbalen Präfixen und Partikeln](#)
B Stiebels - 1996 - [books.google.com](#)

Sie wollen bei dieser Gelegenheit eine Literatursammlung „Linguistik“ beginnen.

Sie rufen auf der Schaltfläche rechts unten Zotero auf, klicken auf das Ikon für eine neue Sammlung und geben dieser den Namen „Linguistik“

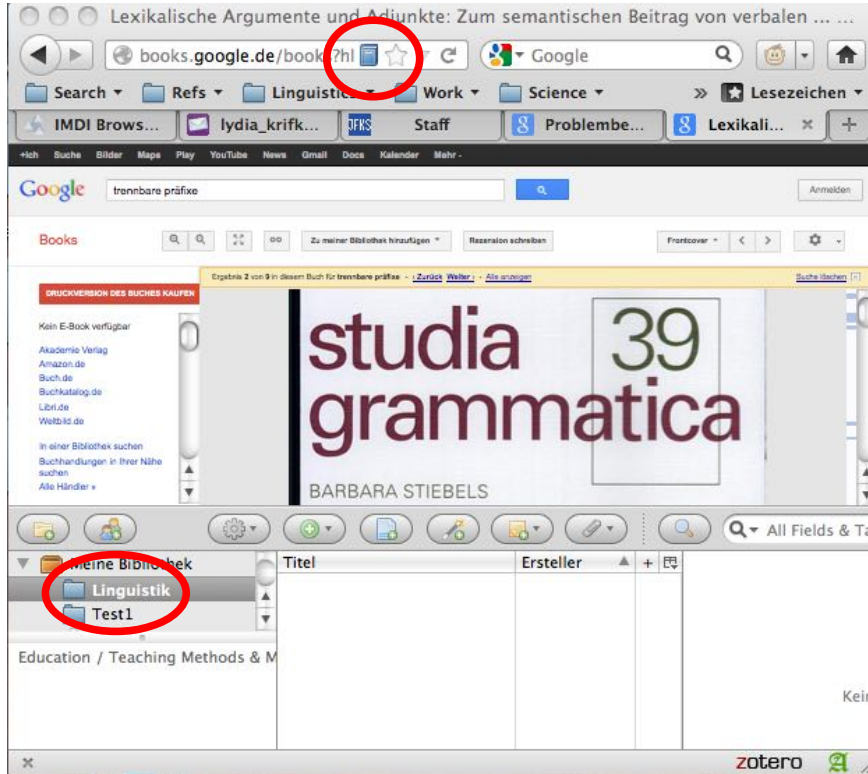


Angabe des Namens der neuen Sammlung

Ikon für Erstellen einer neuen Sammlung

Schaltfläche zum Aufruf von Zotero

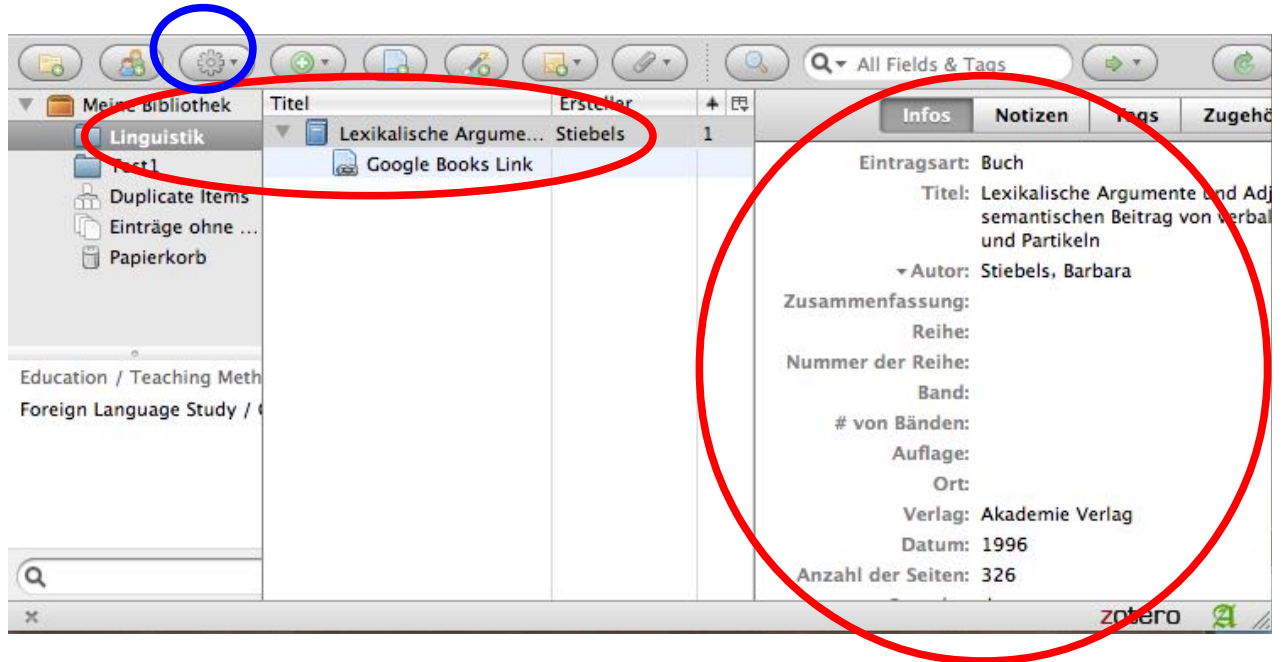
Sie finden das Buch von Barbara Stiebels interessant und klicken auf den Link in Google Scholar. In der URL-Zeile befindet sich ein Ikon, das anzeigt, dass die Referenz übernommen werden kann.



Ikon für Erstellen
der Referenz
in Zotero

Angewählt:
Sammlung „Linguistik“

Die Referenz wird übernommen:



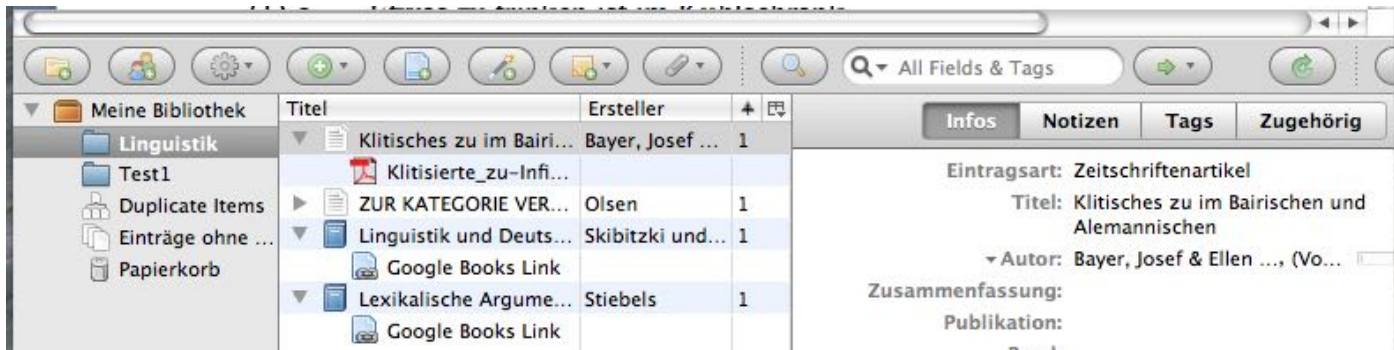
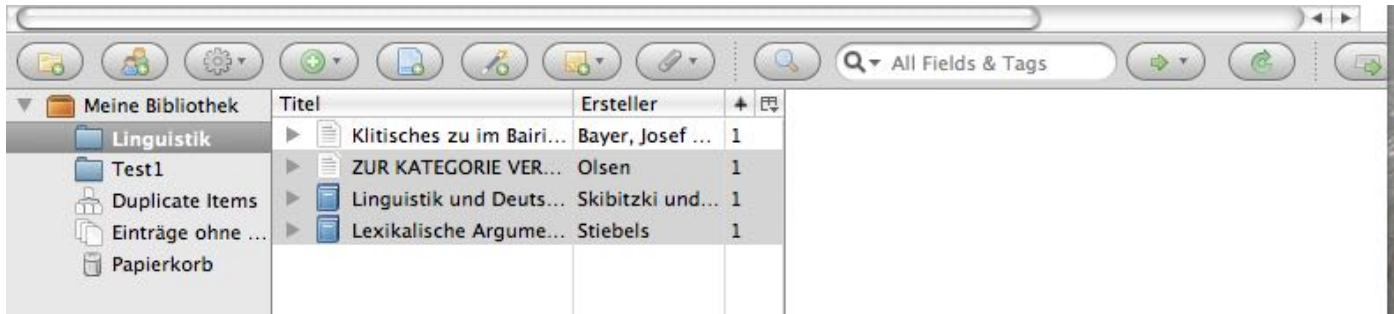
Der Link auf den Text (hier: auf Google Books) wird ebenfalls gespeichert.

Wenn die Referenz nicht automatisch übernommen wird, können Sie die Daten **per Hand** eintragen.

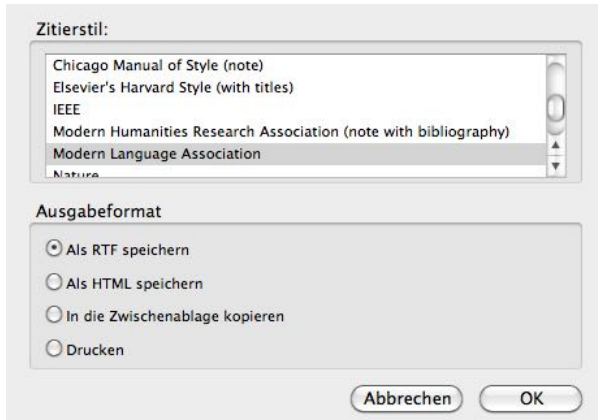
So entsteht eine kleine Literaturliste:

Übrigens: Dokumente wie z.B. pdf-Dateien kann man per drag & drop einbinden.

Auswahl aus der Datenbank durch Shift (Grossbuchstaben) + Klick:



Erstellen einer Literaturliste: Rechtsklick (beim Mac: Control + Click) auf die ausgewählten Einträge.
Auswahl eines Stiles, z.B. MLA, und Speichern als RTF-Datei:



Der gewählte Zitationsstil ist nicht ideal; unter <http://www.zotero.org/styles> kann man weitere Stile finden und implementieren (z.B. für die Zeitschrift *Language*.) Diese Stile kann man auch selbst weiter bearbeiten.

Olsen, Susan. „ZUR KATEGORIE VERBPARTIKEL“. *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 119.1 (1997): 1–32. Web. 28 Oct. 2012.

Skibitzki, Bernd, und Barbara Wotjak. *Linguistik und Deutsch als Fremdsprache: Festschrift für Gerhard Helbig zum 70. Geburtstag*. Walter de Gruyter, 1999. Print.

Stiebels, Barbara. *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Akademie Verlag, 1996. Print.

Textverarbeitung I

Überblick

Ziel der nächste Abschnitte:

Einführung in die technischen Aspekte des Schreibens von Texten, insbesondere Einsatz von Textverarbeitungsprogrammen.

Möglichkeiten:

- LaTeX:
Sehr gute Lösung für das Erstellen von wissenschaftlichen Arbeiten vor allem auch für Formeln, Tabellen etc., kann hier allerdings nur erwähnt werden.
- Microsoft Word:
Gute Lösung, industrieller Standard, allerdings kostenpflichtig.
- OpenOffice / LibreOffice:
Mindestens gleichwertig zu Microsoft Word, in mancher Hinsicht überlegen, Open Source Programm.

Elementares zu LaTeX

- Exzellente typografische Darstellung von Texten (Textsatzprogramm TeX)
- Hervorragend geeignet für Formeln, Tabellen, Baumdarstellungen (z.T. mittels Plug-Ins).
- Trennung zwischen Schreiben des Textes und dessen Darstellung (kein WYSIWIG: What you see is what you get).
- Man kann jeden Texteditor verwenden, um Texte zu produzieren, die von LaTeX gesetzt werden können.
- Es gibt jedoch spezielle Entwicklungsumgebungen, die es erlauben, auch den gesetzten Text am Bildschirm zu sehen.
- Weitere Hinweise, auch auf Programme und Plug-Ins:
<http://de.wikipedia.org/wiki/LaTeX>

Beispiel für Schreibumgebung und Text:

```

%% Erläuterungen zu den Befehlen erfolgen unter
%% diesem Beispiel.
\documentclass{scrartcl}
\usepackage[latin1]{inputenc}
\usepackage[T1]{fontenc}
\usepackage[ngerman]{babel}
\usepackage{amsmath}

\title{Ein Testdokument}
\author{Otto Normalverbraucher}
\date{05. Januar 2004}
\begin{document}

\maketitle
\tableofcontents
\section{Einleitung}

Hier kommt die Einleitung. Ihre Überschrift kommt
automatisch in das Inhaltsverzeichnis.

\subsection{Formeln}

\LaTeX{} ist auch ohne Formeln sehr nützlich und
einfach zu verwenden. Grafiken, Tabellen,
Querverweise aller Art, Literatur- und
Stichwortverzeichnis sind kein Problem.

Formeln sind etwas schwieriger, dennoch hier ein
einfaches Beispiel. Zwei von Einsteins
berühmtesten Formeln lauten:
\begin{align}
E &= mc^2 && \\\
m &= \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} && \\
\end{align}
Aber wer keine Formeln schreibt, braucht sich
damit auch nicht zu beschäftigen.
\end{document}

```

Ein Testdokument

Otto Normalverbraucher

05. Januar 2004

Inhaltsverzeichnis

1 Einleitung	1
1.1 Formeln	1

1 Einleitung

Hier kommt die Einleitung. Ihre Überschrift kommt automatisch in das Inhaltsverzeichnis.

1.1 Formeln

\LaTeX ist auch ohne Formeln sehr nützlich und einfach zu verwenden. Grafiken, Tabellen, Querverweise aller Art, Literatur- und Stichwortverzeichnis sind kein Problem.

Formeln sind etwas schwieriger, dennoch hier ein einfaches Beispiel. Zwei von Einsteins berühmtesten Formeln lauten:

$$E = mc^2 \quad (1)$$

$$m = \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} \quad (2)$$

Aber wer keine Formeln schreibt, braucht sich damit auch nicht zu beschäftigen.

Ausgabe des links stehenden Codes



Einführung in LibreOffice (OpenOffice)

- Freies Software-Paket für verschiedene Betriebssysteme (Windows, Unix, Mac-OS); früher OpenOffice, jetzt aus Lizenzgründen weiterentwickelt als LibreOffice
- Funktionen in einem Programm:
 - Writer (Textverarbeitung)
 - Calc (Tabellenkalkulation)
 - Base (Datenbank)
 - Impress (Präsentation)
 - Draw (Zeichen)
 - Math (Formel-Editor)
- Besitzt viele Funktionen von MS Word, übertrifft es in einigen Punkten, gute Benutzerführung.
- Enthält eine Literaturdatenbank.
- Kann auch komplexe MS Word-Dateien lesen und doc-Dateien produzieren. Allerdings kann es bei Bildern und komplexeren Formaten Schwierigkeiten geben.
- Download: <http://de.libreoffice.org>
- Handbücher: <https://de.libreoffice.org/get-help/documentation/>; empfehlenswert: Einführung in LibreOffice; Einführung in Writer;
- LibreOffice fürs Studium:
David Paenson, Techniken der Textverarbeitung anhand OpenOffice.org
http://de.openoffice.org/doc/howto_2_0/writer/ooo_fuer_studenten.pdf
- Barbara Slawig & Michael Voges, Tutorium für Aufzählungen und Nummerieren:
http://de.openoffice.org/doc/howto_2_0/writer/tutorium-aufzaehlungen-nummerierungen.pdf

Auf der Moodle-Seite wird eine Formatdatei bereitgestellt.

Grundeinstellungen

Die Grundeinstellungen von LibreOffice sind gut.

Vorschlag: Gleiche Schrift für Überschriften und Text verwenden, nämlich

a. Times New Roman (eine Serifenschrift)

T

b. Arial (eine serifenlose Schrift).

T

Über LIBREOFFICE ⇒ EINSTELLUNGEN ⇒ LIBREOFFICE WRITER ⇒ GRUNDSCHRIFTARTEN

Machen Sie sich sonst mit den Einstellungsmöglichkeiten vertraut.

Dokumentvorlagen

Damit ein Text möglichst einheitlich aussieht: Dokumentvorlagen (engl. Templates)

Die Vorgaben für Überschriften, Fließtexte, Beschriftungen, Fußzeilen usw. werden in einer Datei, der Dokumentvorlage, gespeichert. Für verschiedene Textarten kann man verschiedene Dokumentvorlagen erstellen.

Für LibreOffice kann man Vorlagen aus dem Internet herunterladen:

➤ <https://www.openofficetemplates.com/> (deutsche Vorlagen)

Wir wollen hier eine neue Dokumentvorlage anlegen.

Hierzu öffnen wir ein neues Dokument, definieren einige wichtige Eigenschaften, und speichern es als Dokumentvorlage ab.

- Neues Dokument öffnen: DATEI ⇒ NEU ⇒ TEXTDOKUMENT.
- FORMAT ⇒ FORMATVORLAGEN ⇒ SEITENVORLAGEN (4. Schaltfläche) ⇒ STANDARD (rechtsklicken oder Ctrl+Klicken) ⇒ ÄNDERN
- Sie können jetzt das Format der Seite bestimmen. Vorschlag: Din A4, Hochformat, Rand 2 cm. FUSSZEILE einschalten.
- Sie können auch die erste Seite, oder gerade und ungerade Seiten verschieden formatieren.
- Sichern Sie das Dokument als Dokumentvorlage: DATEI ⇒ SPEICHERN UNTER; speichern Sie die Dokumentvorlagen in einem eigenen Ordner.
- In Fußzeile: Seitenzahl (EINFÜGEN ⇒ FELDBEFEHL ⇒ SEITENNUMMER, zentrieren.)

Absatzvorlagen

Wir definieren nun Dokumentvorlagen für die verschiedenen Absätze.
Dies erlaubt uns eine einheitliche Darstellung von Überschriften, Fußnoten, Zitaten etc.

Grundsatz: Formate nie lokal, sondern immer nur über Vorlagen ändern!
Andernfalls droht Uneinheitlichkeit und Mehrarbeit bei Veränderungen.

Beispiel:

- FORMAT ⇒ FORMATVORLAGEN ⇒ ABSATZVORLAGEN (1. Schaltfläche) ⇒ STANDARD
- Rechtsklicken / Ctrl-Klicken, ÄNDERN
- EINZÜGE UND ABSTÄNDE: Einzeilig, Linksbündig
- Textfluss: Schusterjungenregelung, Hurenkinderregelung 2 Zeilen
- Schrift: Times New Roman
- Tabulatoren: 1 cm, 1,5 cm, 2 cm etc., d.h. alle 0,5 cm, linksbündig

Mit der Absatzvorlage STANDARD sind die anderen Absatzvorlagen verknüpft.

Bemerkung zu Absätzen:

Absätze sind Textabschnitte, die durch eine EINGABE/RETURN getrennt sind.
Sie können einen festen Zeilenumbruch einfügen, ohne einen neuen Absatz zu erzeugen,
indem Sie CTRL+RETURN drücken.

Weitere Absatzvorlagen

- Absatzvorlagen TEXTKÖRPER, TEXTKÖRPER EINRÜCKUNG, TEXTKÖRPER EINZUG usw. sind die auf STANDARD definierten Abschnitte für den normalen Text. Zwischen den Absätzen wird hier automatisch ein Zwischenraum erzeugt.
- Formatieren der Überschriften ÜBERSCHRIFT1, ÜBERSCHRIFT2 etc.
Vorschlag:
Bei Überschriften allgemein: Absatz nicht trennen, Absätze zusammenhalten
ÜBERSCHRIFT1: Fett, Schriftgröße 16pt
ÜBERSCHRIFT2: Fett, Kursiv, Schriftgrad 14pt
ÜBERSCHRIFT3: Fett, Schriftgrad 12pt
- Dezimal nummerierte Überschriften:
FORMAT ⇒ NUMMERIERUNG UND AUFZÄHLUNGSZEICHEN ⇒ GLIEDERUNG ⇒ 1., 1.1. etc.
- Einführen einer neuen Absatzvorlage BEISPIEL, mit Nummerierung in Klammern, für linguistische Beispiele.
Hierzu müssen Sie eine Nummerierung wählen (z.B. Nummerierung 1) und diese entsprechend formatieren:
FORMAT ⇒ NUMMERIERUNG UND AUFZÄHLUNGSZEICHEN ⇒ NUMMERIERUNGSART
- Einführung weiterer Absatzvorlagen, z.B. für Listen;
das Aufzählungszeichen kann ebenfalls wie oben gewählt werden.

Querverweise

In wissenschaftlichen Texten werden verschiedene Arten von Textteilen nummeriert, um darauf zu verweisen.

- Seiten (dies ist allerdings eine äußere Einheit)
- Zeilen (z.B. in Quellentexten, meist ebenfalls eine äußere Einheit)
- Überschriften
- Abbildungen, Tabellen, Formeln
- Fußnoten, Endnoten
- Bibliographische Hinweise (nach bestimmten Standards)
- In der Sprachwissenschaft: Sprachbeispiele

Automatische Nummerierung und automatische Querverweise auf nummerierte Einheiten erlauben eine größere Flexibilität beim Schreiben.

Insbesondere müssen die Nummern bei nachträglichen Änderungen nicht per Hand verändert werden, was zeitraubend und fehlerträchtig ist.

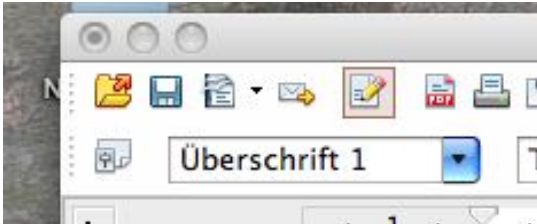
In LibreOffice können auf Abschnitte mit Nummerierung Querverweise setzen:

- EINFÜGEN ⇒ QUERVERWEIS
- Aktualisieren des Querverweises durch CTRL+F9

Ansprechen der Formatvorlagen

Die Formatvorlagen kann man auf verschiedene Weise auswählen:

- Aus dem Fenster links oben:



- Besser und schneller: Über die Tastatur.

LIBREOFFICE hat dabei verschiedene Tastatur-Kürzel vorprogrammiert.

Siehe EXTRAS ⇒ ANPASSEN ⇒ TASTATUR.

Sie können neu definierte Formatvorlagen auf die gleiche Weise zugänglich machen, z.B. die eben definierte Formatvorlage für "Beispiel".

Wichtig: Sehen Sie sich die Tastatur-Abkürzungen an.

Legen Sie sich eine Liste für die häufiger verwendeten Befehle an, das spart Zeit.

Die Liste kann auch als Testdatei abrufbar sein.

Weitere Vorlagen

Bei LIBREOFFICE gibt es ferner die Möglichkeit, für andere Einheiten als Absätze Formatvorlagen zu erstellen. Insbesondere gilt dies:

- Für Text innerhalb von Absätzen, z.B. für Zitate und Hervorhebungen, gibt es Zeichen-Vorlagen. Sie können dann z.B. alle Zitate von unterstrichen auf *kursiv* umstellen.
- Sie können für bestimmte Seiten innerhalb eines Dokuments verschiedene Formatvorlagen definieren.
- Sie können für Textabschnitte verschiedene Formate definieren, z.B. Wechsel ein/zweispaltig.

Weitere Tipps

- Linksbündige Ausrichtung (Flattersatz) oft besser lesbar als Blocksatz.
- Zeilenabstand: genauen Wert einstellen, nicht "einfach", sonst gibt's Probleme mit Sonderzeichen.
- Abstand vor und nach jedem Absatz statt einer Zwischenzeile zwischen Absätzen. Zusätzlich oder alternativ dazu kann man die erste Zeile einziehen.

Automatisches Korrektur

Wie Sie wissen, erlauben Textverarbeitungsprogramme eine automatische Rechtschreib-Korrektur.

Sie müssen dazu die Sprache einstellen:

- Global für Dokumente: LIBREOFFICE ⇒ EINSTELLUNGEN ⇒ SPRACHEINSTELLUNGEN
- Für Dokumente: EXTRAS ⇒ SPRACHE
- Für Formatvorlagen kann man separat die Sprache einstellen und auch verschiedene Sprachen im selben Dokument prüfen lassen.

Überprüfung der Rechtschreibung:

- Nach dem Schreiben
- während des Schreibens; Markierung durch rote wellige Unterstreichung.

Aktivierung unter EXTRAS ⇒ RECHTSCHREIBUNG UND GRAMMATIK ⇒ OPTIONEN.

Wichtig: Richtige, aber als falsch markierte Ausdrücke sollten im Benutzerwörterbuch aufgenommen werden, z.b. Rechtsklick auf rot markierten Wort und Auswahl von HINZUFÜGEN.

Literaturverwaltung

LibreOffice besitzt eine Literaturverwaltung (EXTRAS ⇒ LITERATURDATENBANK).

- Gegenüber anderen Literaturverwaltungen etwas eingeschränkt (keine Internet-Integration, keine Dokumentenspeicherung).
- Reicht für die Zwecke des Bachelor-Studiums aber völlig aus.

The screenshot shows the 'Literaturdatenbank' application window. At the top, there is a search bar with 'Tab biblio' and 'Suc' followed by search icons. Below is a table with columns: Author, Booktitle, Journal, Note, Number, Pages, Publisher, Title, Volume, and Year. The table contains several entries, with the entry for 'Miller, Robin' selected. Below the table, there is a detailed view of the selected entry, showing fields for Autor(en), Verlag, Kapitel, Titel, Adresse, and Seite(n).

	Author	Booktitle	Journal	Note	Number	Pages	Publisher	Title	Volume	Year
	Sun Microsystems Inc					226	luniverse Inc.	StarOffice		2005
	Bain, Mark Alexander					220	Packt Publishing	Learn Op		2006
▶	Miller, Robin					216	Prentice Hall International	Point & Click		2005
	Haugland, Solveig					520	Solveig Haugland	The OpenOffice		2006
	Leete, Gurdy; Finkelstein, Stephen					384	Wiley & Sons	OpenOffice		2003
	Lamprecht, Stephan					102	Remote Verlag	OpenOffice		2005

Datensatz 13 | von 59

Autor(en) Miller, Robin Titel Point & Click OpenOffice.org

Verlag Prentice Hall International Adresse

Kapitel Seite(n) 216

MS Word

Grundsätzliches zu MS Word

- Weit verbreitetes Programm, kommerzielle Software, für Windows und Mac-OS.
- Teil der Suite MS-Office (mit Tabellenkalkulation und Präsentationssoftware)
- Seit ca. 2004 auch sinnvoll für die Textverarbeitung in wissenschaftlichen Texten einsetzbar.
- Die vielen verfügbaren Werkzeuge werden oft nicht eingesetzt; hier konzentrieren wir uns gerade darauf.
- Weitere Hinweise zur Verwendung von MS Word in der Linguistik siehe <http://www.linguistics.ucsb.edu/faculty/cumming/WordForLinguists/Index.htm>

Schreibkonventionen

In diesem Teil geht es um die schriftliche Darstellung wissenschaftlicher Ergebnisse, insbesondere aus der Sprachwissenschaft.

Wir werden ein konkretes Beispiel später ausführlich diskutieren, wenn wir uns genauer die sprachwissenschaftliche Argumentation ansehen.

Hausarbeiten und Bachelorarbeit folgen im wesentlichen diesen Schreibkonventionen.

Aufbau einer wissenschaftlichen Arbeit

In [Klammern]: Optionale Teile

- Titel: Treffende Bezeichnung, kann durch Untertitel spezifiziert werden.
- Abstract: Kurze Zusammenfassung
- Inhaltsverzeichnis, Tabellen- und Abbildungsverz. bei längeren Werken
- Einleitung: Hinführung zum Thema der Arbeit, Begründung aus dem Forschungsstand, Überblick über den Artikel.
- Hauptteil: In Unterteile gegliedert, üblich ist die Dezimalnotation der Überschriften.
- Zusammenfassung: Erinnerung an die Forschungsfrage, wie wurde diese behandelt, welche Ergebnisse hat dies erbracht, was sagen diese aus über die Forschungsfrage, welche neuen Fragen stellen sich, welche neuen Ansätze ergeben sich?
- Literaturverzeichnis: Zitierte Literatur.
- Anhänge: Z.B. untersuchte Texte, verwendete Fragebögen usw.

Schreibkonventionen: Aufbau einer wissenschaftliche

Titel

Autoname

[Arbeitseinrichtung]

[Abstract]

[Keywords]

[Inhaltsverzeichnis]

[Verzeichnis von Tabellen, Abbildungen]

1. Einleitung

Einleitungstext, typischerweise etwa 10% des Gesamttextes.

2. Erste Überschrift des Hauptteils

[evtl. knapper Überblick]

2.1. Erste Unterüberschrift

Text des ersten Unterkapitels.

2.2. Zweite Unterüberschrift

Text des zweiten Unterkapitels
etc.

3. Zweite Überschrift des Hauptteils

(siehe oben)

[evtl. weitere Überschriften]

4. Zusammenfassung / Schlussteil

Zusammenfassung der Arbeit, Resultate, typischerweise 10%.

Literaturverzeichnis (Bibliographien)

[Anhänge]

Weiteres zum Aufbau wissenschaftlicher Arbeiten

- Kapitel- und Unterkapitelteilung:
Nicht zu feinteilig, nicht zu grob.
- Keine Kapitel mit nur einem einzigen Unterkapitel!
- Zu vermeiden:
Überschrift gefolgt von viel Text,
gefolgt von der ersten Unterüberschrift.

Zum Thema Fußnoten

- Fußnoten für zusätzliche Informationen,
die für das Verständnis des Textes entbehrlich sind
und ablenken würden.
- Fußnoten sind leserfreundlicher als Endnoten!
- Fußnoten möglichst sparsam einsetzen.
- Nach der Konventionen in naturwissenschaftlichen und auch linguistischen Texten:
Fußnoten nicht für Literaturangaben verwenden!
Literaturangaben stehen vielmehr in einer Bibliographie am Ende.
- Verwenden Sie die Textverarbeitungsfunktion: EINFÜGEN ⇒ FUSSNOTE
- (Und nicht vergessen: Keine Fußnote zu einer Fußnote!)

2. Überschrift

Text text text text text text text text text text text
text text text text text text text text text text text
text text text text text text text text text text text
text text text.

2.1 Unterüberschrift

Text text text text text text text text text text text
text text text text text text text text text text text
text text.

2.2 Unterüberschrift

Text text text text text text text text text text text
text text text text text text text text text text text
text text.

Linguistische Konventionen

In linguistischen Texten haben sich bestimmte Konventionen herausgebildet, die Sie kennen müssen und die Sie auch einhalten sollen.

- In linguistischen Texten wird häufig über sprachliche Ausdrücke geschrieben. Man unterscheidet hier zwischen **Metasprache** (sprachliche Ausdrücke, die beschreiben) und **Objektsprache** (sprachliche Ausdrücke, die beschrieben werden). Die Referenz auf Objektsprache im Text erfolgt mit *Kursivschrift*, ohne Anführungszeichen. Kursivschrift möglichst für keinen anderen Zweck verwenden (außer im Literaturverzeichnis für selbständige Publikationen, siehe unten).
- Bedeutungsbeschreibungen werden in 'einfache Anführungszeichen' (oben) gesetzt.
- Phonetische Umschrift wird in [eckigen Klammern] wiedergegeben.
- Zur Kennzeichnung phonemischer Darstellungen werden /Schrägstriche/ verwendet.
- Grapheme werden durch <spitze Klammern> markiert.

Ein Beispiel für diese Konventionen:

Das englische Wort *knee* [ni:] 'Knie' zeigt den Verlust eines initialen /k/, das jedoch in der Schreibung noch als <k> vorhanden ist.

- Das **Sternchen** * steht für Ungrammatikalität, das **Fragezeichen** ? oder ?? für schwache Ungrammatikalität, das **Prozentzeichen** % für unterschiedliche Bewertungen in der Sprachgemeinschaft, das **Doppelkreuz** # für pragmatische Unangemessenheit.

- Längere objektsprachliche Beispiele werden in eigenen Abschnitten nummeriert dargestellt. Sie können, müssen dann aber nicht kursiv dargestellt werden.

Das Vorfeld im Deutschen kann manchmal durch ein trennbares Präfix gefüllt werden, wie der Beginn des Gedichts von Conrad Ferdinand Meyer zu dem Brunnen am Petersplatz in Rom zeigt:

(27) Auf steigt der Strahl, und fallend gießt
Er voll der Marmorschale Rund (...)

Wir werden die Behandlung fremdsprachlicher Beispiele (Interlinearübersetzung, idiomatische Übersetzung) später ausführlich behandeln.

- Auf Beispiele und andere nummerierte Elemente kann man durch die Nummern Bezug nehmen:

Beispiele für die Vorfeldfüllung durch ein trennbares Präfix, wie es (27) mit dem Präfix *auch* zeigt (das Verb lautet *aufsteigen*), sind nicht auf poetische Sprache beschränkt. (...)

- Zitate werden in “doppelten Anführungszeichen” wiedergegeben (man kann natürlich auch die „deutschen Anführungszeichen“ verwenden). Längere Zitate werden in einem eigenen Abschnitt, mit Einrückung und oft in etwas kleinerer Schrift, dargestellt.
Zur Kunst des Zitierens kommt später mehr!

Der Literaturverweis

Wie wir gesehen haben, nehmen wissenschaftliche Arbeiten wesentlich auf andere Arbeiten Bezug. Dies geschieht in zwei Schritten:

- Im Text selbst durch Literaturverweise.
- Außerhalb des Texts im engeren Sinn durch das Literaturverzeichnis.

Es gibt verschiedene Traditionen, wie man im Text selbst eine Literaturangabe macht.

Die Fußnote

In der Literaturwissenschaft und in den Geisteswissenschaften üblich.

Zugehörige Fußnoten:

- 1 Zur Terminologie siehe F.K. Stanzel, *Theorie des Erzählens* (Göttingen: UTB 904, 1985), bes. Kap. 6 und *Typische Formen des Romans* (Göttingen, 1981), pp. 39 ff.
- 2 Vgl. Stanzel, *Typische Formen*, *op. cit.*, "Der personale Roman", pp. 39–52.

Abkürzungen:

- a.a.O. "am angeführten Ort" ⇒ Verweis auf frühere Literaturangabe
- ibid. "ibidem" ⇒ Verweis auf unmittelbar vorangegangene Literaturangabe.

Problem: Unübersichtlich, wenn viel Literatur zitiert wird. Man muss oft suchen, um herauszubekommen, was unter a.a.O. oder ibid. steckt.

Autor mit Jahreszahl

In den Naturwissenschaften und in der Linguistik ist die folgende Zitierweise üblich:

einer genaueren Feinabstimmung der Artikulationsbewegungen werden die geometrischen Eigenschaften des Ansatzrohres beeinflusst (s. Kapitel 3). Sowohl Sluiter (1995) als auch Classen et al. (1998) beobachteten, dass ein erhöhter subglottaler Luftdruck bereits Einfluss auf die Quelle, also die Larynx als Ort der Phonation, hinsichtlich der Schwingungseigenschaften der Stimmlippen nimmt (z.B. erhöhte Steilheit des Stimmlippenverschlusses, Öffnungsgrad der Stimmlippen). Zusätzlich dazu bewirkt eine punktuell feinere Abstimmung der Artikulationsbewegungen für die betonte Silbe eine Form von Hyperartikulation (Lindblom 1990). Akustisch gesehen resultiert daraus eine Erhöhung der Amplituden in den höheren Frequenzbereichen des Sprachsignals bei der Realisierung der Wortbetonung (eine detaillierte Untersuchung des so genannten *Spectral tilt* für die Wortbetonung im Deutschen findet sich in Classen et al. 1998).

Bemerke:

- Im Text bezieht man sich auf ein Werk wie folgt:
Sluiter (1995) beobachtet, dass... / Classen et al. (1998) beobachten, dass...
- Zusätze in Klammern: *Dies führt zu Huperartikulation (Lindblom 1990).*
- Bei Mehrfachpublikationen Unterscheidung durch Buchstaben: *Sluiter (1995b)*
- Seitenangaben zu Stellen im Text möglich: *Sluiter (1995: 95-98) beobachtet...*

Das Literaturverzeichnis

Das Literaturverzeichnis muss es erlauben, die Arbeiten, auf die man Bezugnimmt, zu identifizieren. Hierzu benötigt man leicht unterschiedliche Angaben, je nach dem Typ der Publikation.

Grundangaben:

- Verfasser
- Jahr der Veröffentlichung
- Titel der Veröffentlichung
- Erscheinungsort der Veröffentlichung (Zeitschrift, Verlag, Verlagsort).
- Bei unselbständigen Veröffentlichungen: Seitenangabe, evtl. Bandnummer.

Für die Darstellung gibt es unterschiedliche Formate, die oft nur leicht voneinander abweichen.

Für die Sprachwissenschaft: “Unified Style Sheet”, von Herausgebern linguistischer Zeitschriften erarbeitet:

Hier einige Beispiele für Zitationsweisen aus diesem Stylesheet:

<http://www.lsadc.org/info/documents/style-sheet.pdf>

Blevins, Juliette. 2004. *Evolutionary phonology*.
Cambridge:
Cambridge University Press.

Casali, Roderic F. 1998. Predicting ATR activity.
Chicago Linguistic Society (CLS) 34(1). 55-68.

Franks, Steven. 2005. Bulgarian clitics are positioned in
the syntax.
[http://www.cogs.indiana.edu/people/homepages/
franks/Bg_clitics_remark_dense.pdf](http://www.cogs.indiana.edu/people/homepages/franks/Bg_clitics_remark_dense.pdf) (17 May, 2006.)

Johnson, Kyle, Mark Baker & Ian Roberts. 1989. Passive
arguments raised. *Linguistic Inquiry* 20. 219-251.

Lahiri, Aditi (ed.). 2000. *Analogy, leveling, markedness:
Principles of change in phonology and morphology*
(Trends in Linguistics 127). Berlin: Mouton de Gruyter.

McCarthy, John J. & Alan S. Prince. 1999. Prosodic
morphology. In John A. Goldsmith (ed.), *Phonological
theory: The essential readings*, 238-288. Malden, MA &
Oxford: Blackwell.

Oxford English Dictionary, 2nd edn. 1989. Oxford:
Oxford University Press.

Yu, Alan C. L. 2003. *The morphology and phonology of
infixation*. Berkeley, CA: University of California
dissertation.

Selbständige Veröffentlichung:
Titel kursiv, Englisch: Kleinschreibung.

Konferenzbeitrag:
Titel recte, Konferenz kursiv, Kürzel.

Internet-Veröffentlichung, mit URL

Zeitschriftenartikel:
Titel recte, Zeitschrift kursiv,
Jahrgangsnummer.

Herausgegebenes Werk, mit Reihentitel.

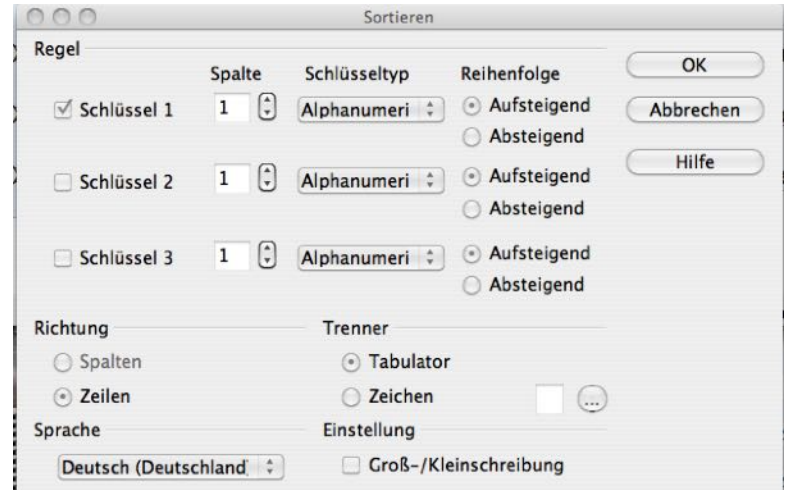
Veröffentlichung in einem Sammelband.
Hier auch: Mehrfache Autorenschaft.

Werk ohne Herausgeber oder Autor.

Dissertation

Bemerke:

- Die Bibliographie ist sortiert. Hierzu kann man die automatische Sortierfunktion verwenden
LibreOffice: Text markieren, dann EXTRAS ⇒ SORTIEREN
MS Word: Text markieren, dann TABELLE ⇒ SORTIEREN ⇒ Absätze
- Sortierreihenfolge alphabetisch nach Nachnamen.
Konventionen: Deutsch *von* wird zum Nachnamen gerechnet (sortiert unter *von*), niederländisch *van* nicht (sortiert unter dem folgenden Namen).
- Umlaute ä, ö, ü und ß werden nach dem deutschen System wie ae, oe, ue und ss behandelt, international folgen sie aber a, o, u und s.
- Der Sortieralgorithmus von Textverarbeitungsprogrammen erlaubt die sprachspezifische Einstellung, z.B. in LibreOffice:: EXTRAS ⇒ SORTIEREN



Textverarbeitung II

Sonderzeichen: Grundsätzliches

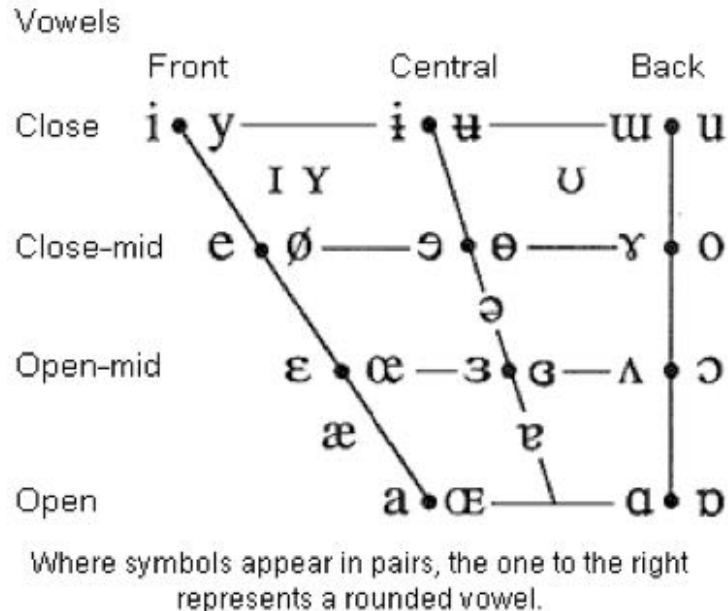
In der Sprachwissenschaft verwendet man verschiedene Typen von Sonderzeichen:

- phonetische Zeichen
- logisch-mathematische Zeichen
- griechische Buchstaben, lateinische Zeichen mit diversen Akzenten
- die Schriften anderer Sprachen.

Phonetische Zeichen: International Phonetic Alphabet (IPA)

- http://de.wikipedia.org/wiki/Internationales_Phonetisches_Alphabet mit detaillierten Informationen und der Möglichkeit, Zeichen herauszukopieren
- <https://www.internationalphoneticassociation.org/> für die Beschreibung des Phonetischen Alphabets durch die International Phonetic Association.
- Eine IPA App von John Esling: <https://itunes.apple.com/de/app/ipa-phonetics/id869642260?mt=8> .

Beispiel: IPA Vokalzeichen.



IPA Konsonantenzeichen

Consonants (pulmonic)

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or flap		ɸ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Logisch-mathematische Zeichen

Die wichtigsten in der Linguistik gebräuchlichen Zeichen:

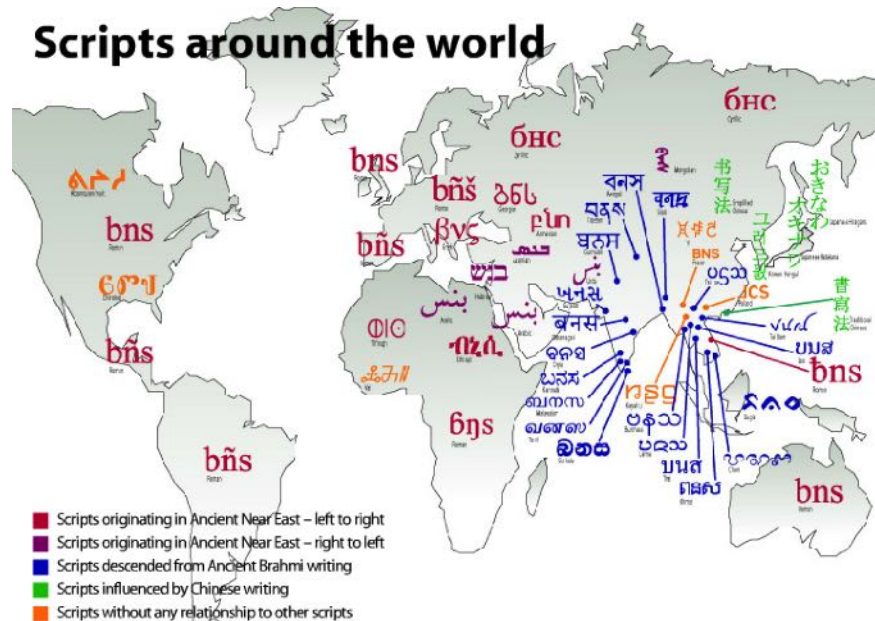
Konjunktion	\wedge	Element	\in
Disjunktion	\vee	Teilmenge	\subseteq
Konditional	\rightarrow	echte Teilmenge	\subset
Bikonditional	\leftrightarrow	Ungleichheit	\neq
Allquantor	\forall		
Existenzquantor	\exists	Summenindividuum	\oplus, \sqcup
iota	ι		
lambda	λ	Logische Folgerung (turnstile)	\vdash, \vDash

Fremdsprachliche Schriften

Heute können zahlreiche fremdsprachliche Schriften geschrieben werden:

- die ideographischen Zeichen des Chinesischen,
- ägyptische Hieroglyphen,
- die Keilschrift,
- die Maya-Schrift usw.
- Natürlich können auch rechtsläufige Schriften (Arabisch, Hebräisch) dargestellt werden.

Weitere Informationen unter der Webseite der *Non-Roman Script Initiative*, http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&cat_id=Home



Sonderzeichen: Kodierung und Eingabe

Unicode

Unicode ist ein international vereinbarter Standard, umfassend genug für alle jemals verwendeten Zeichensysteme (17 Codebereiche, jeweils 65536 Zeichen, d.h. > 1 Mio).

Bis heute wurden erst die ersten drei Codebereiche (planes) belegt, und auch das nur teilweise.
















- Codebereich 0: Gegenwärtige Schriften, Symbole usw.
- Codebereich 1: Historische Schriften (z.B. Hieroglyphen)
- Codebereich 2: Seltene Ideogramme

Für weiterführende Informationen siehe <http://de.wikipedia.org/wiki/Unicode>.

So sieht die Belegung des Codebereichs 0 aus.
Die Zeichen werden mit 16 bit kodiert.



00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

	Lateinische Schriften und Symbole
	Lautschriften
	Andere europäische Schriften
	Nahost- und Südwestasiatische Schriften
	Afrikanische Schriften
	Südasiatische Schriften
	Südostasiatische Schriften
	Ostasiatische Schriften
	CJK-Ideogramme
	Kanadische Silben
	Symbole
	Diakritika
	UTF-16-Surrogates und privater Nutzungsbereich
	Verschiedene Zeichen
	Nicht belegte Codebereiche

Unicode-Zeichensätze

Unicode ist ein Kodierungs-Standard, kein Zeichensatz (Font)!

Es gibt keinen Zeichensatz, der alle Unicode-Zeichen enthält.

Wichtige Zeichensätze mit zahlreichen Zeichen:

- Arial Unicode
- Lucida Grande (Macintosh)
- Symbol Σψμβολ
(eigentlich kein Unicode-Zeichensatz, mit griechischen Buchstaben, mathematischen Symbolen)
- Doulos SIL: Ein spezialisierter Zeichensatz, u.a. mit allen IPA-Symbolen

Weitere Informationen: http://en.wikipedia.org/wiki/Unicode_fonts

Doulos SIL kann heruntergeladen und installiert werden:

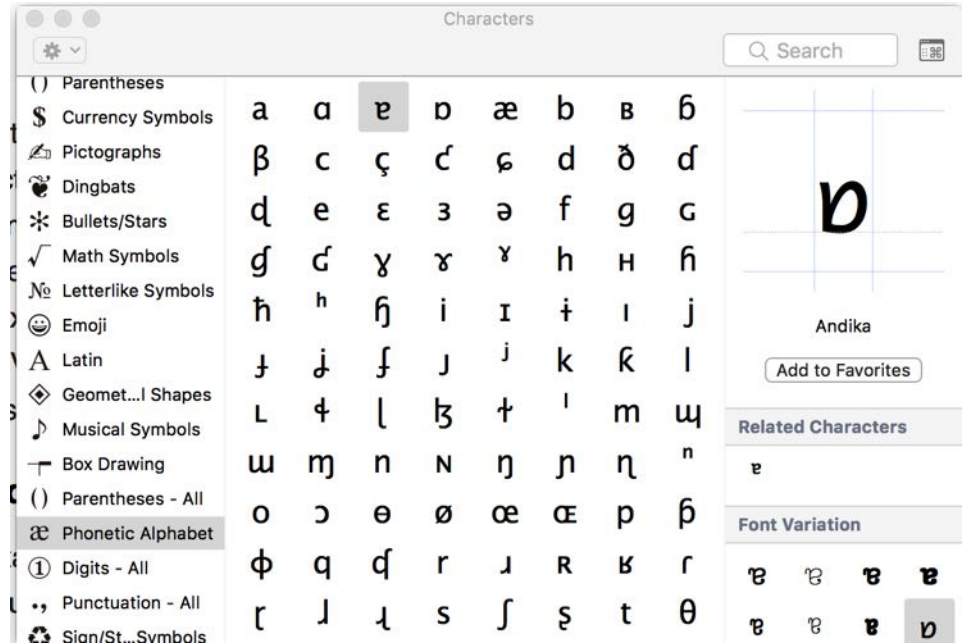
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=DoulosSILfont

Eingabe von Unicode-Sonderzeichen

Der Unicode-Standard UTF8 wird von Textverarbeitungsprogrammen, Webbrowsern, E-mail-Programmen usw. unterstützt.

Es gibt verschiedene Eingabemöglichkeiten, unter anderem:

- Über das Betriebssystem:
 - Character Palette (Mac), siehe links.
 - Character Map (Windows)
 - PopChar (kostenpflichtig)
- Über die Auswahl in Menüs von Textverarbeitungsprogrammen
Beispiel: LibreOffice
EINFÜGEN ⇒ SONDERZEICHEN, Schrift/Bereich auswählen.



Eingabe von Sonderzeichen über die Tastatur

Die Eingabe von Sonderzeichen über Menüs ist oft langwierig.

Besser: Eingabe über die Tastatur!

Dies kann man innerhalb des Textprogramms machen:

- In **MSWord** kann man beim EINGABE ⇒ SONDERZEICHEN Tastenkürzel definieren;
in **LibreOffice** kann man hierfür Makros verwenden.
- In **MSWord** und in **LibreOffice** kann man die Funktion AUTOKORREKTUR dazu gebrauchen, die eigentlich dazu da ist, häufig falsch geschriebene Wörter zu ersetzen.
Sie definieren ein Umschaltzeichen, z.B. “/”, gefolgt von Zeichenbeschreibungen:
z.B. /alpha ⇒ α, /all ⇒ ▼, usw.

Empfehlenswert ist es allerdings die Definition durch ein externes Hilfsprogramm, dann stehen die Sonderzeichen für alle Programme zur Verfügung.

- Für Windows: Keyboard Layout Manager,
<http://www.klm32.com/>
- Für das Macintosh-Betriebssystem: Ukelele,
<http://scripts.sil.org/ukelele>

Damit kann man die Tastatur neu belegen, wobei auch Umschalt-Tasten möglich sind.

Daneben gibt es Programme, die Sequenzen von Tastatureingaben speichern und wiedergeben:

- z.B. Keyboard Maestro für Macintosh.

Beispiel: Meine Kodierung eines Teils der mathematischen Sonderzeichen, unter Option-F abrufbar:



Akzente und diakritische Zeichenpolarity

- Gängige Akzente können durch Tastenkombinationen eingegeben werden (mit alt/option für Macintosh OSX, andere Kombinationen mit Windows, siehe Symbol-Menü; wenn diese Tasten nicht anders belegt sind.)

Gravis à	opt+`	Trema ä	opt+u
Akut á	opt+e	Makron ā	opt+a
Zirkumflex â	opt+^	Brevis ă	opt+b
Bogen (Arch) â	opt+shift+a	Tilde ã	opt+n
Hachek ǎ	opt+v	Ogonek ą	opt+m
Ring å	opt+k	Doppel-Akut ũ	opt+j
Punkt unten ȁ	opt+x	Punkt oben é	opt+w

So können auch weitere sprachspezifische Sonderzeichen eingegeben werden:

Strich ł,	opt+l	œ:	opt+q
Thorn þ	opt+t	ß:	opt+s
Eth ð	opt+d	ø:	opt+o
Cedille ç	opt+c	f:	opt+f

Sie können weitere Zeichen auf diese Weise eingeben.

Steuerungszeichen

Mithilfe von EINFÜGEN ⇒ FORMATIVIERUNGSZEICHEN (für Open Office)
bzw. EINFÜGEN ⇒ SYMBOL (SONDERZEICHEN) (für MS Word)
können auch eine Reihe von Steuerzeichen eingegeben werden, z.B.

- Geviertstriche — (im Deutschen als Spiegelschrift, im Englischen als Gedankenstrich)
- Halbgeviertstriche – (minus, Gedankenstrich)
- geschützte Trennstriche und Leerzeichen (bei denen nicht getrennt wird),
- bedingte Trennstriche (bei denen getrennt wird, wenn sie am Wortende stehen)

Eine weitere wichtige Steuerungsfunktion ist der “weiche” Zeilenbruch, der keinen eigenen Paragraphen erzeugt. Er wird durch Shift + Return ausgelöst.

Schriftauszeichnungen

Eingabe von Schriftauszeichnungen

Sie haben die Möglichkeit, eine Reihe von Schriftauszeichnungen zu verwenden.
Beispiele mit typischen Verwendungsweisen:

- *Kursiv (italics)* für Objektsprache (Symbolleiste K; Tastatur option-I/K)
- **Fettschrift (boldface)** für Hervorhebungen (Symbolleiste F, Tastatur option-F/B)
- Unterstreichung (underline) für Hervorhebungen, Spezialzwecke – sparsam verwenden!
- KAPITÄLCHEN (SMALL CAPS) für Fachausdrücke, Bedeutungen, in manchen Stilen auch Autornamen.

Weiteres:

- Doppelt unterstrichen,
- punktiert unterstrichen,
- überstrichen
- durchstrichen,
- schattiert
- Konturschrift

Textmanipulation

Suchen und Ersetzen

Man kann in allen gängigen Textverarbeitungsprogrammen suchen und Such-Ausdrücke durch andere ersetzen.

In LibreOffice geschieht das in BERARBEITEN ⇒ SUCHEN & ERSETZEN.

Hierbei gibt es eine Reihe von wichtigen Optionen:

- Groß/Kleinschreibung beachten (sonst wird nicht unterschieden)
- Nur ganze Wörter (sonst auch Wortteile oder Folgen von Wörtern)
- Mehr Optionen:
 - Ähnlichkeitssuche
 - Attribute (z.B. Schriftfarbe) und Formate (z.B. Schriftarten, Absatzformate)
 - Reguläre Ausdrücke

Suche nach regulären Ausdrücken

Diese Option erlaubt es, Suchschemata zu definieren

- . steht für ein beliebiges Zeichen: "Schmi.t" findet "Schmidt" und "Schmitt"
- * steht für kein, ein oder mehr Zeichen: "aa*cc" findet "aacc", "aabcc", "aabbcc" etc.
- \n findet neue Zeilen (harter Zeilenumbruch)
- \$ findet einen neuen Absatz, ^\$ findet einen leeren Absatz.
- \t findet Tabulatorzeichen
- \< und \>: Findet Suchbegriff nur am Wortanfang oder Wortende, z.B. "buch\>" findet "Lesebuch" aber nicht "Buchmesse"
- [aeiou] steht für ein Zeichen in der Klammer, [a-e] steht für ein Zeichen zwischen a und e im ASCII-Code, z.B. [0-9] für eine Ziffer.

Beispiel für den Einsatz:

- Sortierte Wortliste aus einem Text erstellen:
 - a. Ersetze Leerzeichen durch \n
 - b. Sortiere das Resultat (EXTRAS ⇒ SORTIEREN)
- Leere Absätze entfernen: ^\$ durch nichts ersetzen.
- Absätze verdoppeln: \$ durch \n\n ersetzen.
- Postleitzahl (fünfstellig) finden: Suche nach [0-9]{5}

Erweiterte Suche / Ersetzen: Plugin „Alternative Find & Replace for Writer“

Änderungen nachverfolgen

Vor allem für die Zusammenarbeit mehrerer Autoren an einem Text ist die Funktion BEARBEITEN ⇒ ÄNDERUNGEN ⇒ AUFZEICHNEN sinnvoll.

Damit kann man einen Text redigieren, wobei alle Veränderungen angezeigt werden und der ursprüngliche Autor den Veränderungen zustimmen oder sie ablehnen kann.

Funktion “Bearbeiten ⇒ Rückgängig”

Hiermit können Sie Veränderungen wieder ungeschehen machen, und zwar über viele einzelne Veränderungen hinweg. Sie können diese Veränderungen wieder rückgängig machen und Schritt für Schritt den originalen Zustand wieder rekonstruieren.

Automatische Rechtschreibkorrektur

- Unter EXTRAS ⇒ RECHTSCHREIBUNG UND GRAMMATIK
- Gewünschte Sprache einstellen (kann je nach Formatvorlage variieren).
- Verlassen Sie sich keinesfalls allein auf die Rechtschreibkorrektur (z.B. *das* – *dass*)
- Ergänzen Sie das eigene Wörterbuch regelmäßig bei auftretenden Wörtern, die nicht erkannt werden, auch Namen – die Rechtschreibkorrektur wird dadurch wesentlich besser. .
- Für das Schreiben von Texten kann auch die Thesaurus-Funktion sinnvoll eingesetzt werden.

Kollaboratives Schreiben

Wenn mehrere Personen an einem Dokument schreiben wollen, empfiehlt sich die Verwendung einer geeigneten Plattform.

Beispiel: Google Docs, <https://docs.google.com/document/u/0/?hl=de&pli=1&showDriveBanner=true>

Tabellen

Systematische Daten können oft klarer in Tabellenform dargestellt werden.

LibreOffice (und auch MS Word) bietet dafür sehr gute Möglichkeiten mit der Tabellenfunktion; bei LibreOffice unter TABELLE ⇒ EINFÜGEN.

Sie können zunächst eine einfache Tabelle einfügen und diese dann nachbearbeiten.
– geben Sie sich nicht mit dem ersten Erscheinungsbild zufrieden!

Beispiel: Lateinische Deklination

	o-Stämme		u-Stämme	
	SG	PL	SG	PL
NOM	<i>hort-us</i>	<i>hort-ī</i>	<i>grad-us</i>	<i>grad-ūs</i>
ACC	<i>hort-um</i>	<i>hort-ōs</i>	<i>grad-um</i>	<i>grad-ūs</i>
GEN	<i>hort-ī</i>	<i>hort-ōrum</i>	<i>grad-ūs</i>	<i>grad-uum</i>
DAT	<i>hort-ō</i>	<i>hort-īs</i>	<i>grad-ui</i>	<i>grad-ibus</i>
ABL	<i>hort-ō</i>	<i>hort-īs</i>	<i>grad-ū</i>	<i>grad-ibus</i>

Grafiken

Für manche Zwecke sind grafische Darstellungen ein wichtiges Kommunikationsmittel.

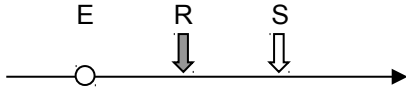
LibreOffice enthält ein leistungsfähiges Grafikwerkzeug (ANSICHT ⇒ SYMBOLLEISTEN ⇒ ZEICHNEN), mehr Möglichkeiten gibt es, wenn man die Zeichnung unter DATEI ⇒ NEU ⇒ ZEICHNUNG bearbeitet und dann in den Text kopiert.

Im einzelnen:

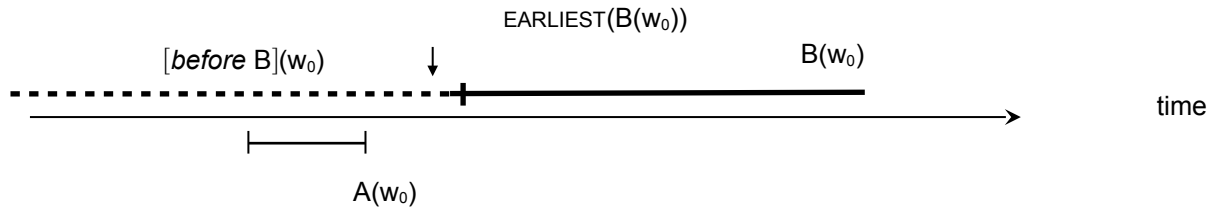
- Standardformen, Blockpfeile, Flussdiagramme, Legenden
- Striche und Pfeile
- Dynamische Verbindungen – besonders hilfreich!
- Veränderungen von Eigenschaften: Einzelne oder in Gruppen.
- Textfelder; Ausrichtung des Textes.
- Verankerung einer Grafik relativ zu einem Paragraphen, zur Seite, als Zeichen im Text; bei Verankerung als Zeichen im Text: Zeilenabstand einrichten.

Beispielgrafiken:

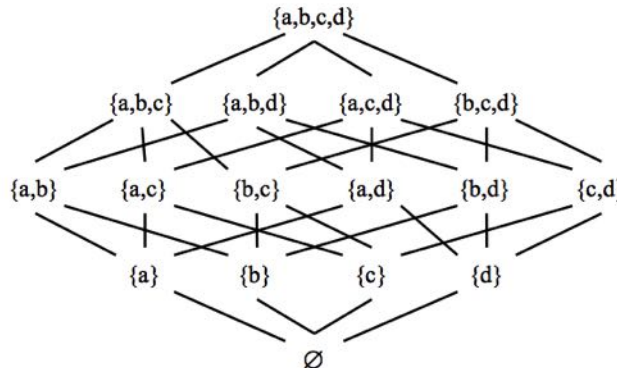
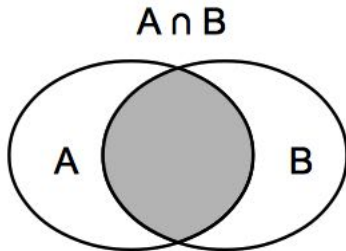
- Zeitverhältnisse: Plusquamperfekt



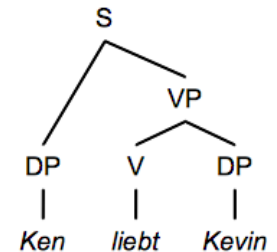
- Zeitverhältnisse: *before*



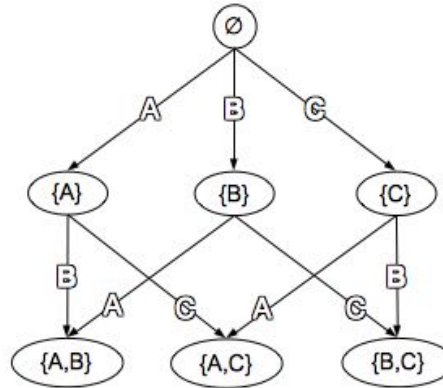
- Mathematische Diagramme



- Syntaktischer Baum



➤ Spieltheoretischer Baum



➤ Darstellung von Situationen: Verben mit kontinuierlicher (*ziehen*) vs. initialer Kontrolle (*werfen*)



➤ Bilder

Sie können in MS Word und in LibreOffice Bilder (Fotos) integrieren, in verschiedenen Formaten:

- GIF (graphic interchange format) ist ein weit verbreitetes nicht-komprimiertes Format, auch für Webseiten geeignet.
- TIFF ist ein proprietäres nicht-komprimiertes Format
- JPEG (joint photographic expert group) ist ein Format, das mit verschiedenen Graden der Datenkompression arbeitet.

Häufige Wege der Bildeinbettung:

- Sie können dies mithilfe von EINFÜGEN ⇒ Bild ⇒ AUS DATEI... bewerkstelligen, wenn sie eine Datei haben, die das Bild (und nichts anderes) enthält.
- Sie können Bilddateien aus dem Internet gezielt herunterladen (Mac: Ctrl+Klick, Windows: Rechtsklick).
- Sie können Bilder oder Bildausschnitte auch von ihrem Bildschirm selbst kopieren; in Windows verwenden Sie die PrintScreen-Taste und das Paint-Programm, beim Mac das Dienstprogramm Grab (Beispiel!)

Sie können das Bild vergrößern (was die Bildqualität beeinträchtigt) und verkleinern (was ebenfalls die Bildqualität verschlechtern kann).

Quelle für Bilder:

- Google Picture, mit Suchbegriffen

Bilder bearbeiten und in den Text einbinden

Bearbeitung von Bilddateien:

- Einige Möglichkeiten stellt das Textprogramm zur Verfügung (Zuschneide-Optionen, Helligkeit, Kontrast, Graustufen oder farbig)

Bei LibreOffice:

Bei ANSICHT ⇒ SYMBOLLEISTEN ⇒ BILD einstellen,
dann erscheint das Schaltfeld



Abbildung 1: Schaltfläche Bildbearbeitung

- Für weitere Optionen:
eigene Bildbearbeitungs-Programme, z.B. GIMP: <http://www.gimp.org/>
- Beschriftung von Bildern: Bild rechtsklicken / Control-Klicken, BESCHRIFTUNG wählen.

Wie Bilder angeordnet werden können:

- Als Zeichen (dann müssen Sie als Option “einzeilig” wählen)
- Im Quadrat (umliegender Text fließt um das Bild)
- Vor dem Text (Sie müssen selbst im Text für das Bild “Platz schaffen”)

Nicht die Option “Hinter dem Text” wählen – schwer, das Bild zu entfernen oder zu verändern!

Verankerung des Bildes:

- An der Seite, an einem Paragraphen, als Zeichen im Text.
- Vor oder hinter anderen Objekten.

Mischung von Bildern und Grafiken ist möglich, z.B. Beschriftungspfeile.

Textverarbeitung 3

Darstellung linguistischer Objekte

Phonetik

Zur einzelsprach-unabhängigen Darstellung von Sprachlauten werden die Zeichen des Internationalen Phonetischen Alphabets verwendet.

Liste mit Aussprachebeispielen: http://de.wikipedia.org/wiki/Liste_der_IPA-Zeichen

Konsonanten

Bei pulmonischen Konsonanten wird der ausströmende Luftstrom unterbrochen. Beschreibung nach Artikulationsort und –art. Aus <http://de.wikipedia.org/wiki/IPA>

	bilabial		labio-dental		dental		alveolar		post-alveolar		retroflex		palatal		velar		uvular		pha-ryngal		glottal	
	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.	stl.	sth.
Plosive	p	b					t	d			ʈ	ɖ	c	ɟ	k	g	q	ɢ			ʔ	
Nasale		m		ɱ				n				ɳ		ɲ		ŋ		ɴ				
Vibranten		ʙ						r										ʀ				
Taps/Flaps				ɸ				ɾ				ɽ										
Frikative	ɸ	β	f	v	θ	ð	s	z	ʃ	ʒ	ʂ	ʐ	ç	ʝ	x	ɣ	χ	ħ	ʕ	h	ɦ	
laterale Frikative							ɬ	ɮ														
Approximanten				ʋ				ɹ					ɻ	j		ɰ						
laterale Approximanten							l						ɭ	ʎ		ʟ						

Neben den pulmonischen Konsonanten gibt es auch Konsonantenverbindungen und Implosive, Ejektive und Klicks.

Non-pulmonic consonants

Clicks	⦿	ǀ	ǃ	ǂ	ǁ	
Implosives	ɓ	ɗ	ɟ	ɠ	ɡ	
Ejectives	p'	t'	c'	t'	k'	q'
	f'	θ'	s'	ʃ'	x'	χ'
	ts'	tʃ'	cʎ'	tʃ'	tʃ'	kx'

Affricates

pf	ts	dz	tʃ	dʒ	tʂ	dʑ	tʂ	dʑ
	tʃ	dʒ	cç	ɟʝ				

Co-articulated consonants

Fricatives	ɛ	z	ɸ	
Approximants	ɱ	w	ɥ	ɹ
Stops	kp	gb	ŋm	

Vokale

Bei Vokalen wird der Luftstrom nicht unterbrochen, aber durch Modulation des Mund-, Nasen- und Kehlräume modifiziert.

	vorne		fast vorne		zentral		fast hinten		hinten	
	ung.	ger.	ung.	ger.	ung.	ger.	ung.	ger.	ung.	ger.
geschlossen	i	y			ɪ	ʏ			ɯ	u
fast geschlossen			ɪ	ʏ				ɯ		
halbgeschlossen	e	ø			ə	ɐ			ɤ	o
mittel					ə					
halboffen	ɛ	œ			ɜ	ɞ			ʌ	ɔ
fast offen	æ				ɐ					
offen	a	œ							ɑ	ɒ

Suprasegmentalia

Eigenschaften, die über Sprachlaute (Segmente) hinausgehen, typischerweise silbenbezogen; z.B. : für Länge: [fal] Fall vs. [fa:l] fahl, für Töne etc.

Baumdarstellungen

Syntaktische Strukturen werden typischerweise durch Bäume dargestellt.
Hierfür gibt es im verschiedene Möglichkeiten:

Die Verwendung der Grafik-Option des Textverarbeitungsprogramms

In LibreOffice verwendet man hierzu am besten den Programmteil „Draw“,
denn damit kann man Textkästen verbinden und so den Syntaxbaum formatieren.

Formatierung der Textkästchen:

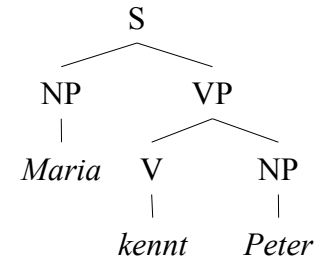
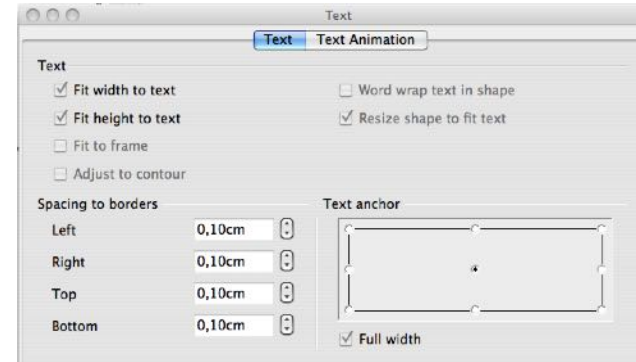
- Rechtsklick auf Kästchen, „Text“ auswählen.
- Textanker: Mittig,
- Randabstand: gering, 0,10 cm; Kästchen minimieren.
- Höhe, Weite mit Text abstimmen
- Unter „Schrift“: Schriftgröße und -Typ einstellen, z.B. 12pt TimesNewRoman
- Diese Einstellung muss nur einmal gemacht werden, dann kann man Kästchen kopieren und einfügen; verschieben mit Shift + Pfeiltaste für weite Schritte

Für die Verbindungslinien „Verbinder“ wählen:

Die Kästchen können dann verschoben werden, wobei die Verbindungslinien haften bleiben.

Am Schluss:

- den gesamten Baum markieren (mit Pfeilkursor) und gruppieren (Rechtsklick + Gruppieren)
- kopieren und
- in das Textdokument einfügen.
- Beachte: Bei Größenveränderung bleibt die Schrift konstant groß, was zu Verzerrungen führt – daher vermeiden.



Die Verwendung eines Web-Dienstes

php Syntax Tree -- <http://ironcreek.net/phpsyntaxtree/>?

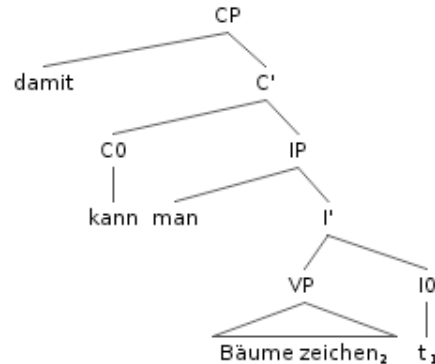
erlaubt die Darstellung von Bäumen aus einer Klammerstruktur. Die Bäume können dann heruntergeladen werden. Beispiel:

Satz od. Satzteil (mit eckige Klammern gekennzeichnet):

```
[CP damit [C' [C0 kann] [IP man [I' [VP Bäume zeichnen_2] [IO t_1]]]]]
```

Zeichnen

Offene Klammern: 7 Geschlossene Klammern: 7



Die Verwendung eines speziellen Programmes für Bäume: TreeForm

Siehe

<http://sourceforge.net/projects/treeform/?abmode=1>.

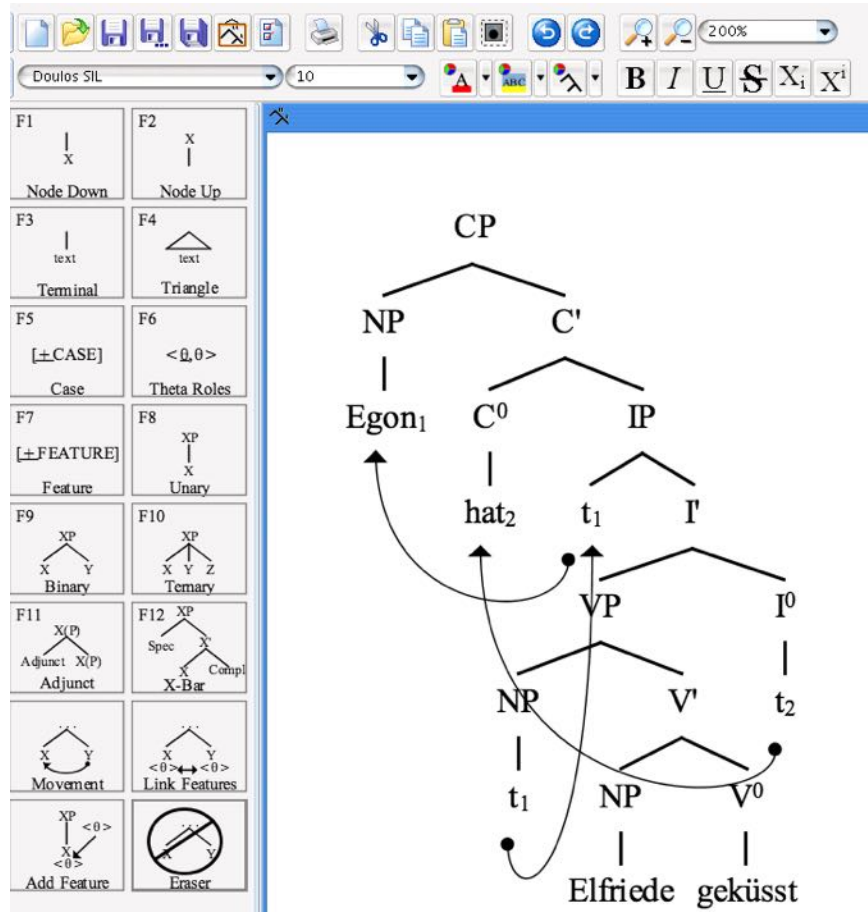
OpenSource-Software, kostenlos.

Läuft unter Windows, Macintosh OS X und Linux.

Man hat in TreeForm Einfluss auf die Gestaltung (Breite) eines Baumes, indem man vor und hinter den Knoten Leerzeichen eingibt.

Bei den Linien für syntaktische Bewegung gibt es keine direkten Einflussmöglichkeiten.

Hier kann man z.B. auch mit der Grafikfunktion des Textverarbeitungsprogramms arbeiten.



Darstellung von linguistischen Glossen

Interlinearübersetzungen: Die Leipzig Glossing Rules

Das am weitesten verbreitete Format:

siehe <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>
und als pdf-Datei auf der Moodle-Seite.

Die Beispiele werden typischerweise in drei Zeilen dargestellt:

- Das sprachliche Beispiele in einer zugänglichen Umschrift, oft mit Angabe der morphologischen Struktur.
- Eine Morphem-für-Morphem-Übersetzung; Empfehlung: kleinere Schrift
- Eine idiomatic Übersetzung, in einfachen Anführungszeichen.

Oft ist es sinnvoll, auch der Name der Sprache und die Quelle mit anzugeben.

Einige Regeln:

➤ Wort-für-Wort-Alinierung

(1) Indonesian (Sneddon 1996:237)

Mereka di Jakarta sekarang.

they in Jakarta now

'They are in Jakarta now.'

➤ Morphem-Korrespondenz

(2) Lezgian (Haspelmath 1993:207)

Gila abur-u-n ferma hamišaluğ güğüna amuq'-da-č.

now they-OBL-GEN farm forever behind stay-FUT-NEG

'Now their farm will not stay behind forever.'

Bei der Bezeichnung der Morpheme sollten möglichst standardisierte Abkürzungen verwendet werden. Lexeme werden in normaler Schrift angegeben, grammatische Morpheme in KAPITÄLCHEN.

➤ Verwendung von Punkten, wenn eine morphologische Zergliederung nicht möglich ist. Person und Numerus werden dabei ohne Punkt zusammengefasst.

(6) Turkish

çık-mak

come.out-INF

'to come out'

(9) German

unser-n Väter-n

our-DAT.PL father.PL-DAT.PL

'to our fathers'

(20) Italian

and-iamo

go-PRS.1PL (not: go-PRS.1.PL)

'we go'

Darstellungsmöglichkeiten nicht-overter Elemente und von Reduplikationen (Tilde)

(22) Latin

puer
boy[NOM.SG]
'boy'

or: *puer-Ø*
boy-NOM.SG
'boy'

(29) Hebrew

yerak-rak-im
green~ATT-M.PL
'greenish ones'

ATT: Attenuative

Darstellungsmöglichkeiten für Zirkumfixe und Infixe

(26) German

ge-seh-en
PTCP-see-PTCP
'seen'

or: *ge-seh-en*
PTCP-see-CIRC
'seen'

(28) Latin

reli<n>qu-ere
leave<PRS>-INF
'to leave'

(stem: *reliqu-*)

Darstellung inhärenter Merkmale wie Genus

(23) Hunzib (van den Berg 1995:46)

oz#-di-g *xõxe* *m-uq'e-r*
boy-OBL-AD tree(G4) G4-bend-PRET
'Because of the boy the tree bent.'

(G4 = 4th gender, AD = adessive, PRET = preterite)

Standard-Abkürzungen für grammatische Morpheme:

1 first person	DET determiner	OBL oblique
2 second person	DIST distal	P patient-like argument of canonical transitive verb
3 third person	DISTR distributive	PASS passive
A agent	DU dual	PFV perfective
ABL ablative	DUR durative	PL plural
ABS absolutive	ERG ergative	POSS possessive
ACC accusative	EXCL exclusive	PRED predicative
ADJ adjective	F feminine	PRF perfect
ADV adverb(ial)	FOC focus	PRS present
AGR agreement	FUT future	PROG progressive
ALL allative	GEN genitive	PROH prohibitive
ANTIP antipassive	IMP imperative	PROX proximal / proximate
APPL applicative	INCL inclusive	PST past
ART article	IND indicative	PTCP participle

AUX auxiliary	INDF indefinite	PURP purposive
BEN benefactive	INF infinitive	Q question particle/marker
CAUS causative	INS instrumental	QUOT quotative
CLF classifier	INTR intransitive	RECP reciprocal
COM comitative	IPFV imperfective	REFL reflexive
COMP complementizer	IRR irrealis	REL relative
COMPL completive	LOC locative	RES resultative
COND conditional	M masculine	S single argument of intrans.
COP copula	N neuter	SBJ subject
CVB converb	N- non- (e.g. NSG nonsingular)	SBJV subjunctive
DAT dative	NEG negation	SG singular
DECL declarative	NMLZ nominalizer	TOP topic
DEF definite	NOM nominative	TR transitive
DEM demonstrative	OBJ object	VOC vocative

Erzeugung von interlinearen Glossierungen in Textverarbeitungsprogrammen

Bei der Realisierung von interlinearen Glossierungen bereiten vor allem die Alinierungen von Wörtern Schwierigkeiten. Möglichkeiten:

- Tabstops. Definieren Sie Tabstops alle 0,5 cm; damit kann man in vielen Fällen eine ansprechende und lesbare Alinierung erzeugen.

(1) → *der* → → *grau-haar-ig-e* → → *Mann* → → *ist* → → *wieder-ge-komm-en*↵
 → → the.M.NOM → grey-hair-ADJ-NOM → man(M) → COP.3S → return-PTCP-come-PTCP↵
 → → 'The grey-haired man has returned.'[☞]

- Tabellen. Dieses Verfahren ist etwas umständlich, führt aber zu einem besseren Resultat. Schreiben Sie die beiden ersten Zeilen mit jeweils einem Tab zwischen Worten, definieren Sie den Abschnitt und gehen Sie auf TABELLE ⇒ EINFÜGEN

der *grau-haar-ig-e* *Mann* *ist* *wieder-ge-komm-en*
 the.M.NOM grey-hair-ADJ-NOM man(M) COP.3S return-PTCP-come-PTCP

<i>der</i>	<i>grau-haar-ig-e</i>	<i>Mann</i>	<i>ist</i>	<i>wieder-ge-komm-en</i>
the.M.NOM	grey-hair-ADJ-NOM	man(M)	COP.3S	return-PTCP-come-PTCP

Die angepasste Tabelle entsteht nach TABELLE ⇒ AUTOMATISCH ANPASSEN

Entfernen Sie die Ränder der Tabelle (Option: Kein Rand)

Darstellung von Formeln

Einfache Formeln können einfach im Text eingegeben werden.

Bei komplexeren Formeln empfiehlt sich die Verwendung des Formel-Editors,

z.B. bei LibreOffice: EINFÜGEN → OBJEKTE → FORMEL,

Informationen unter <http://www.oowiki.de/FormelEditor.html>,

<https://wiki.documentfoundation.org/images/3/37/MG40-MathGuide.pdf>

Beispiele für Klammerdarstellungen:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

```
left[ stack{a # b # c} right] `→` left[ stack{1 # 2 # 3} right]
```

Beispiele für Teilungsstrich:

$$\frac{3}{6} = \frac{1}{2}$$

```
3 over 6 `=` 1 over 2
```

$$p(H|B) = \frac{p(H) \cdot p(B|H)}{p(H) \cdot p(B|H) + p(\neg H) \cdot p(B|\neg A)}$$

```
p(H divides B) `=` {p(H) `cdot` p(B divides H)} over {p(H) `cdot`  
p(B divides H) `+` p(neg H) `cdot` p(B divides neg A)}
```

Seminararbeiten und Abschlussarbeiten

Ziel: Arbeitstechniken für die Herstellung eigener schriftlicher Arbeiten.

Wozu?

Vordergründig dienen schriftliche Arbeiten und Referate zur Bewertung von Studienleistungen. Aber sie sind mehr als das!

- Selbständige Erarbeitung eines Gebiets.
 - Man durchdenkt Daten und ihre möglichen Interpretationen genauer, wenn man sie kohärent und verständlich darstellen muss.
 - Man bezieht Stellung zu den Annahmen, die vertreten werden, und beginnt selbst Hypothesen zu entwickeln. Man entwickelt eigene Interessen.
 - Man überlegt sich, welche Methoden man einsetzen kann, um Hypothesen zu stützen oder zu widerlegen, und lernt auch, diese Methoden anzuwenden.
- Einübung in die Darstellung der Resultate der eigenen Arbeit.
 - Möglicherweise wichtiger als die vertiefte Kenntnis des Stoffes selbst!
- Vorbereitung für weiterführende Arbeiten.
- Die Abschlussarbeit kann wichtig für zukünftige Bewerbungen werden.

Fazit: Es ist wichtig, Referate und Hausarbeiten ernst zu nehmen – nicht nur wegen der Noten, sondern weil sie die besten Chancen liefern, Wissen und Fertigkeiten zu erwerben.

Worüber?

- Themenwahl: So eng wie möglich!
Sie sollen in der Arbeit zwar deutlich machen, dass Sie den wissenschaftlichen Hintergrund kennen, dann aber ein ganz bestimmtes Thema auswählen, das für diesen relevant ist.
Also nicht: *Höflichkeit in der Sprache*, sondern: *Der Sprechakt des Sich-Entschuldigens im Deutschen und im Ukrainischen*.
Und nicht: *Die Wortstellungsfreiheit im Deutschen*, sondern: *Fokussierung und Gegebenheit als konkurrierende Faktoren für die Stellung des direkten vs. indirekten Objekts*.
- Das Referat im Seminar stellt typischerweise eine wissenschaftliche Arbeit vor.
Um das gut zu machen, müssen Sie den Forschungshintergrund kennen.
Sie sollten nicht jeden einzelnen Punkt des Forschungshintergrundes darstellen,
Sie sollten vielmehr raffen, auswählen, zuspitzen, auf den Punkt bringen.
- Die Abschlussarbeit (z.B. Bachelorarbeit, Masterarbeit) geht typischerweise über die Wiedergabe einer einzigen wissenschaftlichen Arbeit hinaus: Es werden z.B. verschiedene Auffassungen gegenübergestellt und bewertet.
- Eine gute Abschlussarbeit sollte darüber hinaus auch auf eigenständigem empirischen Material beruhen und/oder versuchen, unser theoretisches Verständnis voranzubringen.

Wie?

- Fangen Sie frühzeitig mit der Vorbereitung und der Ausführung an.
Für die Bachelorarbeit haben sie zwei Monate Zeit, für die die Masterarbeit 5 Monate, Dissertationen werden in der Regel über 3 Jahre gefördert.
- Trennen Sie die Phasen des Stoffsammlens, des Organisieren des Stoffes, des Schreibens, des kritischen Wieder-Durchlesens, des Verbesserns und der Zuspitzung dessen, was Sie geschrieben haben.
- Behalten Sie hierbei den Hörer bzw. Leser im Auge: Der Zuhörer beim Referat sind neben dem Seminarleiter oder der Seminarleiterin vor allem ihre Mitstudent(inn)en, denen Sie sich verständlich machen wollen, denen Sie etwas sagen wollen.
- Bei der Hausarbeit wirkt es oft gezwungen, wenn Sie mit Hinblick auf den Seminarleiter schreiben. Stellen Sie sich einen bestimmten Adressatenkreis vor (das sollten Sie vorher mit dem Leiter absprechen), z.B. die Leser einer wissenschaftlichen Fachzeitschrift, Leser mit einem breiteren wissenschaftlichen Interesse oder auch interessierte Laien. Mögliche Formen:
 - Wissenschaftsjournalistischer Artikel
 - Forschungsantrag
 - Wikipedia-Artikel

Materialsammeln

Jede wissenschaftliche Arbeit schließt Phasen des Materialsammelns ein:

- Kommentierende Lektüre wissenschaftlicher Literatur:
Zusammenfassen der wesentlichen Argumentation,
evtl. mit eigenen bewertenden Kommentaren;
wichtige Daten oder Zitate im Original,
- Festhalten von eigenen empirischen Befunden:
eigene Beobachtungen, Resultaten von Experimenten und Korpusanfragen.
- Festhalten und Ausarbeiten von eigenen Ideen.
Ideal dazu: ein gebundenes Notizbuch.

Das Sammeln und Aufarbeiten von Material soll so erfolgen, dass es bei Bedarf abrufbar ist.

Hierfür gibt es verschiedene Techniken.

Der Zettelkasten

Das klassische Instrument war der **Zettelkasten**.

Zettel oder Karteikarten bestehen aus zwei Teilen:

- Einen Thema-Teil, in dem Sie z.B. den Autor und das Werk bezeichnen (wenn Sie publikationsweise exzerpieren) oder das Thema (z.B. *Historisches Präsens*) wenn Sie themenspezifisch exzerpieren. Der Thema-Teil dient zur Sortierung der Karten.
- Einen inhaltlichen Teil, in dem Sie Ihre Bemerkungen zu dem Werk bzw. zu dem Thema notieren. Hierbei ist wichtig, dass Sie zwischen (seltenen) Originalzitate, freien Inhaltsangaben und Ihren eigenen Bemerkungen klar unterscheiden (z.B. Zitate in Anführungsstrichen, mit Seitenzahlen; eigene Bemerkungen in eckigen Klammern).

Datensammlung mithilfe des Computers

Heute sind die Karteikästen und die damit einhergehende "Zettelwirtschaft" veraltet, man findet sie vielleicht noch billig beim Trödelmarkt.

Stattdessen nehmen wir entsprechende Daten auf dem Computer auf.

Das hat verschiedene Vorteile:

- Wir können die Daten automatisch durchsuchen,
- wir können die Daten unter verschiedenen Sortier- und Filterkriterien ansehen,
- wir können Text und auch Bilder direkt aus Quellen hineinkopieren (wenn der Text elektronisch zur Verfügung steht)
- wir sind nicht durch die physische Größe einer Karteikarte beschränkt.

Verwendung von Textverarbeitungsprogrammen

Sie können mit Ihrem Textverarbeitungs-Programm Informationen festhalten.

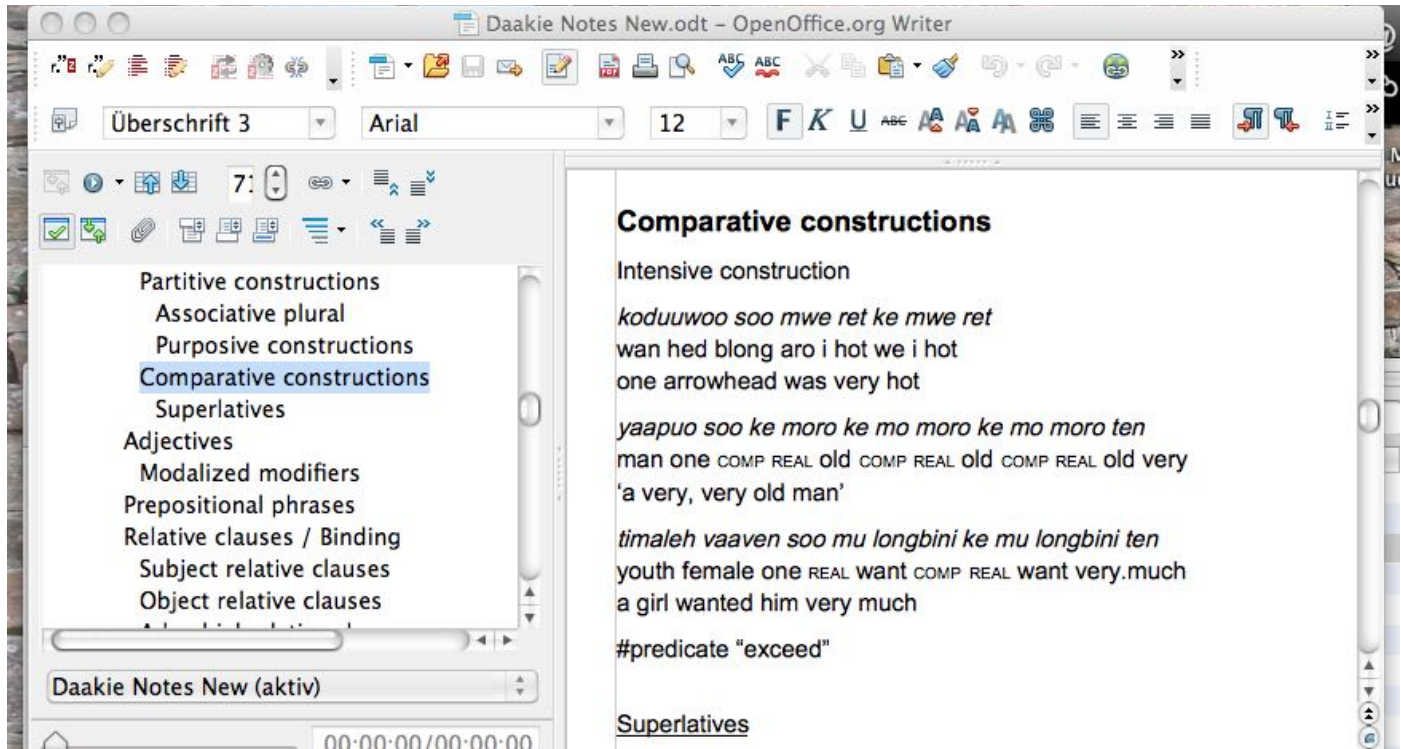
Es ist sinnvoll, dabei zumindest in einem Punkt dem Karteikarten-System zu folgen: Jede neue Informationseinheit beginnt mit einer neuen Seite (EINFÜGEN \Rightarrow WECHSEL \Rightarrow SEITENWECHSEL), und hat eine Überschrift (oder auch Unterüberschrift).

Vorteil: Sie können den Haupttext ausblenden:

- In LibreOffice: Mit Navigator arbeiten (ANSICHT \Rightarrow NAVIGATOR)
- In MS Word: (ANSICHT \Rightarrow NAVIGATIONSBEREICH \rightarrow DOKUMENTSTRUKTUR)

Die Einträge im Navigator sind kollabierbar.

Beispiel für Verwendung von Open Office zum Datensammlung:



Daakie Notes New.odt – OpenOffice.org Writer

Überschrift 3 Arial 12

Partitive constructions
Associative plural
Purposive constructions
Comparative constructions
Superlatives
Adjectives
Modalized modifiers
Prepositional phrases
Relative clauses / Binding
Subject relative clauses
Object relative clauses

Daakie Notes New (aktiv)

00:00:00/00:00:00

Comparative constructions

Intensive construction

koduuwoo soo mwe ret ke mwe ret
wan hed blong aro i hot we i hot
one arrowhead was very hot

yaapuo soo ke moro ke mo moro ke mo moro ten
man one COMP REAL old COMP REAL old COMP REAL old very
'a very, very old man'

timaleh vaaven soo mu longbini ke mu longbini ten
youth female one REAL want COMP REAL want REAL want very.much
a girl wanted him very much

#predicate "exceed"

Superlatives

Vor- und Nachteile der Datensammlung mithilfe des Textverarbeitungsprogramms:

- Sie können in MS Word und in LibreOffice Bilder kopieren und auch Links auf Internet-Adressen oder auf andere Dateien auf Ihrem Computer setzen, die dann nur mehr angeklickt werden müssen (EINFÜGEN ⇒ HYPERLINK).
- Ein weiterer Vorteil: Sie können z.B. bei Sprachbeispielen, syntaktischen Bäumen usw. bereits auf die Mittel zugreifen, die Sie später brauchen.
- Nachteil: Eingeschränkte Such- und Sortiermöglichkeiten. Sie können nur in geöffneten Dokumenten suchen (aber in allen geöffneten Dokumenten gleichzeitig).

Computer-Zettelkästen

Beispiel: Zkn³, nach Niklas Luhmann, <http://zettelkasten.danielluedecke.de/>

Für Windows, Mac und Linus, kostenlos.

Hauptansicht mit

- Überschrift
- Zettelinhalt
- Quellangabe
- Schlagwörter
- Sonstige Vermerke und Anhänge
- Schlagwortkatalog

Man kann

- nach Zetteln suchen
- Zettel in bestimmter Reihenfolge ordnen (z.B. als Basis für Gliederung einer Arbeit)

The screenshot shows the Zettelkasten software interface. The main window displays a note titled "Zettel 24 (112 Wörter)" with the heading "Historische Altersbilder (in der Kunst)". Below the heading is a painting depicting a group of people in historical attire. The text below the painting reads: "In dem hier gezeigten Bild über 'Lebensalter' erreicht das Paar erst ab 50 Jahren seinen Zenit, nach dem 90. Lebensjahr stirbt es erst." Below this is the source information: "Quellangabe (Autor, Literatur): Tews, Hans Peter (1991): Altersbilder. Über Wandel und Beeinflussung von Vorstellungen vom und Einstellungen zum Alter. Köln: Kuratorium Deutsche Altershilfe".

On the right side, there is a "Schlagwörter (4)" section listing: "Altersbilder", "Altersbilder in der Kunst", "historische Altersbilder", and "Lebenserwartung".

At the bottom right, there is a "Schlagwörter" section with a table showing the frequency of various keywords:

Schlagwörter	Häufigkeit
Abduktion	21
Abkopplung des Verstehens	5
Ablehnung der Pflege	7
Absicht	0
Abtreibung	1
Abweichungsverstärkung	2
Adressabilität	0
Adressabilität - Probleme	10
Adressenformular	2
Advanced Nursing Practitioner	2
Ageism	3
Agent	0
aggressives Verhalten	5
AGIL-Schema	1
Akten	1
Akteur	21
aktives Altern	11
aktivierende Pflege	6
Aktivitäten des täglichen Lebe...	8

At the bottom of the interface, it shows "Zettel 24 von 1906" and "(1955 Schlagwörter)".

Zettelkasten-Funktion können auch durch andere Programme erfüllt werden:

Synapsen

Hypertext-basierte Karteikartensimulation: <http://www.verzetteln.de/synapsen/>

Zotero

Möglichkeit von Notizen, bibliographischen Angaben, Sortieren, Suchen

DevonThink

Text (in verschiedenen Schriftauszeichnungen; Unicode), Bilder, Tondateien, Links usw. können integriert werden. Suchmöglichkeiten: Relevanz der Trefferstellen (Häufigkeit des Suchbegriffs) werden angezeigt. Die einzelnen Dateien können in Sammlungen und Unter-Sammlungen angeordnet werden.

Bibliographieprogramme

erlauben in der Regel die Angabe von Schlüsselwörtern, Notizen, Attachments.

Formales zu schriftlichen Arbeiten

Siehe http://fakultaeten.hu-berlin.de/philfak2/lehre/bachelor/ba0708/ba_arbeit_html
und auf der Moodle-Seite.

Einige wichtige Punkte:

Allgemeines

- Einmal gewählte Form beibehalten (Einheitlichkeit)
- DIN A 4, Blatt einseitig beschrieben, Zeilenabstand 1,5 Schriftgröße 12 Pkt, Rand links 4 cm, rechts 3 cm.
- Seiten fortlaufend nummerieren, außer Titelseite.
- Klammern (Seminararbeiten) oder binden (Abschlussarbeiten), keine losen Blätter abgeben.

Aufbau der Arbeit

- Titelblatt: Universität, Institut, Titel des Seminars, Semester, Name Seminarleiter
Titel der Arbeit, Name Verf., Matrikelnummer, Anschrift, E-Mail, Ablieferungstermin.
Für Abschlussarbeiten gibt es herunterzuladende Deckblätter.
- Inhaltsverzeichnis.
Empfohlen wird die Dezimalgliederung.
Gliederung muss folgerichtig aufgebaut sein.
Überschriften in Text und Inhaltsverzeichnis müssen übereinstimmen, Überschriften nur einmal vergeben.
Alle Gliederungspunkte mit Seitenzahlen im Inhaltsverzeichnis aufführen.
Bibliographie, Anhang, Abkürzungsverzeichnis etc. nicht in die Nummerierung einbeziehen, aber in das Inhaltsverzeichnis aufnehmen.
- Weitere Gliederung in:
 - Einleitung. Fragestellung, Ziel der Arbeit, Erklärung der Verfahrensweise und des Aufbaus.
 - Hauptteil.
 - Zusammenfassung und Ergebnisse.
- Textgestaltung:
 - Orthografie, Interpunktion, Setzung von Leerzeichen etc. Nach der "Duden-Rechtschreibung der deutschen Sprache", darin auch: Richtlinien für den Satz.
 - Die in der Linguistik üblichen Konventionen müssen eingehalten werden.

➤ Zitate:

Wörtliche Zitate originalgetreu wiedergeben.

Bei Auslassungen werden drei Punkte in eckigen Klammern gesetzt: [...].

Zusätze werden durch eckige Klammern markiert, z.B. [Hervorhebung M.K.].

Nicht satzförmige Zitate in den Satzzusammenhang einbetten.

Auch Gedankengänge, die nicht wörtlich wiedergegeben werden, sind als fremde kenntlich zu machen, z.B. (nach Haspelmath 1990: 47, vgl. Matzel 1970: 151).

Längere Zitate, die über mehrere Zeilen gehen, zur Hervorhebung links einrücken und einzeilig wiedergeben.

Unmittelbar nach dem Zitat ist die Quelle anzugeben.

➤ Fußnoten

Keine Endnoten. Der primäre Zweck ist nicht der Literaturverweis. Weiterführende Erläuterungen, Kommentare, Detailfragen. Sparsam einzusetzen!

➤ Bibliographische Verweise (bereits abgehandelt).

Umfang und Bearbeitungszeit

➤ Bachelorarbeit: 40 Seiten, Bearbeitungszeit zwei Monate, keine Überschneidung mit Hausarbeit.

➤ Masterarbeit: 60 Seiten, Bearbeitungszeit fünf Monate; Vorbereitung durch Modul 15: Forschungsliteratur.

Ein unangenehmes Thema: Plagiate

Selbständigkeitserklärung

Mit der Abgabe von Hausarbeiten und Abschlussarbeiten versichern Sie, dass Sie die Arbeit selbständig erstellt haben und alle Hilfsmittel angegeben haben. Sie müssen insbesondere bei Abschlussarbeiten die Selbständigkeitserklärung ausfüllen und beilegen, siehe http://fakultaeten.hu-berlin.de/philfak2/selbststaendigkeiterklaerung_neu2011.pdf

Damit versichern Sie, dass es sich um eine erstmalig, selbständig und ohne fremde Hilfe verfasste Arbeit handelt.

Sie erklären, dass Sie alle verwendeten fremden Quellen als solche kenntlich gemacht haben.

Sie bestätigen, dass Sie bei wörtlich übernommenen Aussagen, Tabellen, Grafiken als auch bei in eigenen Worten wiedergegebenen Aussagen und abgewandelten Tabellen und Grafiken (Paraphrasen) die Quelle angegeben haben.

Verstöße werden als Täuschung betrachtet und nach der Allgemeinen Satzung für Studien- und Prüfungsangelegenheiten der HU mindestens durch die Entscheidung des Nichtbestehens der Prüfung geahndet.

Plagiat

ist die Übernahme fremden geistigen Eigentums, die als eigene Leistung ausgegeben wird.

Dazu gehört:

- Die wörtliche Übernahme von Texten und Grafiken, ohne Urhebernennung.
- Die enge Umschreibung von Texten und Abwandlung von Grafiken ohne Urhebernennung.
- Ein Plagiat liegt auch dann vor, wenn fremde Arbeiten in ihrer Argumentation und ihrem Resultat bewusst übernommen werden, auch wenn die Textgestalt selbst sich nicht an das Original anlehnt.
- Bei Prüfungsleistungen gilt auch, dass eigene Werke oder Teile von Werken nicht, auch nicht in Teilen, wiederholt eingereicht werden dürfen. Im Zweifelsfall sollte man dies mit dem Betreuer / der Betreuerin der Arbeit besprechen.

Aber Achtung: Eine Arbeit, die vor allem aus Zitaten und Paraphrasen besteht, auch wenn diese korrekt Autoren zugeschrieben werden, ist problematisch. Sie sollten schon versuchen, selbst, mit eigenen Worten, zu formulieren. Wörtliche Zitate sollen vor allem dann gebraucht werden, wenn es wirklich auf den Wortlaut ankommt.

Aufmerksame Prüfer bemerken mögliche Plagiate durchaus, insbesondere an der Uneinheitlichkeit des Textes. Sie werden im Zweifelsfall „verräterische“ Stellen über das Internet identifizieren oder Plagiatsoftware einsetzen. Es lohnt sich also nicht, zu plagiierten!

Beurteilungskriterien für Bachelor-Arbeiten

1. Gliederung

Logik (Vollständigkeit und Relevanz der einzelnen Punkte). Durchsichtigkeit in Aufbau und Formulierung. Angemessenes Verhältnis der Einzelpunkte im Text

2. Inhalt

Erfassen des Themas. Vollständigkeit. Logik der Darstellung und Stringenz der Argumentation. Klarheit und Ausgewogenheit der Darstellung. Konzentration auf das Wesentliche.

3. Zitierweise

Angaben korrekt. Angaben vollständig, sinnvoll, in ausreichender Anzahl vorhanden. Form korrekt und systematisch.

4. Objektsprache

Einhaltung der fachüblichen Konventionen (Glossierung, Übersetzung, Nummerierung, Transliteration).

5. Bibliographie

Literatur selbständig recherchiert und kompetent ausgewählt. Angaben vollständig. Titel korrekt (richtige Auflage etc.). Äußere Form korrekt und systematisch.

6. Sprache

Grammatisch korrekter Sprachgebrauch. Klarheit der Ausdrucksweise. Angemessenheit des Stils.

7. Äußere Form

Anordnung und Form (Seitenanordnung etc.) sauber, einheitlich und systematisch.

Korrekte Orthographie und Interpunktion.

Einhaltung der Satzkonventionen (Kursive, Zitatkonventionen etc.).

Der Textumfang sollte 40 Seiten nicht überschreiten (für Bachelorarbeiten).

Referate

Zweck des Referants

Wichtig: Adressat des Referats sind Ihre Mitstudierenden! Ihnen sollen Sie Aspekte eines Themas aus dem Rahmen des Seminars vermitteln. Zugleich hilft es Ihnen, sich schwerpunkthaft mit diesem Thema vertraut zu machen, das Wichtige von dem weniger Wichtigen zu trennen und zu einer eigenen, neuen Darstellung zu gelangen. Ferner üben Sie ein, was Sie später in vielen Berufen tun werden: Die Aufarbeitung und öffentliche Darstellung von Information.

Bei der Erarbeitung des Referats sollten Sie sich fragen:

- Das Stoffgebiet ist meist zu groß, um jedes Detail darstellen zu können.
Was ist für die Zuhörer wichtig, was halte ich selbst für wichtig, wie kann ich es nachvollziehbar und interessant, vielleicht sogar spannend darstellen?
- Die Zeit für das Referat ist beschränkt – Wie kann ich diese Zeit optimal nutzen?
Bleibt dann auch noch genügend viel Zeit für die Diskussion übrig?
- Referate sollten nicht abgelesen werden!
Man kann sich an eine schriftliche Struktur halten oder geeignete Wendungen einprägen, das Referat sollte jedoch ein mündlicher, frei gehaltener Beitrag sein.
Gerade dadurch gewinnt es an Verständlichkeit.
- Wie kann ich das Referat visuell oder auch auditiv unterstützen?
Handout, Folien, Beamer-Präsentation,
auch der Tafelanschrieb haben ihre Vor- und Nachteile.

Tischvorlagen

Für Handouts spricht:

- Zuhörer können sie annotieren und später wieder lesen. Sie zwingen den Zuhörer nicht sklavisch an die jeweilige Stelle der Darstellung.

Handouts sollen das mündliche Referat unterstützen, nicht ersetzen. Sie sollten also nicht den Text des ganzen Referats enthalten.

Es gibt zwei extreme Formen des Handouts, eine Minimal- und eine Maximalversion:

- Handouts, die nur das Material zeigen (Beispielsätze, Zitate, Bilder, Tabellen)
- Handouts, die auch Fragestellungen, Behauptungen, Argumentationen darstellen.

Der zweite, explizite Typ ist zu bevorzugen. Der schriftliche Kanal sollte jedoch den mündlichen möglichst wenig stören. Deshalb:

- Verknappte sprachliche Darstellungen: Stichpunkte, kurze Sätze!
- Unterstützung der Argumentation durch das Layout, z.B. Abschnittseinteilungen, Spiegelstriche ("Bullet lists"), Nummerierung von Beispielen, Abbildungen usw.
- Eine feinteilige Numerierung von Abschnitten, Unterabschnitten, Beispielen usw. hilft, die Aufmerksamkeit der Zuhörer auf bestimmte Textstellen zu lenken.
- Sparsam mit Platz umgehen, aber bitte die Schrift des Handouts nicht zu sehr reduzieren (z.B. beim doppelseitigen Drucken). Mindestgröße: 10pt.

Das Handout soll darüber hinaus auch die verwendeten Quellen angeben.

Präsentations-Software

Eine Alternative zu Handouts besteht in der Verwendung von Overhead-Folien und von Präsentations-Software wie MS Powerpoint, LibreOffice Impress oder Apple Keynote.

Alle Punkte, die bei Referaten erwähnt wurden, treffen auch hier zu.

Darüber hinaus hat die Präsentations-Software aber auch einige Nachteile:

- Man kann die visuelle Unterstützung nicht später nochmal durchlesen.
- Man kann während des Vortrags nicht zurückspringen, um einen Punkt nachzulesen, oder vorausspringen, um Aspekte des Gesamtplans zu erkennen.
- Man hat als Zuhörer oft kein Gefühl, an welcher Stelle im Vortrag man sich befindet.

Beide Nachteile kann man wettmachen, wenn man auch ein Handout mit dem Inhalt der Präsentation verteilt. Dann allerdings hat die Präsentations-Software Vorteile:

- Der Vortragende kann die Aufmerksamkeit auf einen bestimmten Punkt (auf die Stelle der gegenwärtigen Folie) lenken.
- Man kann Schaubilder und komplexe Grafiken dynamisch entwickeln und so ein komplexes Gebilde verständlicher darstellen als in einer statischen Abhandlung. (Allerdings ist hierzu sorgfältige Planung nötig, und man läuft wiederum Gefahr, dass man die Zuhörer durch grafische Mätzchen verwirrt oder abstößt).

Einige Hinweise zu Präsentations-Software:

- Hintergrund möglichst weiß (manchmal wird blau mit weißer Schrift empfohlen)
- serifenlose Schrift,
- Zeilenfall sinnvoll einsetzen, d.h. Zeilenbruch nach Sinneinheiten,
- kurze, knappe Formulierungen,
die Folie ist kein Vortragsmanuskript!
- Zusammengehörendes möglichst auf eine einzige Seite.
- Die Möglichkeit des Mediums nutzen: Grafiken, Bilder, Diagramme.
- Es ist sinnvoll, auch einen Ausdruck mitzugeben
(ideales Layout: 4 Folien auf eine DIN A 4-Seite).

Hinweis für nächste Sitzung zur linguistischen Argumentation:

Lesen Sie den folgenden Artikel (wird auf der Moodle-Seite bereitgestellt):

Katrin Axel (2009), „Die Entstehung des *dass*-Satzes“, in V. Ehrich, Ch. Fortmann, I. Reich, M. Reis (Hg.), *Koordination und Subordination im Deutschen*. Sonderheft 16 der *Linguistischen Berichte*, 21-42.

Argumentation in Fachartikeln

Zielsetzung

Wir wollen uns an einem Beispiel die linguistische Argumentation und die Darstellung linguistischer Argumente ansehen; hierzu wird es auch die **zweite Hausaufgabe** geben.

Wir wählen dazu aus:

- Katrin Axel (2009), „Die Entstehung des *dass*-Satzes – Ein neues Szenario“, in V. Ehrich, Ch. Fortmann, I. Reich, M. Reis (Hg.), 2009. *Koordination und Subordination im Deutschen*. Sonderheft 16 der *Linguistischen Berichte*, 21-42

Der Artikel ist eine Zeitschriftenpublikation in den *Linguistischen Berichten*, allerdings in einem Sonderheft (einer Themenausgabe) zu Koordination (Satzverknüpfung) und Subordination (Satzeinbettung).

Es handelt sich damit dem Typ nach eher um einen Beitrag zu einem Buch (Sammelband) als um einen Zeitschriftenbeitrag.

Bei solchen Beiträgen sollte man:

- sich die Zielsetzung des Sammelbandes insgesamt deutlich machen,
- das Vorwort und die Einleitung der Herausgeber lesen.

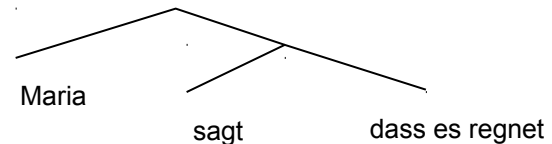
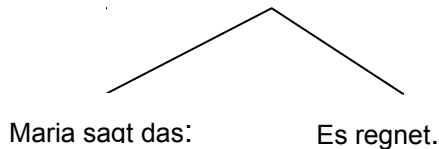
1 Fragestellung

Argumentrealisierende *dass*-Sätze gehören im **Gegenwartsdeutschen (Gwd.)** zu den häufigsten Nebensatztypen überhaupt. Die **Lehrmeinung** ist, dass der *dass*-Satz im Deutschen aus einer parataktischen Konstruktion hervorgegangen ist. Dieses **Parataxe-zu-Hypotaxe-Szenario** lässt sich an den nhd. Beispielen in (1) verdeutlichen. Ausgangspunkt ist eine Sequenz aus zwei Hauptsätzen. Am Ende des ersten Hauptsatzes steht das Demonstrativpronomen *das*, das kataphorisch auf den Inhalt des zweiten Hauptsatzes verweist. Aus dieser Konstruktion hat sich, so die Annahme, durch Übertritt von *das* über die Satzgrenze ein komplexer Satz entwickelt bestehend aus einem Haupt- und einem Argumentsatz, der nunmehr durch die Konjunktion *dass* eingeleitet wird.

Abkürzung:
Gwd.
nhd: neu-
hochdeutsch
kataphorisch:
vorausweisend
Parataxe etc.
siehe unten.
Übertritt:
Veränderung im
Sprachwandel

(1) **Maria sagt das: Es regnet.** → **Maria sagt, dass es regnet.**

- Gegenstand: *dass*-Sätze, die ein Argument eines Verbes realisieren.
- Darstellung der allgemeinen Meinung (“Lehrmeinung”) der historischen Entstehung: Parataxe (Satznebenordnung) reanalysiert als Hypotaxe (Satzunterordnung)



Nachschlagewerke

Bei Verständnisschwierigkeiten sollten Sie Nachschlagewerke benutzen,
z.B. Hadumod Bußmann, *Lexikon der Sprachwissenschaft*:

- **Argument:** (2) In der ⇒ GB-Theorie NOAM CHOMSKYS ein referenzfähiger Ausdruck, dem eindeutig eine ⇒ Thematische Rolle, d.h. eine logische Argumentstelle eines Prädikats entsprechen muss.
(...)
- **Parataxe:** [griech. *parátaxis* 'Beiordnung'] Syntaktische Verknüpfung von Sätzen durch Nebenordnung (im Unterschied zur unterordnenden Verknüpfung, vgl. ⇒ Hypotaxe).
- **Katapher.** [griech. *kata-phérein* 'hindurchtragen'] [...] ein Sprachelement [...] das auf folgende Information innerhalb eines Äußerungskontexts vorausweist. [...]

Beschreiben und Exzerpieren

Es ist günstig, den Argumentationsverlauf mit entstehenden Fragen zu skizzieren,
(am Rand des Artikels, auch im PDF, in einem Notizbuch, auf einer Karteikarte oder in einem Programm wie Zkn³).

Thema: Entstehung von dass-Sätzen.

Autor: Axel 2009

1. Standardthese, Lehrmeinung: Parataxe zu Hypotaxe (PzH);
[Maria sagt das:] [es regnet] → Maria [sagt [dass es regnet]]

(Wer vertritt PzH?)

Bei dem **Standardszenario** gibt es, wie im folgenden Abschnitt argumentiert werden wird, eine Reihe von **empirischen und konzeptuellen Problemen**. In diesem Beitrag wird daher eine neue Hypothese vorgestellt und argumentiert, dass sich **der argumentrealisierende *dass*-Satz aus einer korrelativen Konstruktion entwickelt hat, in der *dass* als Relativpartikel fungierte.**

Im Gegensatz zum Standardszenario werden in dem neuen Ansatz die folgenden Fragen getrennt betrachtet: **Erstens, die Entstehung der Konjunktion *dass*** (d.h. einer Kopfkategorie) aus dem Demonstrativpronomen *das* (d.h. einer Phrasenkategorie). **Zweitens, die Entstehung der Komplementationsstruktur**, in der der *dass*-Satz die Schwester des Matrixverbs ist.

Hinweise auf Probleme der allgemeinen Meinung:

- empirische: Daten können nicht erklärt werden.
- konzeptuelle: Erklärung ist nicht vereinbar mit gängigen theoretischen Vorstellungen

Vorwegnahme der neuen Erklärung:

- *dass*-Sätze haben sich aus **korrelativen Konstruktionen** entwickelt (was das genau ist, wird hoffentlich später klar werden).

Wesentliche argumentative Unterschiede:

- Wie entstand die Konjunktion *dass* aus dem Pronomen *das*?
- Wie wurde der *das*(s)-Satz zu einem Komplement des Verbs (z.B. *sagen*)?

*Standard
szenario* =
Lehrmeinung
Matrixverb:
Verb des
Matrix-
satzes, d.h. des

übergeordneten
Satzes
Schwester:
Schwesterknote
n

Der Argumentationsverlauf kann nun weiter skizziert werden:

Thema: Entstehung von *dass*-Sätzen.

Autor: Axel 2009

1. Standardthese, Lehrmeinung: Parataxe zu Hypotaxe (PzH);
[*Maria sagt das:*] [*es regnet*] → *Maria [sagt [*dass es regnet*]]*
- Wer vertritt PzH?
2. Neue These: Entwicklung aus korrelativer Konstruktion (KK).
 - 2a) Entstehung von *dass* (Kopfkategorie (?)) aus *das* (Phrasenkategorie (?))
 - 2b) Entstehung der Komplementationsstruktur (*dass*-Satz Schwester des Matrixsatzes)

Wir erwarten also Argumente gegen PzH und für KK.

Grundannahmen für die interne Syntax des *dass*-Satzes

Was die syntaktische Position der Konjunktion *dass* im Gegenwartsdeutschen angeht, so wird in der generativen Literatur seit der Analyse von den Besten (1983)¹ angenommen, dass sie in der so genannten **COMP- oder C-Position basisgeneriert** wird:

(2) [_{CP} [_C *dass*] es geregnet hat]

Im Gwd. gibt es eine **komplementäre Distribution zwischen der Basisgenerierung eines Komplementierers in C und der Verbbewegung**: Die Basisposition des finiten Verbs befindet sich am Satzende; denn das Deutsche ist eine OV-Sprache. Wenn kein Komplementierer in C eingesetzt wird, bewegt sich das **finite Verb obligatorisch in die C-Position**:²

(3) [_{CP} [_C hat_i] es geregnet t_i]?

Also:

- *dass* in Nebensätzen befindet sich in der Position, in der das finite Verb in Hauptsätzen steht.
- (das wird später klar): *dass* ist eine C⁰-Kategorie, d.h. keine expandierbare Phrase, sondern Kopf einer Phrase (der CP).

Hier werden die Begriffe der X-bar-Syntax verwendet, ein prominentes Modell innerhalb der generativen Grammatik.

generative Literatur:
ein prominentes Grammatik-Modell.

komplementäre Distribution:
Auftrittsbeschränkungen
in verschiedenen Kontexten

Es geht nun offensichtlich um Hintergrunds-Annahmen für die Diskussion der beiden Thesen:

Thema: Entstehung von *dass*-Sätzen.

Autor: Axel 2009

1. Standardthese, Lehrmeinung: Parataxe zu Hypotaxe (PzH);
[*Maria sagt das:*] [*es regnet*] → *Maria [sagt [dass es regnet]]*

Wer vertritt PzH?

2. Neue These: Entwicklung aus korrelativer Konstruktion (KK).

2a) Entstehung von *dass* (Kopfkategorie (?)) aus *das* (Phrasenkategorie (?))

2b) Entstehung der Komplementationsstruktur (*dass*-Satz Schwester des Matrixsatzes)

Hintergrund-Annahmen:

Syntaktische Position von *dass* im GwD: Basisgeneriert in C von CP:

[_{CP} [_C *dass*] [*es geregnet hat*]]

Thema: Entstehung von *dass*-Sätzen.

Autor: Axel 2009

1. Standardthese, Lehrmeinung: Parataxe zu Hypotaxe (PzH);
[*Maria sagt das:*] [*es regnet*] → *Maria [sagt [*dass es regnet*]]*

Wer vertritt PzH?

2. Neue These: Entwicklung aus korrelativer Konstruktion (KK).

2a) Entstehung von *dass* (Kopfkategorie (?)) aus *das* (Phrasenkategorie (?))

2b) Entstehung der Komplementationsstruktur (*dass*-Satz Schwester des Matrixsatzes)

Hintergrund-Annahmen:

Syntaktische Position von *dass* im GwD: Basisgeneriert in C von CP (*dass* in C)

[_{CP} [_C *dass*] [*es geregnet hat*]],

Externe Position von *dass*-Sätzen im GwD: Nachgestellte Adjunkte zu V (NachAdj)

Maria hat [_{VP} [_{CP} *dass es regnet*] *gesagt*] ⇒ *Maria hat* [[_{VP} _{t_i} [_{V0} *gesagt*]] [_{CP} *dass es regnet*]_i]

umstritten. Problematisch ist beispielsweise, dass in manchen süddeutschen Varietäten **aus einem *dass*-Satz extrahiert werden kann**. Bei bestimmten Matrixprädikaten (den sog. Brückenprädikaten) ist es in diesen Varietäten möglich, eine *w*-Phrase (oder eine nicht-interrogative XP) aus dem *dass*-Satz in die SpecC-Position des Matrixsatzes zu bewegen:

(4) Wen_i hat Peter behauptet, [_{CP} t_i [_C *dass*] Maria t_i getroffen hat]?

Nach Huangs (1982: 505) *Condition on Extraction Domain* gilt, dass **nur aus strikt regierten Domänen extrahiert werden kann**.³ Das Extraktionsargument⁴ und weitere Evidenzen haben Weibelhuth (1992), Bayer (1995), (2001a), (2001b) und Haider (1996) dazu bewogen einer alternativen Analyse den Vorzug zu geben. Sie gehen davon aus, dass ***dass*-Sätze im Nachfeld als Komplemente von V⁰ in nicht-kanonischer Position basisgeneriert** sind:

- Alternative Meinung: Der *dass*-Satz ist im Nachfeld basisgeneriert, da aus dem *dass*-Satz Konstituenten bewegt werden können und dies nur aus selbst nicht bewegten Konstituenten möglich ist.

Maria hat [_{VP} [_{V0} *gesagt*] [_{CP} *dass es regnet*]_i]

Brückenprädikat:
z.b. *behaupten*

strikt regiert:
abhängig von
einem Prädikat
wie *behaupten*

nicht-kanonisch:
nicht-standard

basisgeneriert:
in dieser Position
erzeugt, d.h. nicht
nachträglich in
diese Position
bewegt.

Darstellung von Gegenargumenten:

Thema: Entstehung von *dass*-Sätzen.

Autor: Axel 2009

1. Standardthese, Lehrmeinung: Parataxe zu Hypotaxe (PzH);
[*Maria sagt das:*] [*es regnet*] → *Maria [sagt [dass es regnet]]*

Wer vertritt PzH?

2. Neue These: Entwicklung aus korrelativer Konstruktion (KK).

2a) Entstehung von *dass* (Kopfkategorie (?)) aus *das* (Phrasenkategorie (?))

2b) Entstehung der Komplementationsstruktur (*dass*-Satz Schwester des Matrixsatzes)

Hintergrund-Annahmen:

Syntaktische Position von *dass* im GwD: Basisgeneriert in C von CP (*dass* in C)

[_{CP} [_C *dass*] [*es geregnet hat*]],

Externe Position von *dass*-Sätzen im GwD: Nachgestellte Adjunkte zu V (*dass* NachAdj)

Maria hat [_{VP} [_{CP} *dass es regnet*] *gesagt*] → *Maria hat* [[_{VP} _{t_i} [_{V0} *gesagt*]] [_{CP} *dass es regnet*]]_i

Gegenargument zu NachAdj: Dann sollte man nicht extrahieren können, im Süddtsch. aber möglich:

Wen₁ hat Peter behauptet, [dass Maria t₁ getroffen hat]?

Gegenposition zu NachAdj: *dass*-Satz als Komplement basisgeneriert. (*dass*-Basis)

Maria hat [_{VP} [_{V0} *gesagt*] [_{CP} *dass es regnet*]]

Auf jeden Fall ist der *dass*-Satz ein **Komplement** (und nicht nur ein Adjunkt).

Dafür spricht auch, dass *dass*-Sätze als Komplement von Präpositionen wie *bis* und *ohne* auftreten:

- (5) **Bis daß der Tod uns scheidet. / Ohne daß eine Zahlungsaufforderung ergeht.**
(Reis 1997: 130, Bsp. (29))

Nicht expliziert gemachter Hintergrund:

Es gibt vier lexikalische X^0 -Kategorien (Verben, Adjektive, Nomina, Präpositionen).

Dass-Sätze treten als Komplemente von allen vier Kategorien

auf:

- V^0 : *behauptet, dass Hans die Silbermedaille gewonnen hat*
- A^0 : *Maria ist stolz, dass Hans die Silbermedaille gewonnen hat.*
- N^0 : *das Gerücht, dass Hans die Silbermedaille gewonnen hat*
- P^0 : *bis dass Hans die Silbermedaille gewonnen hat*

Konkretisierung der Fragestellung

Somit gibt es im Gwd. sowohl Evidenzen dafür, dass (auch nachgestellte) *dass*-Sätze syntaktische Komplemente im engeren Sinne sind als auch dafür, dass *dass* eine X^0 -Kategorie ist.⁵ Zwar scheinen dies triviale Feststellungen zu sein, aus diachroner Sicht ist jedoch beides erklärungsbedürftig: Zum einen kann kein Zweifel daran bestehen, dass die Konjunktion *dass* aus dem Nominativ/Akkusativ des Pronomens *das* hervorgegangen ist. Es muss also eine Reanalyse von einer XP- zu einer X^0 -Kategorie stattgefunden haben. Zum anderen gibt es Evidenzen, dass sich argumentrealisierende *dass*-Komplementsätze aus einer korrelativen Konstruktionen entwickelt haben, in denen der *dass*-Satz ein attributives oder explikatives Verhältnis zur Korrelats-XP hatte. Das würde bedeuten, dass der *dass*-Satz ursprünglich ein Adjunksatz war, der als Komplementsatz reanalysiert wurde. Mit diesen beiden Aspekten (bzw. Reanalysen) beschäftigt sich der vorliegende Artikel. Dabei wird nur auf *dass*-Sätze in Ob-

Komplemente im engeren Sinn:
Komplemente, die keiner nachträglichen Adjunktion unterliegen.
Argument:
Komplement
Adjunkt:
angalagert, Adverb

Evidenz für:

- *dass*-Sätze sind Komplemente, die nicht nachträglich adjungiert werden.
- *dass* ist eine X^0 -Kategorie, keine Phrase (XP)

Historische Evidenz (noch zu erbringen):

- *dass*-Sätze sind aus Korrelativen entstanden, also ursprünglich aus Adjunkten
- *dass* ist aus dem Demonstrativ *das* entstanden, also ursprünglich eine Phrase (DP)

Überblick über die folgende Untersuchung

Die Untersuchung gliedert sich wie folgt: In Abschnitt 2 werden zwei ältere Szenarien zur Entstehung der Konjunktion *dass* bzw. von *dass*-eingeleiteteten Argumentsätzen vorgestellt und diskutiert. In Abschnitt 3 werden zwei verschiedene Diagnostika für Komplementsatzstatus (Extraktionstransparenz, Kombinierbarkeit mit Präpositionen) auf *dass*-Sätze im Althochdeutschen (Ahd.) und Mittelhochdeutschen (Mhd.) angewandt. Dies alles mündet in Abschnitt 4 in einem neuen Vorschlag zur Entstehung des *dass*-Argumentsatzes im Deutschen.

Vorausschau am Ende der Einleitung erleichtert die folgende Lektüre.

2. Ältere Szenarien: Standardmeinung

Wie eingangs erwähnt, hat sich der Lehrmeinung zufolge der **Komplementierer *dass* (ahd. *thaz*)** aus dem Akkusativ/Nominativ Neutrum des Demonstrativpronomens ***der/die/das*** entwickelt. Diese Entwicklung wird immer wieder als ein Beispiel für ein **typisches Parataxe-zu-Hypotaxe-Szenario** diskutiert. Das Standardszenario, das bereits von den Junggrammatikern postuliert wurde, ist in (6) bis (8) illustriert. Ich folge der Darstellung in Wunder (1965: 255ff.), ähnliche Überlegungen finden sich in Behaghel (1877, 1928: 130), Paul (1920: 241), Müller / Frings (²1959), Horacek (1964), Helgander (1971, Kap. 7 und 156ff., 181ff.), Fleischmann (1973: 18ff.), Ebert (1978: 26), Dorchenas (2005) u.v.a.m.

Die Entstehung des *thaz*-Satzes muss **schon in vorahd. Zeit** erfolgt sein. Jedoch wird in der Regel argumentiert, dass im Ahd. noch Residuen von Vorgängerkonstruktionen belegt sind, welche die Entstehungsgeschichte synchron widerspiegeln.

Als Belege werden v.a. Zitate aus *Otfrid* gebracht, d.h. dem Evangelienbuch von Otfrid von Weißenburg (ca. 790-875).

Hier: Faksimile der Heidelberger Handschrift:

<http://diglit.ub.uni-heidelberg.de/diglit/cpl52/>



Drei Formen von Satzanbindungen im *Otfrid*; metrische Zäsur

- (6) a. [HS ...] [HS ...]
b. [kuntā imo,] [er iz wólta,]
verkündete ihm er es wollte
,(er) verkündete ihm: Er wollte es'
(Otfrid I 25,10)

Nebengeordnete
Sätze ohne
verbindenden
Ausdruck.

- (7) a. [HS ... thaz] [HS/NS? ...]
b. [joh gizálta in sar tház,] \ [thiu sálida
und erzählte ihnen sogleich das die Seligkeit
untar in was.]
unter ihnen war
,und er erzählte ihnen sogleich das: Die Seligkeit war unter ihnen'
(Otfrid II 2,8)

Nebengeordnete
Sätze mit
verbindendem
kataphorischen
Pronomen *tház*

- (8) [[HS ...] [NS thaz ...]]
[[Druhtin Krist irkánta,] \ [thaz er mo war zalta,]]
Herr Christus erkannte dass er ihm Wahrheit erzählte
,Christus, unser Herr, erkannte, dass er ihm die Wahrheit erzählte'
(Otfrid II 12,11)

Untergeord-
neter Satz
mit Kon-
junktion
thaz,
Evidenz:
Position der
metrischen

Gegenargument: Zäsur ≠ Satzgrenze

Es ist allerdings fragwürdig, ob die Zäsur in Otfrid ein zuverlässiger Indikator für die syntaktisch-strukturellen Verhältnisse darstellt. Denn es finden sich sogar Belege dafür, dass die Zäsur Konstituenten spalten kann. Zum Beispiel folgt die Zäsur manchmal auf einen Determinierer in einer Nominalphrase:

- (10) Nust siu gibúrdinot [**thes**⁷ \ kindes so dfures]
nun-ist sie beladen des Kindes so teures
,nun ist sie mit dem so teuren Kind schwanger'
(Otfrid I 5,61)

Gegenargument: Komplementierlose Sätze bereits veraltet

Das beschriebene Szenario setzt weiterhin voraus, dass es selbständige komplementiererlose Sätze mit Verbendstellung gab, vgl. die Struktur in (6).

Der einzige Text, in dem eine substantielle Anzahl von Belegen anzutreffen ist, ist Otfrid, wobei sich hier das Problem ergibt, dass in diesem Text der relativ hohe Anteil an Endstellung auch durch den Endreim ausgelöst worden sein kann. In den Prosatexten ist die Verbstellungsasymmetrie zwischen eingeleiteten und uneingeleiteten Sätzen im Ahd. bereits fest etabliert (Axel 2007: Kap.

Gegenargument: Theoretisch fragwürdige Annahmen

Was allerdings aus theoretischer Sicht sehr problematisch ist, ist der Übergang von Struktur (7) zu Struktur (8), denn dieser umfasst genau betrachtet drei Entwicklungsschritte: Eine Reanalyse der Satzgrenzen, im Zuge derer *thaz* eine Konstituente des nachfolgenden Satzes wird. Eine Reanalyse der Satzfügung, im Zuge derer sich der ehemals selbständige zweite Satz zum Komplement des vorangehenden Satzes entwickelt. Schließlich eine Reanalyse des Pronomens *thaz*, welches sich zu einer X^0 -Kategorie (Komplementierer) entwickelt. Sowohl die Verschiebung der Satzgrenze als auch die Degradierung eines Hauptsatzes zum Komplementsatz sind sehr radikale Reanalyseschritte. Ein Szenario, das auf solche radikalen Reanalysen verzichtet, wäre wesentlich plausibler.

Reanalyse:
syntaktische
Veränderung
im Sprach-
wandel

Fragwürdige Annahmen:

- Verschiebung der Satzgrenze: [A *tház*] [B] ⇒ [A] [*tház* B]
- Degradierung zum Komplementsatz: [*Krist irkanta*] [B] ⇒ [*Krist* [*irkanta* [B]]]
- Kategorienwechsel (nicht so kritisch): [A [_{DP} *tház*]] [B] ⇒ [A [[_{Co} *tház*] B]]

Alternatives Szenario: Reanalyse von Relativsätzen

Als alternatives Szenario wurde vorgeschlagen, dass sich der *dass*-Satz aus einem freien Relativsatz entwickelt hat und dass entsprechend die Konjunktion *dass* aus dem Relativpronomen *das* hervorgegangen ist. Horacek (1957: 422f.) diskutiert diese Möglichkeit mit Bezug auf gotische Daten. Ausgangspunkt der Entwicklung seien Sätze des folgenden Typs:

- (11) hausideduþ þatei qiþan ist.
hörte das gesagt ist
(Matth. 5,38; zit. nach Horacek 1957: 422)

Horacek führt aus, dass der *þatei*-Satz synchron sowohl als ein freier Relativsatz analysiert werden könne (‘ihr hörte, was gesagt worden ist’) als auch als ein Komplementsatz (‘ihr hörte, dass gesagt worden ist’). Im ersten Fall sei *þat* ein Pronomen gefolgt von der Relativpartikel *ei*, im zweiten Fall sei *þatei* eine Konjunktion. Sie behauptet, dass aus diachroner Sicht die letztere Konstruktion aus der ersteren hervorgegangen ist. Veranschaulichen kann man sich den von ihr postulierten Wandel vom Pronomen zur Konjunktion wie folgt:

- (12) [_{CP} þat_i [_C ei] [_{IP} t_i qiþan ist]] → [_{CP} [_C þatei] [_{IP} qiþan ist]]

Freier
Relativsatz:
Relativsatz ohne
Kopf, z.B.
*er hört, was
gesagt wurde*

Vgl.: *hört
das(jenige) was
gesagt wurde*

So hat das
Horacek 1957
sicher nicht
gesagt – eine
moderne
Rekonstruktion

Neue Evidenz aus dem Althochdeutschen:

Daten wie der von Horacek angeführte gotische Beleg finden sich auch in ahd. Quellen, (13). Wie die Paraphrasen veranschaulichen, ist bei dem Tatianbeleg jeweils sowohl die Lesart möglich, bei der *thaz* ein Relativpronomen ist wie auch die Lesart, bei der es eine Konjunktion ist.

- (13) ther heilant thô gisehenti thaz her spâhlihho antlingita
der Heiland da sehend das/dass er vernünftig antwortete
quad Imo. ...
sagte ihm
,Der Heiland, als er da sah, was/dass er vernünftig geantwortet hatte,
sagte zu ihm:'
ihesus autem uidens quod sapienter respondiss& dixit illi
(Tatian 433,3)

➤ Vorteile des alternativen Szenarios:

Dieses Szenario hat gegenüber dem im letzten Abschnitt beschriebenen den Vorteil, dass keine Verschiebung der Satzgrenze und kein radikaler Wandel des zweiten Satzes vom Haupt- zum Nebensatz angenommen werden muss: *thaz* ist auch in der Relativsatzkonstruktion eine Konstituente des zweiten Satzes⁸ und letzterer ist auch vor der Reanalyse ein Nebensatz.

Was sich geändert hat, ist lediglich, dass das Pronomen zum Komplementierer re kategorisiert wurde ($[_{CP} \text{thaz}_i [C] \ t_i \dots] \rightarrow [_{CP} [C \ \text{thaz}] \dots]$) und dass der zweite Nebensatz, der ursprünglich ein Adjunktsatz war, als Komplementsatz reanalysiert wurde ($[_{VP} [_{VP} \dots] [\text{thaz} \dots]] \rightarrow [_{VP} [\text{thaz} \dots]]$). Allerdings gibt es

Allerdings bestehen diese Vorteile nicht bei Horacek wörtlich genommen.

Dies ist wird in einem Nebenargument gezeigt, das wie andere nicht zentralen Argumente in eine Fußnote ausgelagert wurde:

⁸ Horacek (1957) geht allerdings davon aus, dass das Relativpronomen ursprünglich dem ersten Satz angehörte und von dort in den zweiten Satz übertrat. Sie paraphrasiert dies in Bezug auf das in (11) zitierte gotische Beispiel wie folgt: „Ihr hörtet das: Gesagt ist worden ...“, woraus „über die Sinnggebung“ (ibid. 423) „Ihr hörtet das, was gesagt worden ist ...“ geworden sei. Diese Entwicklung setzt ebenfalls eine radikale Verschiebung der Satzgrenzen und einer Herabstufung des zweiten Satzes vom Haupt- zum Nebensatz voraus. Dieses Entstehungsszenario für das demonstrative Relativpronomen ist jedoch umstritten. Alternativ wurden auch Szenarien diskutiert, in denen das Relativpronomen schon immer dem Nebensatz angehörte (vgl. Lenerz 1984). Für die Frage, wie der Komplementierer *thaz* entstanden ist, ist diese Problematik zunächst irrelevant. Denn Horaceks Annahme, dass das Relativpronomen als Komplementierer re kategorisiert wurde, ist eine separate, eigenständige Hypothese, unabhängig davon wie das Relativpronomen entstanden ist.

- Nachteil des alternativen Szenarios: Unplausible Einschränkung für das Verb des zweiten Satzes.

auch hier ein **Problem**: Es muss nämlich die Bedingung erfüllt sein, dass das **Verb des Folgesatzes ein fakultatives Akkusativobjekt selegiert** wie etwa *ant-lingōn*. Genauer ausgedrückt scheint uns aus gwd. Perspektive dieser Reanalyse-Schritt deshalb plausibel, weil im Gwd. beim Verb *antworten* in der neuhochdeutschen Übersetzung des Nebensatzes in (13) das interne Argument nicht syntaktisch realisiert ist. Ob das im Ahd. auch möglich war, kann man ohne Grammatikalitätsurteile durch Muttersprachler nicht mit Sicherheit beantworten.

seligiert:
efordert

Erläuterung:

Die Analyse setzt einen Nebensatz mit Akkusativobjekt *das/thaz* voraus:

als er sah, [_{CP} REL_i [*er vernünftig* t_i *geantwortet hatte*]]

⇒ *als er sah*, [_{CP} [_C KONJ] *er vernünftig geantwortet hatte*

dass-Sätze sind aber nicht auf Sätze mit Akkusativobjekt beschränkt.

Es folgt eine Zwischenüberlegung:

Komplementeigenschaft von thaz

Im Ahd. sind *das*-Korrelate häufiger als im Gwd. [Gegenwartsdeutschen] belegt:

- (14) Er tháhta odowila tház, \ thaz er ther dúriwart wás,
er dachte vielleicht das dass er der Türhüter war
,Er dachte vielleicht, dass er der Türhüter war'
(Otfrid II 4,7)

Vgl. *Er wusste das, (nämlich) dass er der Türhüter war.*

Außerdem gibt es im Ahd. mehr Nullpronomina, sodass auch Fälle ohne *thaz*-Korrelat ein Pronomen im ersten Satz haben könnten. Das hieße, dass *thaz*-Sätze keine Komplementsätze wären, sondern Adjunksätze. Man muss daher explizit nachweisen, dass es sich um Komplementsätze handelt.

Korrelat:
Pronomen im
Hauptsatz,
das dem
Nebensatz
entspricht.

Bewegung aus *thaz*-Sätzen

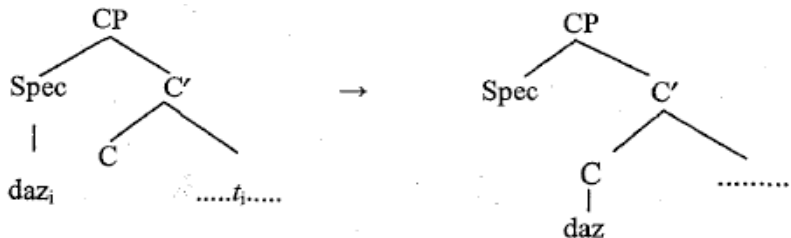
Theoretisches Argument: Extraktion aus Komplementen möglich, aus Adjunkten nicht.

- *Wen hast du vermutet, dass Otto getroffen hat?* (süddeutsch)
- **Wen hast du das vermutet, dass Otto getroffen hat?*

- (16) a. **uuenan_i** uuollet ir *thaz* ih iu **t_i** forlázze, barrabban oda
 wen wollt ihr dass ich euch begnadige Barabas oder
 then heilant
 den Heiland
 ,wen wollt ihr, dass ich begnadige, Barabas oder den Heiland?’
 Quem ergo uultis dimittam uobis barabban an ihesum
 (Tatian 631,30)

Eigene Befunde
 aus dem Ahd.,
 neue empirische
 Evidenz

Hinweis für den Komplementstatus und für die Reanalyse von *thaz* von einer Phrase in Spec-CP zu einem Koipf in C⁰.



Selektion von *thaz*-Sätzen durch Präpositionen

Beispiel: Notker (Spät-Ahd.)

- (21) Tie niht ánderes ze_demo tôde nebráhta . . . Áne dáz sie
die nichts anderes zu_dem Tod NEG.brachte ohne dass sie
uuâren gerárte nâh mînemo síte ...
waren aufgezogen nach meiner Sitte
,denen nichts anderes den Tod brachte, als dass sie nach meiner Sitte wa-
ren ...'

Dies spricht ebenfalls für den Komplementstatus von *thaz*-Sätzen.

Neue Hypothese

Eine bisher nicht berücksichtigte Klasse von Belegen

- (29) a. nist nīaman thero frīunto, \ thaz mīr zi thiū gihēlfe ...
NEG-ist niemand- derer Freunde THAZ mir zu dem helfe
,Es ist keiner unter den Freunden, der mir dabei hilft'
(Otfrid III 4,23)

thaz ist hier kein Pronomen (da das Antezedens maskulin ist), sondern Einleitung eines Relativsatzes [ähnlich Englisch *that* in *the man that came*].

Vgl. auch Beispiele der folgenden Art, in denen *thaz* kein Pronomen sein kann, da das entsprechende Satzglied im Satz vorkommt:

- (30) a. Wanta āllaz thaz sies thénkent, \ sie íz
denn alles THAZ sie-es.GEN denken sie es
al mit góte wirken;
alles mit Gott wirken
,denn alles, auf das sie sinnen, wirken sie mit Gott'
(Otfrid I 1,105)

Reanalyse Relativpronomen ⇒ Relativsatz-Subordinator

Da es also Evidenzen dafür gibt, dass *thaz* im Ahd., zumindest in der **Varietät Otfrids von Weissenburg**, neben seiner unstrittigen Verwendung als **demonstratives bzw. relatives Pronomen und Konjunktion** auch als **Relativpartikel** belegt ist, liegt ein alternatives Szenario für die Entstehung der Konjunktion nahe, das auch aus komparativ-typologischer Perspektive unabhängige Bestätigung findet.

Es liegt auf der Hand, dass sich die Relativpartikel durch eine Reanalyse des Relativpronomens entwickelt hat:

(32) a. [_{CP} *thaz*_i [_C] [... *t_i*]] → b. [_{CP} *Op_i* [*thaz*] [... *t_i* ...]]

Hinweis, dass dies eine häufig zu beobachtende Erscheinung ist, z.B. auch in deutschen Dialekten:
der Ort, wo das passiert ist ⇒ *der Mann, wo ich kenne*.

Neues Szenario für Entstehung von *dass*-Sätzen aus Relativsätzen

Man kann nun spekulieren, dass in vorahd. Zeit die *thaz*-Relativsätze quasi-explikativ zu einer ‚dummy NP‘ wie z.B. *thaz* im vorausgehenden Satz hinzutreten konnten.²⁴ Der ko-indizierte Relativsatz wurde durch die Relativpartikel *thaz* eingeleitet, (33a). Wenn man nun annimmt, dass die ‚dummy NP‘ nicht immer overt realisiert werden musste, sondern als (evtl. kontextuell lizenziertes) Nullpronomen auftreten konnte (als Objekt- oder Subjekt-*pro*), dann liegt folgendes Reanalyseszenario nahe:

- (33) a. ... $thaz_i V^0 [_{CP} OP_i [thaz] \dots]_i$
b. ... $pro_i V^0 [_{CP} OP_i [thaz] \dots]_i \rightarrow \dots V^0 [_{CP} [thaz] \dots]$

Mit anderen Worten:

- *thaz* wurde vom Relativpronomen zum Relativsatz-Subordinator.
Damit einher geht die kategoriale Veränderung von *thaz* von XP (DP) zu X⁰ (C⁰)
- Pronomen *thaz* im Hauptsatz und Null-Pronomina konnten durch *thaz*-Relativsätze modifiziert werden
- Reanalyse: Das Objektverhältnis wurde von dem Pronomen oder dem leeren Pronomen auf den *thaz*-Satz übertragen.
- Wie die Extraktionsdaten zeigen, war dieser Schritt schon im Ahd. vollzogen.

dummy NP:
Nominalphrase,
die sich auf nichts
bezieht, z.B. *es*
in es kommt
jemand.

Vergleich der Hypothesen

Dieses neue Szenario hat folgende Vorteile: Es gab keine radikale Reanalyse der Satzgrenze bei der Entstehung des *thaz*-Satzes, denn *thaz* gehörte sowohl vor als auch nach der Reanalyse dem Nebensatz an. Es fand kein radikaler Wandel im Status des entstehenden *thaz*-Argumentsatzes statt: Dieser war auch vor der Reanalyse ein Nebensatz. Was sich geändert hat, war lediglich, dass er vom Adjunktsatz zum Komplementsatz wurde. Das neue Szenario findet darüber hinaus indirekte Bestätigung durch die komparative Evidenz, dass sich in zahlreichen indo-europäischen Sprachen Komplementierer aus Relativ- bzw. Interrogativpronomina entwickelt haben. Die bislang herrschende Lehrmeinung, dass sich die Konjunktion *thaz* aus dem Demonstrativpronomen (Neutr., Nom/Akk.) entwickelt hat, muss nach dem hier propagierten Szenario wie folgt revidiert werden: Die Konjunktion *thaz* hat sich aus der Relativpartikel *thaz* entwickelt, die ihrerseits aus dem Relativpronomen (Neutr., Nom/Akk) hervorgegangen ist. Das Relativpronomen ist aus dem Demonstrativpronomen entstanden.

Theorienvergleich:

Annahme	Evidenz / Problem	Standard	Horacek	Axel
Reanalyse der Satzgrenze: [A <i>thaz</i>] [B] \Rightarrow [A] [<i>thaz</i> B]	vielleicht möglich; keine klare Evidenz dafür	✓	(✓)	
Hauptsatz \Rightarrow Nebensatz: [A B] \Rightarrow [A [B]]	oft angenommen: Parataxe \Rightarrow Hypotaxe Einzelheiten unklar	✓	(✓)	
Komplementierer DP \Rightarrow C ⁻	im Nebensatz	keine klare Evidenz dafür		
	im Relativsatz	Viele Beispiele		✓
Freier Relativsatz \Rightarrow Komplementsatz	Problem: Keine Einschränkung auf Akkusativargumente		✓	
Expliziter Kopf \Rightarrow Impliziter Kopf (Nullpronomen)	Nullpronomen gut belegt			✓
Relativsatz mit Null-Kopf \Rightarrow Komplementsatz	Neue Annahme; belegt z.B. in Bengali			✓

Wichtig: Argumente für eine Annahme sollten unabhängig gestützt sein
(durch weitere Evidenz oder durch etablierte theoretische Vorstellungen)

Empfohlen: Nochmalige Lektüre des Artikels

- Wie wird die argumentative Grundstruktur dargestellt?
- Wie werden die Argumente eingeführt?
- Wie wird die empirische Evidenz dargestellt?
- Wie unterstützen Vorausschau und Zusammenfassung die Argumentation?
- Wie unterstützt die Gliederung die Argumentation?
- Welche Teile der Argumentation werden in die Fußnoten verlagert?

Zweiter Test

Hintergrund: In diesem Test geht es um die Verwendung von linguistischen Schreibkonventionen in Textverarbeitungsprogrammen. Sie können dabei LibreOffice / OpenOffice, MS-Word oder LaTeX verwenden.

Aufgabe:

Stellen Sie auf drei Seiten die Argumentation, oder Teile davon, in Katrin Axel (2009), „Die Entstehung des dass-Satzes – Ein neues Szenario“ dar. Dieser Text soll in sich schlüssig sein. Er soll durch Überschriften gegliedert sein (das würde man sonst bei einem einseitigen Text nicht machen!), er soll drei nummerierte linguistische Beispiele enthalten, dabei mindestens eine Interlinearübersetzung, er soll mindestens einen Verweis auf ein Beispiel enthalten, er soll einen Bezug zu Objektsprache im laufenden Text enthalten, er soll Sonderzeichen (z.B. das Thorn, þ) enthalten, und er soll mindestens zwei Literaturangaben enthalten (2 Punkte).

Der Text soll dabei die einschlägigen linguistischen Konventionen der Textdarstellung und der bibliographischen Angabe befolgen (1 Punkt).

Die editierbare Text-Datei soll mit Formatvorlagen (für Titel, Überschriften, den Textkörper) geschrieben sein und einen automatischen Querverweis auf ein nummeriertes Beispiel enthalten (1 Punkt).

Der Text soll eine von Ihnen selbst konstruierte linguistische Baumdarstellung enthalten (1 Punkt).

Benennen Sie diese Datei mit ihrem Nachnamen und laden Sie diese in zwei Versionen auf die Moodle-Datei:

(a) als editierbare Text-Datei mit LibreOffice, MS-Word oder LaTeX-Datei, (b) als pdf-Datei.

Korpuslinguistik

Gegenstand der Korpuslinguistik

Was ist ein Korpus?

- eine Sammlung von Texten, die nach bestimmten Kriterien ausgewählt wurden
- schriftsprachliche oder sprechsprachliche Korpora
- sprechsprachliche Korpora auch mit Videoaufzeichnungen (z.B. für Gesten)

Was macht man mit einem Korpus?

- Korpusbelege dienen als Evidenz für den Sprachgebrauch
(und indirekt für die zugrundeliegende mentale Repräsentation von Sprechern,
obwohl diese Ziele manchmal im Gegensatz zueinander gesehen werden)
- Verwendung von **Einzelbelegen**, z.B. für bestimmte grammatische Phänomene;
- Verwendung des gesamten Korpus zur Untersuchung von **Häufigkeiten**
bestimmter sprachlicher Merkmale.

Was macht die Korpuslinguistik?

- Entwicklung von Methoden, um linguistische Fragestellungen anhand von Sprachdaten empirisch zu überprüfen.
- Aufbau von Korpora: Auswahl von Texten, Digitalisierung
- Befassung mit Urheberrechtsfragen.
- Aufbereitung von Texten: Annotation von grammatischen u. anderen Informationen
- Auswertung von Korpora: Suchprogramme, Suchanfragen, Statistik

Korpus-Arten

- Textkorpora:
Sammlung von Texten (geschriebener Text, transkribierte gesprochene Sprache)
- Annotierte Korpora
Korpora, in denen die Originaldaten mit Annotationen versehen sind.
 - Getaggte Korpora: Korpora mit Annotationen (“tags”) zu Wortarten.
 - Baum(daten)banken: Korpora, in denen Sätze syntaktisch geparkt wurden, d.h. in denen ihnen eine syntaktische Struktur zugewiesen wurde.
- Sprachkorpora
Sprachdaten (Sprachsignal u. U. mit Transkription, phonetische Annotation)
- Multimodale Korpora
Sprachdaten mit Transkription und weiteren Information wie Mimik, Gestik.

Für eine differenziertere Darstellung:

- Lemnitzer, Lothar & Heike Zinsmeister (2010, 2.Aufl.).
Korpuslinguistik. Eine Einführung. Tübingen: Narr.
- Perkuhn, Rainer, u.a. (2012). *Korpuslinguistik*. München: W. Fink, UTB.
- Materialien der korpuslinguistischen Seminare von Prof. Anke Lüdeling und Mitarbeiter(innen).

Schritte zur Korpuserstellung

- Auswahl des Korpus (Kriterien),
z.B. Zeitungssprache, Belletristik, Sachtexte, juristische Texte,
Jugendsprache, Sprache von Lernern des Deutschen, Arzt-Patienten-Kommunikation ...
- Beschaffung und Digitalisierung des Korpus:
Texte, die bereits digitalisiert zur Verfügung stehen;
eingescannte Texte mit OCR oder abgetippte Texte;
transkribierte Sprachaufnahmen
- **Tokenisierung**: Identifikation von kleinsten Einheiten, zum Beispiel:
Text: *Peter sagt, ihm mache schachspielen Spaß.* (6 Wörter)
Tokens: *Peter sagt, ihm mache schach spielen Spaß .* (9 Tokens)
Listen: *Peter sagt, ihm mache schach.spielen Spaß .*

Bei der automatischen Tokenisierung und Listenbildung kann es Probleme geben
– z.B. bei der Satzende-Erkennung, da der Punkt z.B. auch bei Abkürzungen auftritt.

- **Lemmatisierung**: Identifikation der Grundformen von flektierten Formen
Peter sagt[sagen], ihm[er] mache[machen] schach.spielen Spaß .
- **Tagging**: Angabe von Wortarten (“Parts of Speech” – POS)
Peter.NAME sagt[sagen][3.SG.PRÄS] mache[machen][3.SG.KONJ-1] ...
Eine anspruchsvolle Aufgabe, da u.a. Ambiguitäten erkannt werden müssen.
- Zuweisung von syntaktischen Strukturen bei geparsten Korpora.

Mögliche Annotationsebenen

Morphosyntax:	Wortarten
Morphologie:	Flexionsmorphologie
Lemma:	Grundformen
Syntax:	Konstituentenstruktur, Felderstruktur (Vor- und Mittelfeld, Satzklammer...)
Semantik:	Lesarten, thematische Rollen (Agens, Patiens...)
Pragmatik:	Koreferenz Antezedens-Pronomen,
Informationsstruktur:	Topik-Kommentar, Fokus...
Orthographie, Fehler, phonetisch/phonologische Merkmale, Gestik Mimik Blickbewegung	

Zweck der Annotation: Markierung von Eigenschaften, die sich aus dem Text selbst nicht oder nicht leicht automatisch erkennen lassen.

Das DWDS-Korpus

Grundsätzliches

- DWDS: *Digitales Wörterbuch der deutschen Sprache*
- Projekt an der Berlin-Brandenburgischen Akademie der Wissenschaften.
- Verbunden mit dem Wörterbuch der deutschen Gegenwartssprache (WDG)
- Zugang: www.dwds.de
ältere Panel-Ansicht www.eins.dwds.de
- Anonyme Verwendung möglich,
Registrierung empfohlen: Ermöglicht Zugriff auf mehr Ergebnisse und erweiterte Funktionen.

Aufbau Kernkorpus

- Deutsche Sprache des 20. und 21. Jahrhunderts, zeitlich und inhaltlich ausgewogen
- Umfang: > 120 Millionen Wörter aus ca. 80,000 Dokumenten
- Abdeckung: 1900 – 2000 (Erweiterungen sind vorgesehen; Belletristik und Zeitungen bis 2010)
- Jedes Jahrzehnt soll gleich stark vertreten sein.
- Jede Textsorte von vier Textsorten soll gleich stark vertreten sein:
Schöne Literatur (ca. 26%), Journalistische Prosa (ca. 27%), Fachprosa (ca. 22%),
Gebrauchstexte (ca. 20%), transkribierte Texte gesprochener Sprache (ca. 5%)

Deutsches Textarchiv: <http://www.deutschestextarchiv.de/>

- Texte 1600 – 1900
- Belletristik, Gebrauchsliteratur, Wissenschaft

Weitere Korpora:

Corpus Gesprochene Sprache, *ZEIT-Korpus*, das Korpus der *Berliner Zeitung* und des *Tagespiegel*, der *Potsdamer Neuesten Nachrichten*, ein *DDR Corpus*, ein *Wendecorpus*.

Siehe [HTTP://WWW.DWDS.DE/RESSOURCEN/KORPORA/](http://www.dwds.de/ressourcen/korpora/)

Korpus	Tokens	Sätze	Dokumente
Kernkorpus	125.990.080	7.046.937	79.312
Berliner Zeitung	242.046.373	15.951.701	869.023
Bild	121.520.037	12.629.828	548.181
Süddeutsche Zeitung	453.945.194	29.125.790	1.099.920
Tagesspiegel	184.202.717	10.392.257	394.465
WELT	238.403.711	15.787.624	600.007
ZEIT	417.422.714	23.631.230	499.520
Gesamt	1.783.530.826	114.565.367	4.090.428


Beispiel: Suchanfrage *ausgerechnet*

Mit dem neuen Interface:

D W D S Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart. [Anmelden](#)

ausgerechnet


ausgerechnet

Grammatik partizipiales Adverb
Aussprache 
Worttrennung aus-ge-rech-net (computergeneriert)
Grundform ↗ [ausrechnen](#)

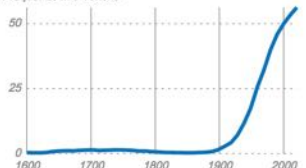
Bedeutung

eWDG, 1967

Worthäufigkeit

selten  häufig

Frequenz / Mio Tokens



Jahr	Frequenz (Mio Tokens)
1600	0
1700	0
1800	0
1900	0
2000	50

Ältere Wörterbücher

Korpusbelege (DWDS-Kernkorpus)

ausgerechnet 🔍 ?

Korpus: DWDS-Kernkorpus

Start: 1900 **Ende:** 1999

Textklassen: Belletristik Wissenschaft Gebrauchsliteratur
 Zeitung

Anzeige: KWIC voll maximal

Sortierung: Datum absteigend

Anzahl Treffer pro Seite: 50

- (1626)
- DWDS-Kernkorpus 21 (620)
- Deutsches Textarchiv (160)
- Zeitungskorpora
 - Berliner Zeitung (11972)
 - Tagesspiegel (8601)
 - Die Zeit (30081)
- Spezialkorpora
 - Blogs (2907)
 - Polytechnisches Journal (185)
 - Filmuntertitel (1020)
 - Gesprochene

1-50 von 913 Treffern [Export als: TSV] (1626 insgesamt)

⏪ ← 1 2 3 4 5 → ⏩

1:	B	1999	moers_kaept	Es war eine doppelt glückliche Fügung, ausgerechnet von den Zwergpiraten gefunden zu werden, denn jeder ...	☰
2:	B	1999	moers_kaept	...war auch das ein Grund, warum Professor Nachtigaller ausgerechnet diesen Ort für seine Nachtakademie ausgewählt hatte.	☰
3:	B	1999	moers_kaept	Leider waren ausgerechnet die in denkbar schlechter Verfassung, vom ewigen Sitze...	☰
4:	B	1999	moers_kaept	Unglaublich, wie vernagelt man ausgerechnet in den größten Krisensituationen sein kann.	☰
5:	B	1999	moers_kaept	Man hat ausgerechnet , daß selbst ein Eydeet ein komplettes Leben benötigt, ...	☰
6:	B	1999	moers_kaept	»Und wieso ausgerechnet ich?«	☰

Beispiel-Suchanfrage: *ausgerechnet*, früheres Interface (eins.dwds.de)

Suchanfrage

Wörterbuch

Wortprofil:
Kombinationen

ZEIT-Korpus

The screenshot shows the DWDS website interface with the search term "ausgerechnet". The top navigation bar includes the DWDS logo and links for "Hilfe zur Suche", "Impressum", and "zur alten DWDS". The search bar contains "ausgerechnet" and the results are displayed in several panels:

- DWDS-Wörterbuch:** Shows "ausgerechnet" as a part. Adv. with pronunciation and usage information. It lists relationships to other words, such as "selbständige flektierte Form zu ausrechnen".
- Etymologisches Wörterbuch (nach Pfeifer):** Indicates "Kein Eintrag vorhanden".
- OpenThesaurus:** Lists synonym groups for "ausgerechnet", including "ausgerechnet, unbedingt".
- Wortprofil 2012 für ausgerechnet:** Shows a word profile with filters for "Adverb", "Dice", and "MI". It lists combinations like "ausuchte droht erwischte feiert gelang geriet erhalten passiert scheiterte stattfinden trifft verdanken verliebt vorwirft wähite".
- Kernkorpus 20:** Shows search results for "ausgerechnet" in the core corpus, with 975 hits. The first few results are listed.
- ZEIT & ZEIT online:** Shows search results for "ausgerechnet" in the ZEIT corpus, with 2518 hits.

Etymol.
Wörterbuch

Thesaurus

Kern-
korpus

Ansicht des Kernkorpus alleine, vergrößert:

Darstellung KWIC (Key Word in Context)

kritika@rz.hu-berlin.de | Abmelden | Profil Hilfe zur Suche | Im

D W D S Ressourcen ▾ Erschließung ▾ Pn

ausgerechnet DWDS Standardsicht ▾ +Ressourcen ▾

Kernkorpus 20 📄 🗑️ ? ✕

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 975

1	1999	BE	...en Sie bitte gar nicht erst, sich diesen Raum vorzustellen! Man hat ausgerechnet , daß selbst ein Eydeet ein komplettes Leben benötigt, um sich auch n
2	1999	BE	...Ich brauchte nur zu spucken! Unglaublich, wie vernagelt man ausgerechnet in den größten Krisensituationen sein kann. Also sammelte ich Spuck...
3	1999	BE	...n diesem Stollen gefangen. Daran kannst du dir ungefähr die Chancen ausrechnen , die du hast. Kähähä!...
4	1999	BE	...daß mir meine Beine zunächst am meisten helfen würden. Leider waren ausgerechnet die in denkbar schlechter Verfassung, vom ewigen Sitzen, der mangeln
5	1999	BE	... ertappt, als er nackt, im Arm einen Jungvolkpimpf - Päule Schabrowski ausgerechnet !, im Farnkraut versteckt lag, hatte ihn in eine der berühmigten ...
6	1999	BE	...cheln um seinen Winzmund gezeigt -, Jochen unterbrechend, sagte: » Daß ausgerechnet du das sagst, der Finanzminister der Familie, der das auf Geld und Wa
7	1999	BE	...o tutte le stanze non era degno di essere abitato. Warum nun Jochen ausgerechnet diese Kinderkrempe labstellkammer aufgesucht, eins der Laufstälchen
8	1999	BE	...erschönere die Welt. Bin schließlich berufshalber damit beschäftigt auszurechnen , wieviel Liter Wasser in Tempelhof und Neuköln jeden Tag durch die F
9	1999	BE	...Trash-Meile gibt's ein paar neue Läden, den » Billigpreis «, warum der ausgerechnet neben » Conny's Container « und schräg gegenüber von » Rudis Reste
10	1999	BE	...Sachen auf und trägt sie zum Müllkübel. Irgendwelche Leute scheinen ausgerechnet die Wiese hinter Gerds Grab auserkoren zu haben, um sich hier nachts




Treffer pro Seite:





975 von 1635 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Textsorte wort: lemmatisiert,
ausgerechnet, ausrechnen, auszurechnen...


Darstellung Zufallsauswahl:

kann unter „Darstellung“ eingestellt werden;
hier auch weitere Suchoptionen, z.B. nach Länge der Sätze

 DWDS Standardsicht ▾ +Ressourcen ▾  

Kernkorpus 20    

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 975 

1	1999	GE	...Ausschaltung der innerparteilichen Reformbewegung geführt hätten, wenn ausgerechnet hier die von der Zentrale vorgegebene Zielstellung nicht erreicht word..
2	1987	ZE	...lisierung. Der Abgeordnete Shevah WEISS erklärte, wenn HERZOG schon ausgerechnet in die Bundesrepublik reisen wolle, dann solle er eine Rede im ehemali
3	1977	BE	...hten Joli-Text meldeten die Druckergesellen eigene Forderungen an, und ausgerechnet ging es auch um Brot und Fleisch. Sicher, sie hungerten nicht, aber..
4	1964	BE	...ubikmeter Beton sind dann zuviel, hat er gesagt. Und er hat's schon ausgerechnet . Sechzig Minuten, nicht mehr..." Balla rüttelte ihn.
5	1958	BE	...sogar. In einer Woche von Remagen bis Frankfurt, dann kannst du dir ausrechnen , wann sie hier sein werden. Mal herhören, was ich kombiniere.
6	1944	GE	...9., als Fulda dran war, auch Gießen zwei Bomben abgekliegt hat. Und ausgerechnet 100 Meter von unserem Haus entfernt ist eine gefallen, und die andere
7	1937	BE	... in den die Straße mündete, beobachteten es staunende Gruppen. Kaum auszurechnen ist es «, wiederholte Gundling nochmals, » es ist so lange her, daß au..
8	1926	BE	...Der Schiffsbauer war ein guter Rechner gewesen. Er hatte ausgerechnet , daß auf einem Schiff immer ein Drittel oder manchmal gar die Hälfte .
9	1916	ZE	... Potsch bist du! die gebratenen Tauben werden dir ins Maul fliegen, ausgerechnet - ich stoß und du kommst nicht vom Fleck - möchten möchtest du viel
10	1908	BE	...chwätzig fort," aber die Männer denken immer, alles geht so, wie sie's ausrechnen . Der Flur des Feuerbachschen Hauses war angefüllt von Menschen.

Treffer pro Seite:

975 von 1635 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Suchfilter eingeschränkt auf wissenschaftliche Texte

Kernkorpus 20 f e ? x

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 72 🔍 1 2 3 4 >>|

1	1994 WI	...nrichtung einsichtsvoll bejahen, aber tausend Gründe dafür finden, daß ausgerechnet hier der dafür denkbar ungünstigste Platz wäre. Diese Wendung wird ...
2	1983 WI	...Menschen sich zumuten lassen? Es gibt keine Mathematik, mit der man ausrechnen kann, wie lange der Geduldsfaden hält, und es gibt auch keine Arithmet...
3	1971 WI	...e Beatles das Universalgenie und dessen vergeblich objektivierete Werke ausgerechnet vermittelt der Schallplatte, des Symbols von Kunst als saisonbedingter.
4	1961 WI	...b, den man natürlich für alle Felder aufgrund der Rohwertformel vorher ausgerechnet haben muß. b) Bei standardisierten Testen wird heutzutage allgemein...
5	1952 WI	...issenschaft und der Technik, wo der Fortschritt notwendig ist und sich ausrechnen lasse. In der Philosophie gibt es keinen 'Fortschritt', deshalb auc...
6	1941 WI	... gemessen und die resultierende Druckkraft in Strömungsrichtung danach ausgerechnet wird. Für die Kugel und für den Zylinder werden auf S. 158 diese Dr...
7	1932 WI	...hst merkwürdiges Versehen ist dem Verf. mehrmals unterlaufen, indem er ausgerechnet Bibl zitiert, wenn er, wie sich aus dem Inhalt und den angegebenen Sei...
8	1903 WI	...ichtung der Erkenntnis vollständig ausrechenbar, wenn auch bisher nicht ausgerechnet sei, also kann er vom Gegenstand reden, als hätten wir ihn, und kann s...
9	1927 WI	...elegt. Die Entfertheiten sind zunächst und auch da, wo » amtlich « ausgerechnete Maße bekannt sind, umsichtig geschätzt. Weil das Ent-fernte in solc...
10	1910 WI	...ckt im Darwinismus nur eine Fiktion. Nun aber: » Ich kann den Kreis ausrechnen , wenn ich ihn so betrachte, als ob er ein regelmäßiges Unendlicheck w...

Treffer pro Seite: 10 20 50 100 500

72 von 74 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Anklicken eines Schlüsselworts liefert die Satzumgebung (3 Sätze)

Kernkorpus 20

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 72

1 2 3 4 >>|

1 1994 WI ...nrichtung einsichtsvoll bejahen, aber tausend Gründe dafür finden, daß **ausgerechnet** hier der dafür denkbar ungünstigste Platz wäre. Diese Wendung wird ...

2 1983 WI ...Menschen sich zumuten lassen? Es gibt keine Mathematik, mit der man **ausrechnen** kann, wie lange der Geduldsfaden hält, und es gibt auch keine Arithmet...

3 1971 WI ...e Beatles das Universalgenie und dessen vergeblich objektivierte Werke **ausgerechnet** vermittels der Schallplatte, des Symbols von Kunst als saisonbed...

Werckmeister, Otto Karl, Das gelbe Unterseeboot und der
eindimensionale Mensch, in: ders., Ende der Ästhetik,
Frankfurt a.M.: S. Fischer 1971, S. 114

Die zynisch gemeinte Abwertung der eigenen Arbeit entwertet zugleich die große Tradition, auf die sie sich beruft. Daher verhöhnen im Film die Beatles das Universalgenie und dessen vergeblich objektivierte Werke **ausgerechnet** vermittels der Schallplatte, des Symbols von Kunst als saisonbedingter Mode. Wenn Lennon den Sinn vergangener und gegenwärtiger hoher und populärer Kunst unterschiedslos leugnet, benennt er, ohne es gewahr zu werden, die gegenwärtige Krise der künstlerischen Bedeutung, die aus einem historischen Verfallsprozeß resultiert.

4 1961 WI ...b, den man natürlich für alle Felder aufgrund der Rohwertformel vorher **ausgerechnet** haben muß. b) Bei standardisierten Testen wird heutzutage allgemei...

5 1952 WI ...issenschaft und der Technik, wo der Fortschritt notwendig ist und sich **ausrechnen** lasse. In der Philosophie gibt es keinen 'Fortschritt', deshalb auc...

6 1941 WI ... gemessen und die resultierende Druckkraft in Strömungsrichtung danach **ausgerechnet** wird. Für die Kugel und für den Zylinder werden auf S. 158 diese Dr...

7 1932 WI ...hst merkwürdiges Versehen ist dem Verf. mehrmals unterlaufen, indem er **ausgerechnet** Bibl zitiert, wenn er, wie sich aus dem Inhalt und den angegebenen Sei...

8 1903 WI ...ichtung der Erkenntnis vollständig ausrechenbar, wenn auch bisher nicht **ausgerechnet** sei, also kann er vom Gegenstand reden, als hätten wir ihn, und kann s...

9 1927 WI ...elegt. Die Entferntheiten sind zunächst und auch da, wo » amtlich « **ausgerechnete** Maße bekannt sind, umsichtig geschätzt. Weil das Ent-fernte in solc...

10 1910 WI ...ckt im Darwinismus nur eine Fiktion. Nun aber: » Ich kann den Kreis **ausrechnen** , wenn ich ihn so betrachte, als ob er ein regelmäßiges Unendlicheck w...



Treffer pro Seite: 10 20 50 100 500




72 von 74 Treffern anzeigbar

DDC-Query | Darstellung | Suchfilter






Heft-
klammer:
Speicher-
ung von
Belegen

Genauere Wortsuche (nicht lemmatisiert): @ausgerechnet




@ausgerechnet  DWDS Standardsicht  +Ressourcen   

Kernkorpus 20    

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 815     

1	1999	BE	... sagte 16 U. BA-RUMMS! Mir schien es egoistisch, das Bollogg-Gehirn ausgerechnet in diesem wichtigen historischen Augenblick zu verlassen, aber wir hat
2	1999	BE	... nach außen stehenden und empfindlichen Ohren. Schwer zu sagen, was ausgerechnet Nattiftoffen zu so ausgezeichneten Politikern machte, vielleicht war ...
3	1999	GE	...s Bildungssystems für alle - unabhängig von der sozialen Herkunft - es ausgerechnet die Schulen waren, die wieder neue Unterschiede schufen: Diese ware
4	1999	GE	...des geozentrischen Weltbildes (alles dreht sich um die Erde), hatte er ausgerechnet , daß die Bewegungen der Planeten sich besser erklären ließen, wenn
5	1999	BE	... ertappt, als er nackt, im Arm einen Jungvolkpimpf - Päule Schabrowski ausgerechnet !, im Farnkraut versteckt lag, hatte ihn in eine der berühmigten ...
6	1999	BE	...cheln um seinen Winzmund gezeigt -, Jochen unterbrechend, sagte: » Daß ausgerechnet du das sagst, der Finanzminister der Familie, der das auf Geld und Wa
7	1999	BE	...o tutte le stanze non era degno di essere abitato. Warum nun Jochen ausgerechnet diese Kinderkrempe labstellkammer aufgesucht, eins der Laufstälchen
8	1999	BE	...Trash-Meile gibt's ein paar neue Läden, den » Billigpreis«, warum der ausgerechnet neben » Conny's Container« und schräg gegenüber von » Rudis Reste
9	1999	BE	...Sachen auf und trägt sie zum Müllkübel. Irgendwelche Leute scheinen ausgerechnet die Wiese hinter Gerds Grab auserkoren zu haben, um sich hier nachts
10	1999	BE	...h. Astrologin war meine Oma; sie starb in dem Monat, den sie vorher ausgerechnet hat. Sie hat mir ihre Bücher gegeben, kurz vorm Ende, und gesagt, i...

Treffer pro Seite:     

815 von 1389 Treffern anzeigbar DDC-Query | Darstellung | Suchfilter

Suchanfrage (unter „Hilfe zur Suche“)

Suchanfrage	Effekt
Arzt	<i>Arzt, Arztes, Ärzte ...</i> (flektierte Formen von Arzt)
@Arzt	<i>Arzt</i> (nur die Wortform <i>Arzt</i>)
Arzt*	<i>Arzt, Arztbesuch, Arztberuf, ...</i>
*arzt	<i>Sportarzt, Hausarzt, ...</i>
"gute Arzt"	<i>guter Arzt, bester Arzt, gute Ärzte, ...</i>
"das gute Beispiel"	<i>das gute Beispiel, die besseren Beispiele ...</i>
"Kanzler #1 Schröder"	<i>Kanzler Schröder, Kanzler Gerhard Schröder</i> (Abst. 1 Wort)
Kanzler && Schröder	Sätze, in denen <i>Kanzler</i> und <i>Schröder</i> vorkommen
Kanzler Schröder	Sätze, in denen <i>Kanzler</i> oder <i>Schröder</i> vorkommt
Kanzler && !Schröder	Alle Sätze, mit <i>Kanzler</i> , aber ohne <i>Schröder</i>
\$p=NE with Herzog	<i>Roman Herzog, Peter Herzog, ...</i> (Eigennamen)
"Ägide #2 \$p=NE"	<i>Ägide Bush, Ägide von Harald Szeemann</i>
\$p=NN with *zeit	<i>Weihnachtszeit, Übergangszeit, Halbzeit,</i>
"üben #5 aus with \$p=PTKVZ"	<i>... übt er ein Wahlamt aus ...</i> (Präfixverben)

"schälen with \$p=VFIN #5 aus with \$p=PTKVZ"	<i>schälen</i> als Verb gefolgt von einem trennbaren Präfix <i>aus</i>
"sein with \$p=VFIN #20 \$p=VPP #0 @worden":	Auxiliar <i>sein</i> gefolgt von einem Partizip und <i>worden</i>

Verfügbare Operatoren:

@	Wörter (keine Lemmasuche)
*	Beliebig viele Zeichen
?	Ein Zeichen
"..."	Suche nach Wortfolge
#n	Abstand: max. n Wörter
&&	und
	oder
!	nicht
\$p=X	Variable für "part of speech"; X: Tag (Annotation)
\$p=X with Y	Part of speech X enthält Y

Annotationskategorien (“Tag-Set”)

ADJA	attributives Adjektiv
ADJD	adverbiales oder prädikatives Adj.
ADV	Adverb
APPR	Präposition; Zirkumposition links
APPRART	Präposition mit Artikel
APPO	Postposition
APZR	Zirkumposition rechts
ART	Artikel
CARD	Kardinalzahl (Ordinalzahlen: ADJA)
FM	Fremdsprachliches Material
ITJ	Interjektion
KOUI	unterord. Konjunktion + „zu“ u. Inf.
KOUS	Konjunktion mit Satz
KON	nebenordnende Konjunktion
KOKOM	Vergleichskonjunktion
NN	normales Nomen
NE	Eigennamen
PDS	substituierendes Demonstrativpronomen
PDAT	attrbuierend. Demonstrativpronomen

[das] große [Haus]
[er fährt] schnell, [er ist] schnell
schon, bald, doch
in [der Stadt], ohne [mich]
im [Haus], zur [Sache]
[ihm] zufolge, [der Sache] wegen
[von jetzt] an
der, die, das, eine, ein, ...
zwei [Männer], [im Jahre] 1994
[Er hat ``] big [“ gesagt]
mhm, ach, tja
um [zu leben], anstatt [zu fragen]
weil, daß, damit, wenn, ob
und, oder, aber
als, wie
Tisch, Herr, [das] Reisen
Hans, Hamburg, HSV
dieser, jener
jener [Mensch]

PIS	substituierendes Indefinitpronomen	<i>keiner, viele, man, niemand</i>
PIAT	attribuierend. Indef.pron. o. Det.	<i>kein [Mensch], irgendein [Glas]</i>
PIDAT	attribuierend. Indef.pron. mit Det.	<i>[ein] wenig [Ei], [die] beiden [Eier]</i>
PPER	irreflexives Personalpronomen	<i>ich, er, ihm, mich, dir</i>
PPOSS	substituierendes Possessivpronomen	<i>meins, deiner</i>
PPOSAT	attribuierendes Possessivpronomen	<i>mein [Buch], deine [Mutter]</i>
PRELS	substituierendes Relativpronomen	<i>[der Hund ,] der</i>
PRELAT	attribuierendes Relativpronomen	<i>[der Mann ,] dessen [Hund]</i>
PRF	reflexives Personalpronomen	<i>sich, einander, dich, mir</i>
PWS	substituierendes Interrogativpron.	<i>wer, was</i>
PWAT	attribuierendes Interrogativpron.	<i>welche [Farbe], wessen [Hut]</i>
PWAV	adverbiales Inter. od. Rel.Pron.	<i>warum, wo, wann, worüber, wobei</i>
PAV	Pronominaladverb	<i>dafür, dabei, deswegen, trotzdem</i>
PTKZU	``zu" vor Infinitiv	<i>zu [gehen]</i>
PTKNEG	Negationspartikel	<i>nicht</i>
PTKVZ	abgetrennter Verbzusatz	<i>[er kommt] an, [er fährt] rad</i>
PTKANT	Antwortpartikel	<i>ja, nein, danke, bitte</i>
PTKA	Partikel bei Adjektiv oder Adverb	<i>am [schönsten], zu [schnell]</i>
TRUNC	Kompositions-Erstglied	<i>An- [und Abreise]</i>

VVFIN	finites Verb, voll	<i>[du] gehst, [wir] kommen [an]</i>
VVIMP	Imperativ, voll	<i>komm [!]</i>
VVINFINF	Infinitiv, voll	<i>gehen, ankommen</i>
VVIZU	Infinitiv mit ``zu'', voll	<i>anzukommen, loszulassen</i>
VVPP	Partizip Perfekt, voll	<i>gegangen, angekommen</i>
VAFIN	finites Verb, aux	<i>[du] bist, [wir] werden</i>
VAIMP	Imperativ, aux	<i>sei [ruhig !]</i>
VAINFINF	Infinitiv, aux	<i>werden, sein</i>
VAPP	Partizip Perfekt, aux	<i>gewesen</i>
VMFIN	finites Verb, modal	<i>dürfen</i>
VMINFINF	Infinitiv, modal	<i>wollen</i>
VMPP	Partizip Perfekt, modal	<i>gekonnt, [er hat gehen] können</i>
XY	Nichtwort, Sonderzeichen enthaltend	<i>3:7, H2O, D2XW3</i>
\$,	Komma	<i>,</i>
\$.	Satzbeendende Interpunktion	<i>. ? ! ; :</i>
\$(sonstige Satzzeichen; satzintern	<i>- [.]()</i>

Beispiel: Suche nach Partikel (Adverb) und Partizip Perfekt *ausgerechnet*

@ausgerechnet with \$p=ADV

Kernkorpus 20

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 779

1 1999 ZE ...er Übermut eines Regisseurs, um das verwegene Bild vom fliegenden Dach **ausgerechnet** in Tadschikistan Gestalt annehmen zu lassen. Luna Papa wurde in Ver...

2 1985 GE ...terngeneration. Dieses Ziel ist, so haben B.-Statistiker in der DDR **ausgerechnet**, nur dann zu erreichen, wenn von je 1000 Frauen im Laufe ihres Lebens...

3 1970 BE ...n spricht über die Statue der Freiheit, die Frédéric-Auguste Bartholdi **ausgerechnet** in Gestalt seiner lieben Mutter im Hafen von New York ausgestellt hat...

4 1965 BE ...Hauptabteilungsleiter, wie er nicht im Buche steht, dachte Robert, und **ausgerechnet** Jakob Filter. Ein selbstbewußter Mann ohne falsche Jovialität, ohne...

5 1957 ZE ...be. Diese Situation werde nicht dadurch überdeckt, daß man die NATO **ausgerechnet** in Bonn tagen lassen oder wieder einmal auf die europäischen Pläne...

Bemerke: Fehlerhafte Klassifikation, durch automatisches Tagging

@ausgerechnet with \$p=VVPP

Kernkorpus 20

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

Treffer: 35

1 1901 GE ...auswärts vergrößerten Haushalt mitsorgen kann. Da man sich ungefähr **ausgerechnet** haben wird, wie teuer das Hochzeitsmahl mit den Weinen sein darf, so w...

2 1970 BE ...n habe ich das Sterben versucht. Der Traum wußte den Tag des Todes, **ausgerechnet** oder vorausgesagt, im voraus. Richtig hatte ich Schwierigkeiten bei...

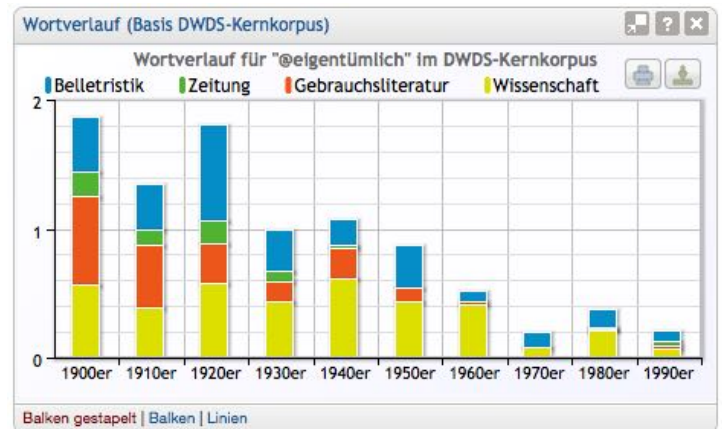
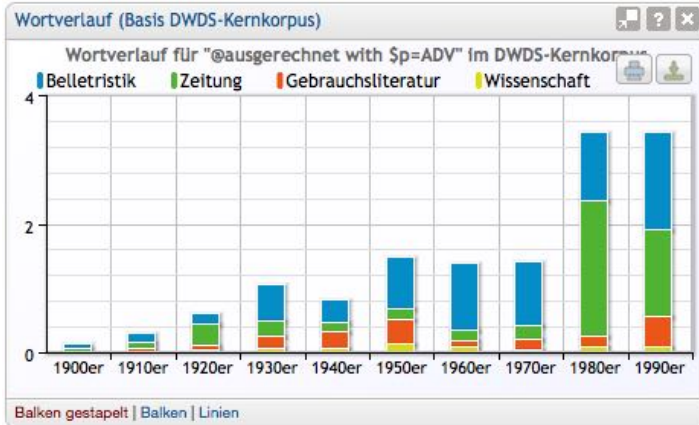
3 1911 BE ...icht war glatt und licht, wie ein Sommertag Ich habe mir die Sache gut **ausgerechnet**«, sagte Doktor Emmerich. Ein reicher Patient - das gibt, bei Wegle...

4 1980 BE ... los bzw. will losfahren. Hoffentlich reicht die Zeit, die Arne mir **ausgerechnet** hat. Vielleicht hätte sie gereicht, wenn mir nicht der Altonaer Bah...

5 1932 ZE ...n allen Regierungen der Welt erlassen worden - das hat ein Statistiker **ausgerechnet**. Mehr als lange Kommentare und Reden beweisen solche Zahlen, unter...

Häufigkeiten

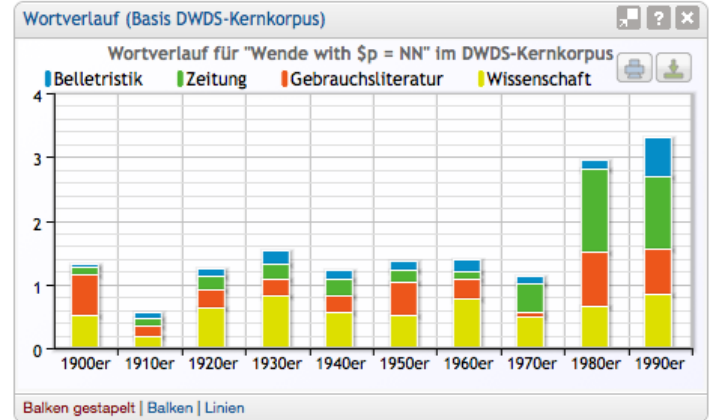
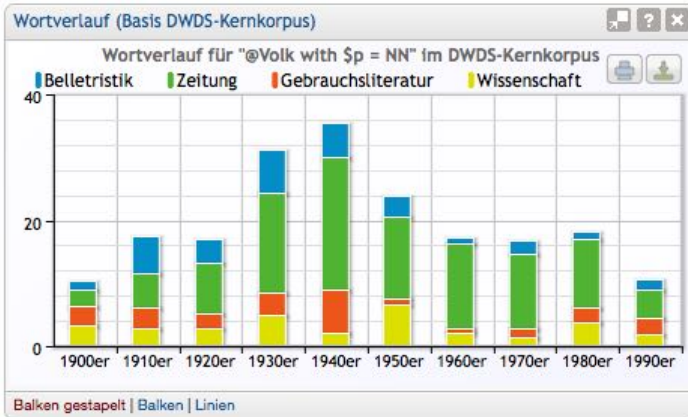
Anzeige von Trefferzahlen nach Jahrzehnten und nach Textsorten über DWDS-Statistiken; („Wortverlauf“), hier ein Vergleich von *ausgerechnet* (ADV) und *eigentlich*:



Weitere statistische Daten,
zu *ausgerechnet*:

Korpus	Hits	Hits [ppm]	Korpusgröße [Mill. Token]
Berliner Tagesspiegel	6653	40.2754	165.19
Berliner Zeitung	9054	35.7348	253.37
C4-Korpus	371	4.6375	80.00
Compact Memory Corpus	29	1.1049	26.25
DDR-Korpus	116	13.4001	8.66
Die ZEIT & ZEIT Online	22634	49.2043	460.00
DWDS-Kernkorpus	779	7.7435	100.60
DWDS-Korpus21	111	59.4017	1.87
Julland-Korpus	4	8.0000	0.50

Volk und Wende, Wortverlauf und -profil



Information über häufige Wortkombinationen (Kollokationen) zu *Volk*, z.B. mit dem Maß log-DICE:



Klicken auf *bedroht*:



Lemmbasierte Kollokationssuche im DWDS-Kerncorpus

Suchbegriff:

Optionen

Statistisches Maß:

n-Gramme: Folgen von n Wörtern
 Bigramme: Folgen von 2 Wörtern

Verschiedene Maße
 für Bigramme

Ihre Suche ergab 4554 Teffer. Das Wort *Volk* kommt 45700 mal im Korpus vor.

#	w1	F(w1)	w2	F(w2)	F(w1,w2)	MI	T-Score	Log-L.	Belege
1	Volk	45700	deutsch	117415	5907	7.1129	76.3018	47617.7734	Suche
2	Volk	45700	unsere	124575	3014	6.0567	54.0752	19645.9316	Suche
3	Volk	45700	ganz	161552	1829	4.9611	41.3939	9133.3594	Suche
4	Volk	45700	Volk	45700	1234	6.2152	34.6555	8264.6064	Suche
5	Volk	45700	Land	69671	1076	5.4091	32.0305	6009.3716	Suche
6	Volk	45700	Recht	38596	1068	6.2505	32.2510	7201.2100	Suche
7	Volk	45700	Staat	58719	1044	5.6123	31.6505	6119.9805	Suche
8	Volk	45700	groß	171120	1032	4.0525	30.1887	3887.7275	Suche

Weitere Beispiele für Abfragen

Beispiel: Diminutiva

Nomina auf *-lein*: &p = NN with *lein

Nomina auf *-chen*: &p = NN with *chen; leider hier viele ungewollte Formen

Nomina auf *-i* ohne *-ei*: \$p = NN with *i && !\$p = NN with *ei

Beispiel: Syntax

Syntaktisches Muster Verbkomplex: "haben \$p=VVINF @können"

Syntaktisches Muster: Präfixverben, *ausüben* "üben #5 aus with \$p=PTKVZ"

Das Deutsche Referenzkorpus (IDS)

Siehe Folien von Stefan Engelberg (Moodle-Seite),
wesentlich größeres, aber weniger ausgewogenes Korpus (IDS Mannheim, Cosmas)
bessere Abfragesprache, feinere Wortverläufe



Das Internet als Korpus?

Der Vorteil:

- Sehr viele Texte verfügbar. Auf diese Weise können auch sehr seltene Phänomene der Schriftsprache entdeckt werden.

Nachteile:

- Kein allgemeines Prinzip der Textzusammenstellung.
- Texte von sehr unterschiedlicher Qualität.
Mögliche Abhilfe: Einschränkung auf Google Print (gedruckte Texte)
- Texte möglicherweise von Nicht-Muttersprachlern (vor allem englische Texte).
Abhilfe: Sehen Sie sich den Kontext von Belegstellen an!
- Suchabfragen gehen über das gesamte Internet, d.h. auch fremdsprachliche Texte werden erfasst.
Abhilfe: Einschränkung auf bestimmte Domänen, z.B. auf “.de”.
Aber auch dabei werden fremdsprachliche Texte erfasst.
- Viele Texte treten mehrfach auf, daher ist die Angabe der Trefferhäufigkeit nicht zuverlässig.

Suche mit Google

- Domänen-Einstellung, z.B. “.de” für deutsche Webseiten
- Suche nach Dateitypen möglich, z.B. “.pdf”
- Suche nach Ketten von Wörtern: “sogar nur“
- Joker (wildcards): “sogar * nur“, “sogar * * nur”;
- Google lemmatisiert, z.B. findet *Auto* auch Suchstellen mit *Autos*.

Suche mit WebCorp

Eine Suchschnittstelle für linguistische Zwecke: <http://www.webcorp.org.uk/>

WebCorp erlaubt verschiedene Suchmethoden auch unterhalb der Wortebene:

- Joker (Wildcards) für Wortteile “run*” für *running, runner* etc; “*ing” für “running” etc.
- Joker für Phrasen: “* sank” für *the boat sank, the ferry sank* (genauer: “the ... sank”, wobei ... für ein bis zwei Wörter steht).
- Mehrfache Joker
- Alternativen in eckigen Klammern: “[ship|boat] sank” , “[ship|boat] s[a|u]nk”;
“r[u|a]n[n|ning|s]” für *running, runs, run, ranning, rans, ran*

Sie können ferner:

- das Output-Format steuern (verschiedene Formen von Konkordanzen)
- die Domäne bestimmen (Sprachen).
- angeben, welche Wörter vorkommen müssen / nicht vorkommen dürfen (+/-).

Es gibt einen Zugang mit zusätzlichen Optionen: <http://wse1.webcorp.org.uk/home/>

WebCorp Live lets you access to the Web as a corpus - a large collection of texts from which examples

Search:



Case Insensitive:

Span:

50 characters



Search API:

Bing

Language:

Not specified



Advanced Options

Zurücksetzen

Search

Google Books als Korpus

Eine Alternative zur Verwendung des Internets als Korpus:
Buchpublikationen, wie sie z.B. durch GoogleBooks verfügbar sind.

Für deutschsprachige Publikationen siehe <http://books.google.com/books?hl=de>

Dieser Korpus ist sehr groß – für englischsprachige Publikationen wird geschätzt, dass er ca. 4% der jemals veröffentlichten englischsprachigen Literatur umfasst.

Auswahlkriterien: Schwerpunkt auf urheberrechtsfreier Literatur,
bzw. Literatur, deren Urheberrechtsstatus ungeklärt ist.

Zu den Möglichkeiten solcher massiver Korpora

vgl. Michel e.a. „Quantitative analysis of culture using millions of digitized books“, in *Science* 2010.

<http://www.sciencemag.org/content/early/2010/12/15/science.1199644>

Für quantitative Daten zu Worthäufigkeiten auf Grundlage von Google Books:

Google Ngram Viewer, <http://books.google.com/ngrams/>

- Zeitverlauf von 1800 bis heute;
aber: es ist unklar, wie genau dies kontrolliert ist – z.B. Neuauflagen von Werken.
- Angabe von relativen Häufigkeiten von Wörtern in Prozent.
- “Smoothing” – z.B. Smoothing mit Wert 3: es werden Mittelwerte über 3 Jahre gewählt.

Beispiel für kulturhistorische Verwendung von Google NGram: *Vaterland, Muttersprache*

Google books Ngram Viewer

Graph these **case-sensitive** comma-separated phrases: between and from the corpus with smoothing of .

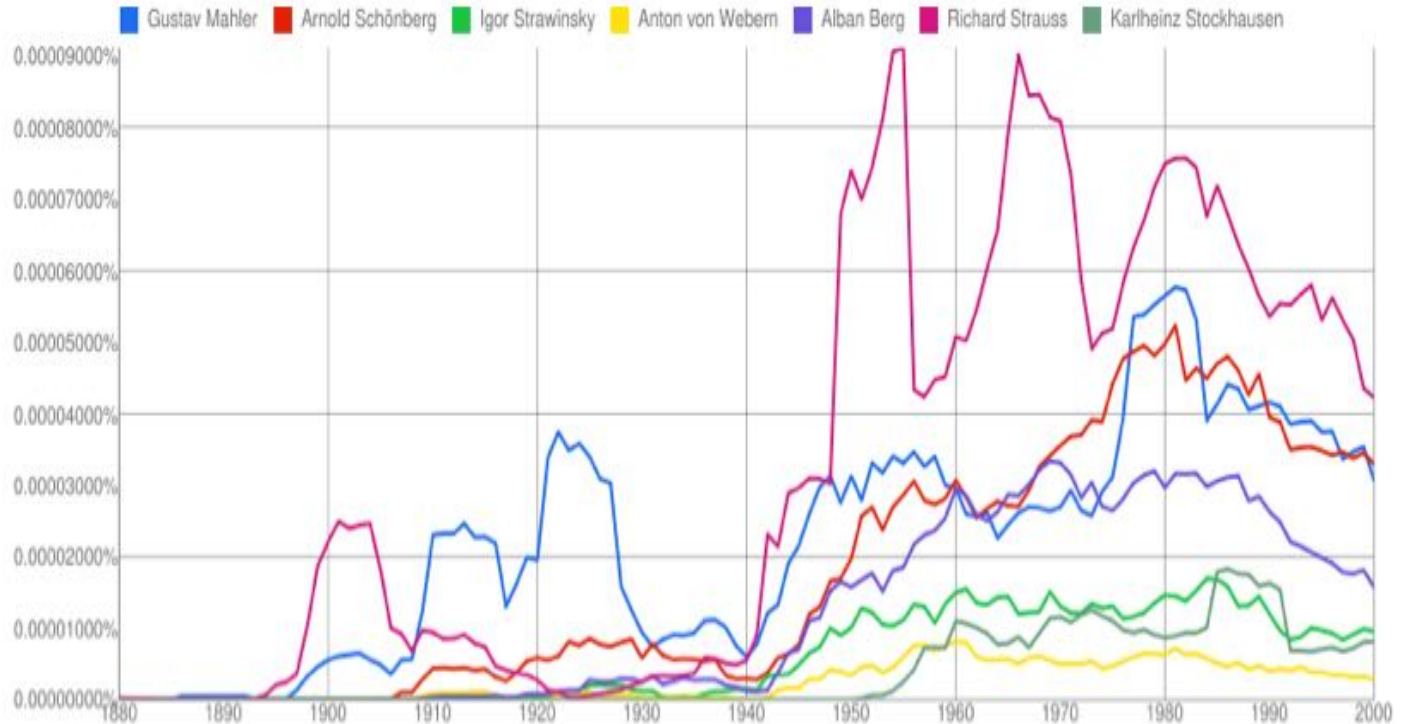
[Search lots of books](#)



Search in Google Books:

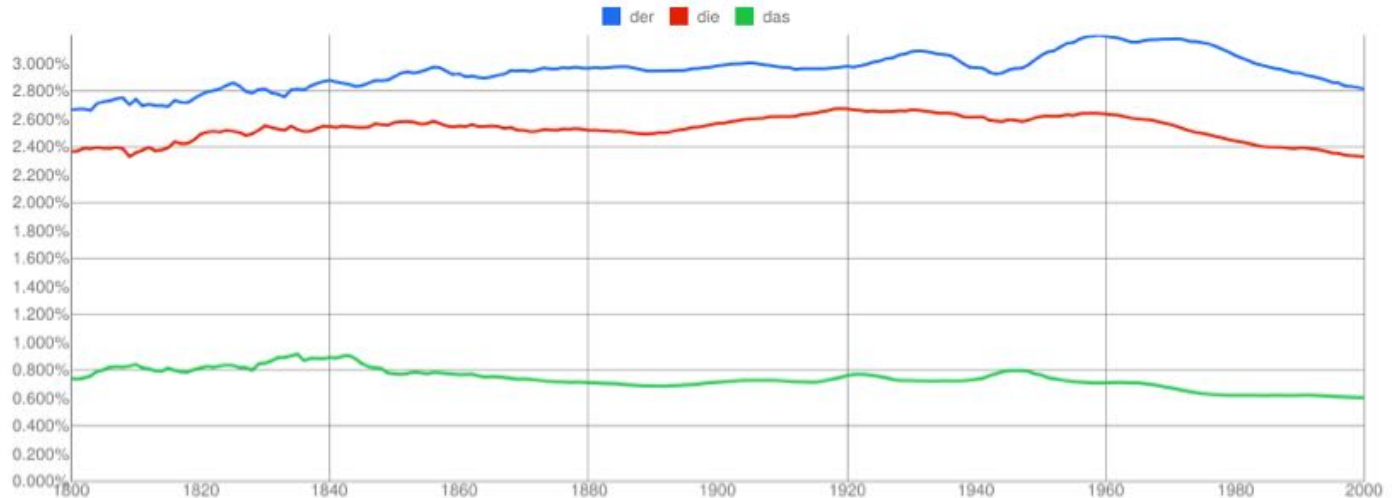
1800 - 1828	1829 - 1958	1959 - 1972	1973 - 1986	1987 - 2000	Muttersprache (German)
1800 - 1811	1812 - 1820	1821 - 1866	1867 - 1970	1971 - 2000	Vaterland (German)

Beispiel für kulturhistorische Verwendung, Google NGram: Komponisten

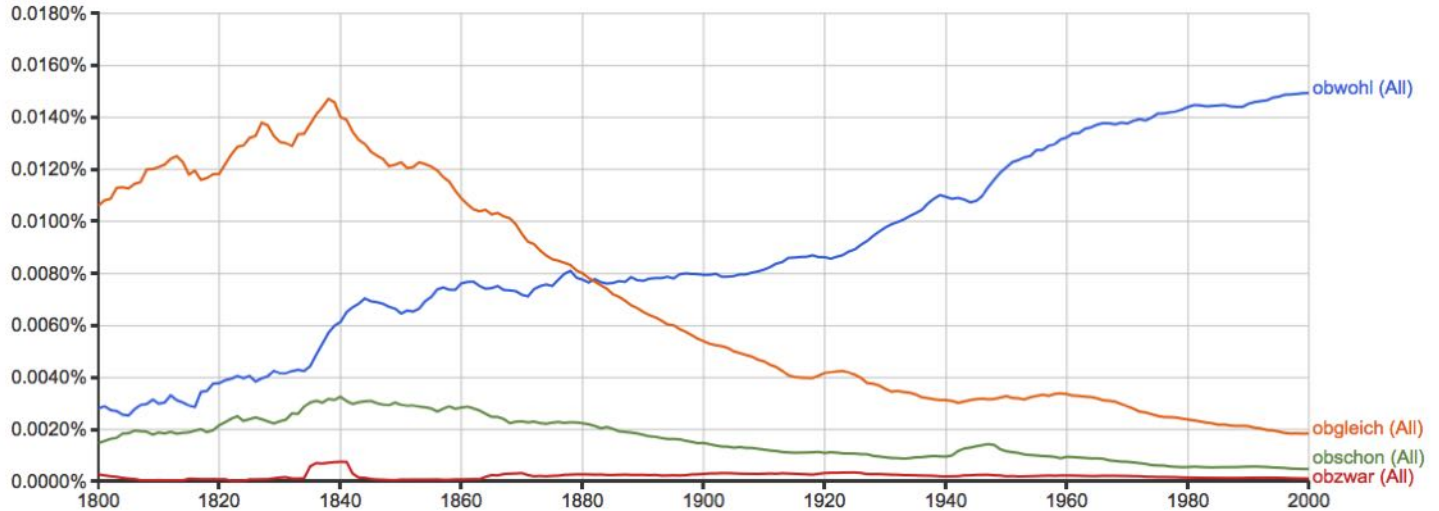


Beispiel für linguistische Verwendung von Google NGram: *der, die, das* – Genus-Häufigkeiten

(man beachte: *die* enthält auch Plural, *die* und *das* auch Akkusativ-Formen, *der* auch fem. Genitiv)



Konjunktionen: *obwohl*, *obgleich*, *obschon*, *obzwar*



Beispielsuche möglich:

[1800 - 1839](#)

[1840 - 1955](#)

[1956 - 1969](#)

[1970 - 1984](#)

[1985 - 2000](#)

[obwohl](#)

German

[1800 - 1837](#)

[1838](#)

[1839 - 1898](#)

[1899 - 1964](#)

[1965 - 2000](#)

[obzwar](#)

German

[1800 - 1817](#)

[1818 - 1841](#)

[1842 - 1851](#)

[1852 - 1947](#)

[1948 - 2000](#)

[obschon](#)

German

[1800 - 1811](#)

[1812 - 1832](#)

[1833 - 1842](#)

[1843 - 1943](#)

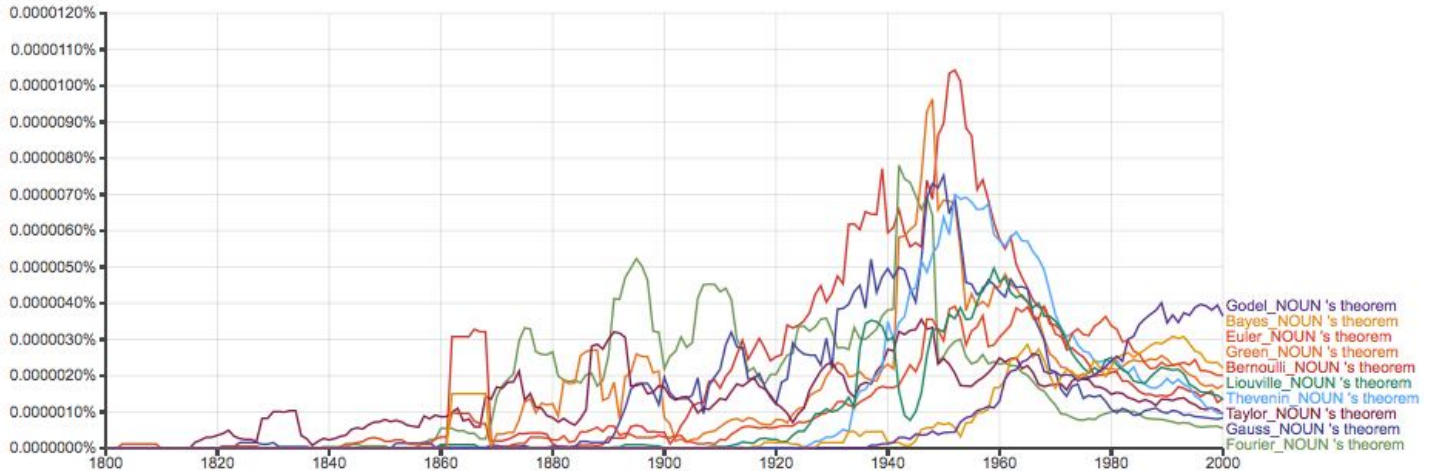
[1944 - 2000](#)

[obgleich](#)

German

Suchmöglichkeiten mit Tagging, Jokers

Tagging nach _, Joker (Wildcards) * für 10 häufigste Werte, z.B: *_NOUN 's theorem



Man kann nun auch die entsprechenden Daten erhalten;
ferner Angabe von:

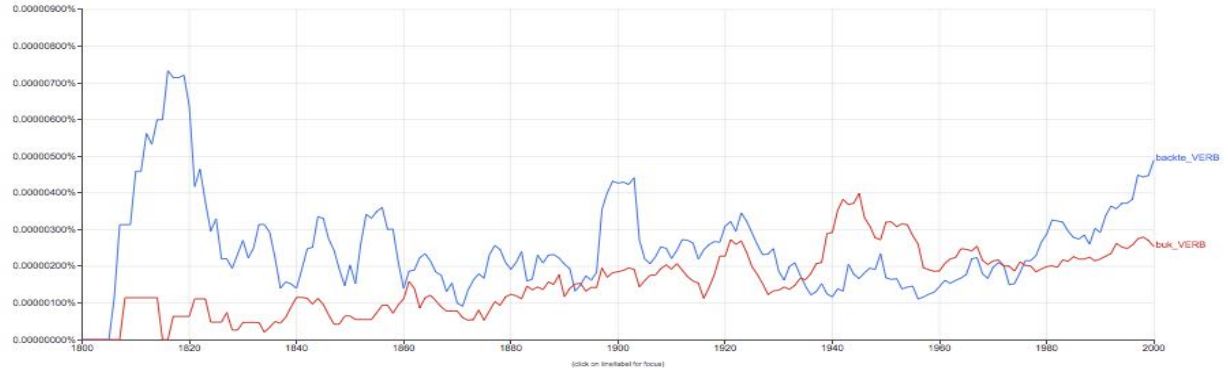
- Wortart (Part of Speech), z.B. help_VERB, help_NOUN
- morphologischer Kategorien, z.B. help_INF (Infinitiv)
- Abhängigkeiten, z.B. drink=>water_NOUN

Info unter: <https://books.google.com/ngrams/info>

NOUN		These tags can either stand alone (_PRON_) or can be appended to a word (she_PRON_)
VERB		
ADJ	adjective	
ADV	adverb	
PRON	pronoun	
DET	determiner or article	
ADP	an adposition: either a preposition or a postposition	
NUM	numeral	
CONJ	conjunction	
PRT	particle	
ROOT	root of the parse tree	These tags must stand alone (e.g., _START_)
START	start of a sentence	
END	end of a sentence	

Beispiel: Suche nach Flexionsformen

backte vs. *buk*

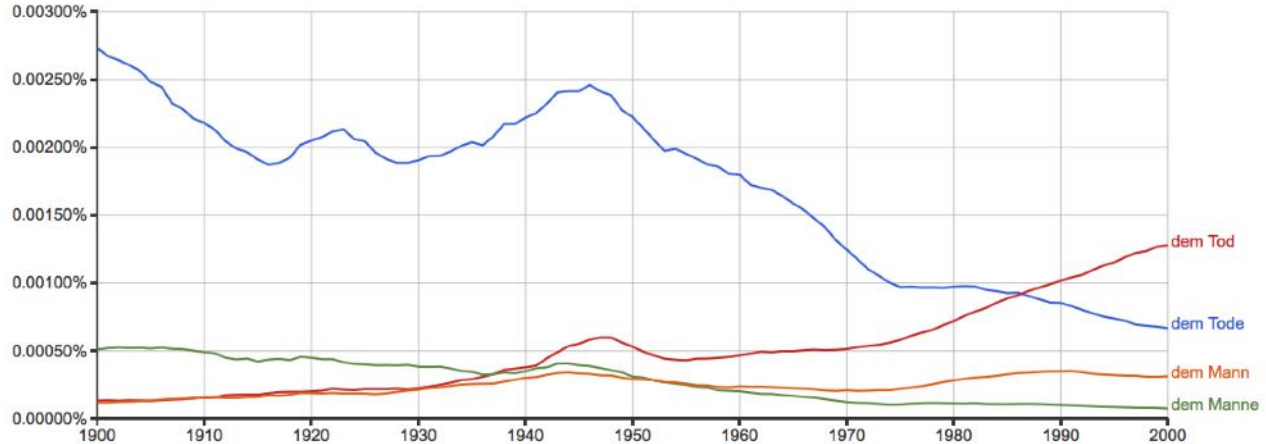


gewinkt vs. *gewunken*

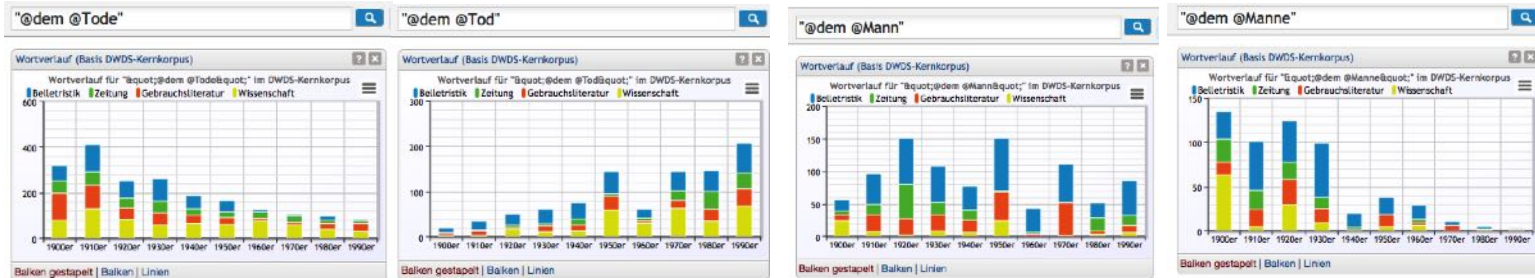


Beispiel Nominalflexion: *dem Tode, dem Tod, dem Manne, dem Mann*

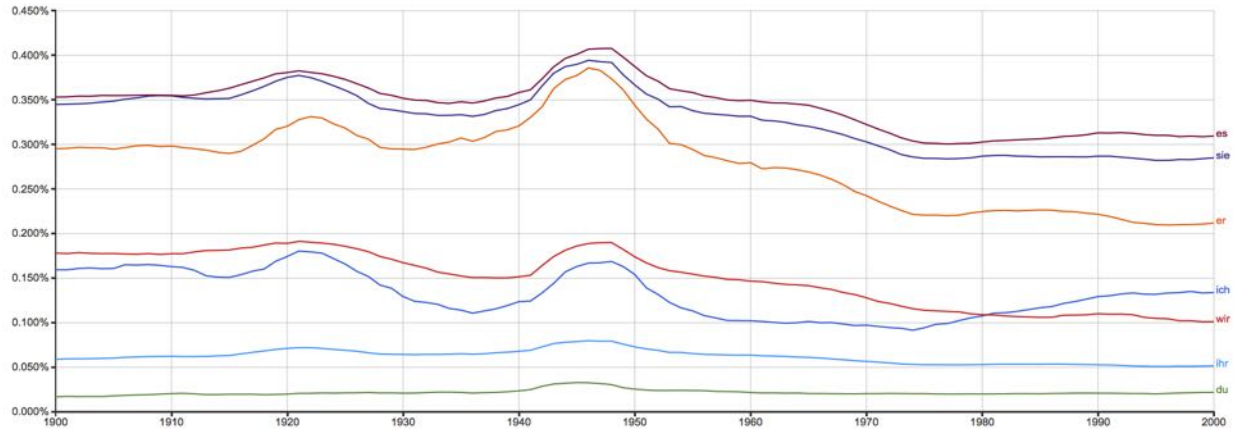
im



Vergleich: DWDS-Korpus



Beispiel: Pronomenhäufigkeit (*du, ihr, wir, ich, er, sie, es*)



Beispiel: *derselbe, dieselbe, dasselbe*



BYU Webkorpus

Google NGram erlaubt keine genauere Steuerung der Abfragesprache; es wird nur nach Wörtern oder Folgen von Wörtern gesucht.

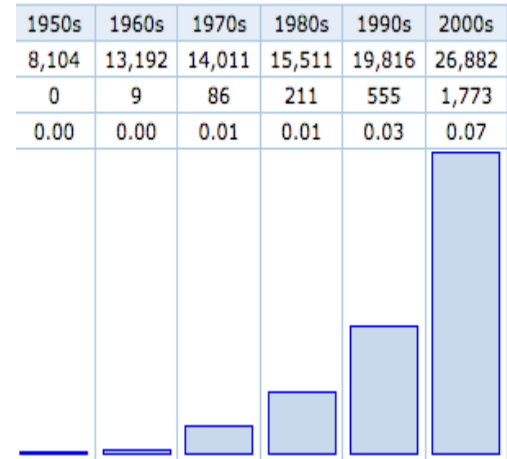
Für amerikanisches Englisch (bald auch für andere Sprachen).

Suchmaschine Brigham Young University: <http://googlebooks.byu.edu/>

- Detaillierte Abfragesprache
 - Wiedergabe von Kollokationen
 - Textvorkommnisse direkt aufrufbar
- Beispiel BYU Webkorpus:

Darstellung der Häufigkeit des Worts

Einfacher Zugang zu den Beispielen



Beispiele für Korpora

Korpus	Zugang	Funktion	Medium	Umfang	Annotation	Verfügbarkeit
Deutsch. Referenz-Korpus DeReKo	http://www.ids-mannheim.de/kl/projekte/korpora/ IDS Mannheim		Schriftl.	>5 Mia.	Textstruktur	online, Verkauf
DWDS-Kernkorpus	www.dwds.de (Berlin-Brandenburgische Akademie der Wissenschaften)	Teil der Textbasis für Digitales Wörterbuch der Deutschen Sprache des 20. Jh.	Mündl., Schriftl.	100 Mio.	Lemma, Morphosyntax	online
Fehlerannot. Lernerkorpus (FALKO)	www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko	Untersuchungen zu Fehlern von DAF-Lernern	Schriftl.	im Aufbau	im Aufbau, flexible Architektur	frei

NEGRA-Korpus	http://www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html	Empirische Basis für ling. Forschung, maschinelles Lernen	Schriftl.	0,36 Mio.	Morphosyntax, Konstituenten,	Lizenz
TIGER-Korpus	www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/	ling. und computerling. Forschung	Schriftl.	0,9 Mio.	Morphosyntax, Morphologie, Syntax, Semantik	Kostenlose Lizenz
Tübinger Baumbank (TüBa-D/Z)	http://www.sfs.uni-tuebingen.de/ascl/ressourcen/corpora/tuebadz/	Datenbasis für maschinelles Lernen	Schriftl.	1,4 Mio.	Morphosyntax, topologische Felder, Semantik, Koreferenz	Lizenz

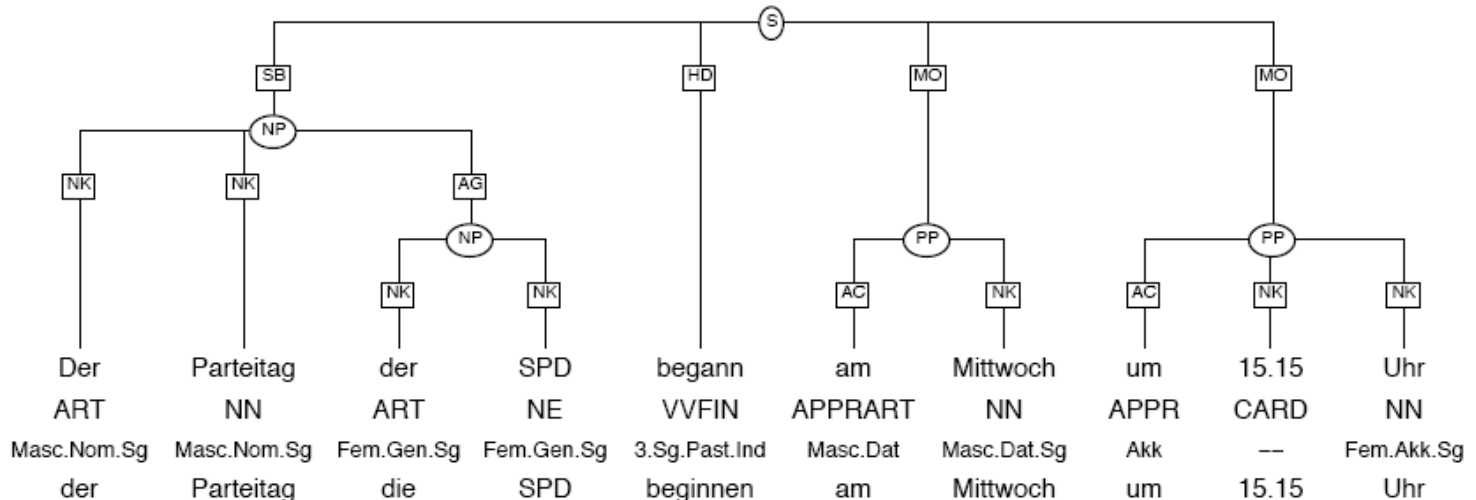
Syntaktisch annotiertes Korpus: Tiger

Dieses Korpus enthält mehr als 40,000 Sätze aus Zeitungstexten und ist nicht nur morphologisch annotiert, sondern auch syntaktisch geparkt. Informationen unter:

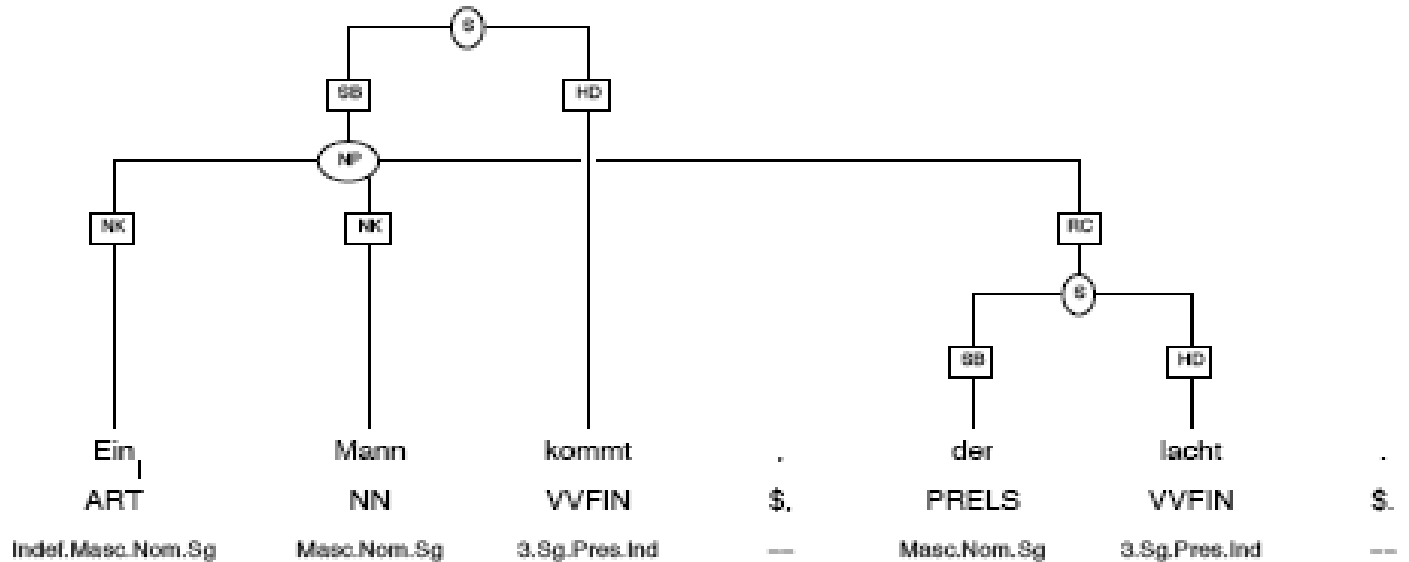
<http://www.ims.uni-stuttgart.de/forschung/projekte/tiger.html>

<http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/icarus.html>

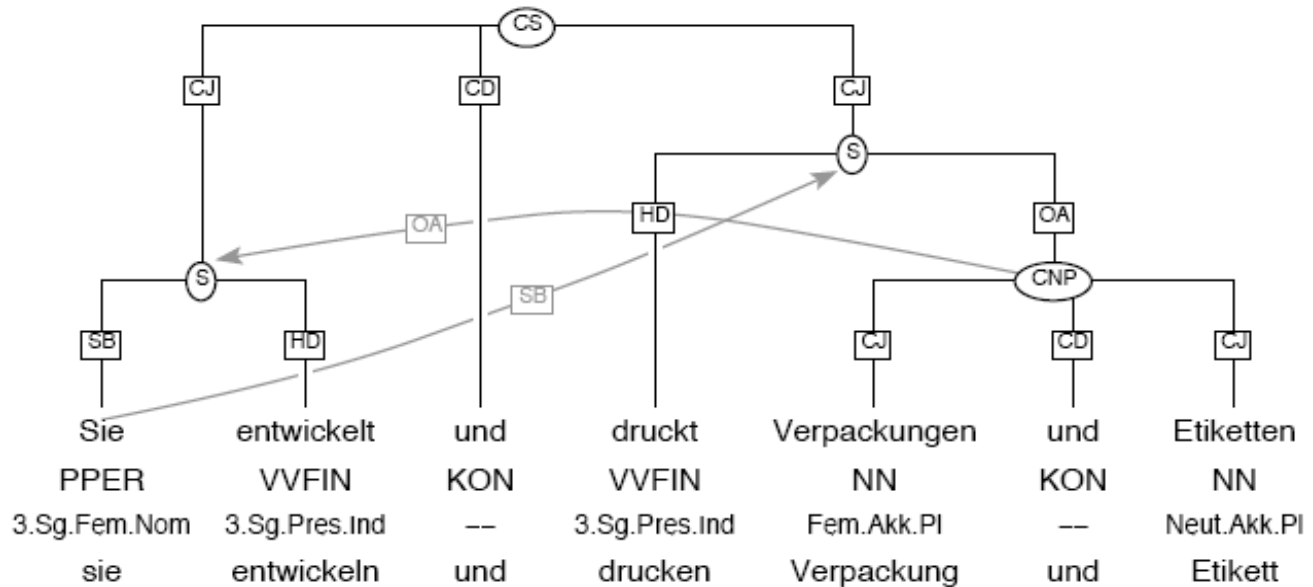
Beispiel einer syntaktischen Annotation: Graphen.



Beispiel mit extraponiertem Relativsatz, kreuzende Kanten:



Beispiel für Koordination, geteilte Konstituente:



Ziel der syntaktischen Annotation ist nicht die Analyse nach bestimmten syntaktischen Theorien (Theorie-Neutralität).

Abfragesprache für TIGER:

Für die Suche nach Strukturen wurde eine Abfragesprache entwickelt:

- > Dominanzbeziehung,
- >* verallgemeinerte Dominanz,
- . Präzedenz,
- \$ Geschwisterbeziehung

Die folgende Frage identifiziert Sätze mit pronominalen Subjekten:

```
[cat = "S"] >SB [pos = "PPER"]
```

Die folgende Frage will solche Sätze mit finitem Verb:

```
([cat = "S"] >SB [pos = "PPER"] & [cat = "S"] > [pos = "VVFIN"])
```

Siehe: <http://homepage.ruhr-uni-bochum.de/stephen.berman/Korpuslinguistik/TIGER.html>

Abfragesprachen

Reguläre Ausdrücke

Damit bezeichnet man Verfahren der Beschreibung von Strings, die – mehr oder weniger vollständig – in Abfragesprachen übernommen werden.

Hinweise siehe http://de.wikipedia.org/wiki/Regul%C3%A4re_Ausdr%C3%BCcke

Corpus Query Processor (CQP)

Siehe Einleitung von Amir Zeldes auf der Moodle-Seite.

Informationen zu linguistischen Korpora

- Korpus-Seite der *Linguist List*: <http://www.linguistlist.org/>
- Spezialkorpora gesprochener Sprache: <http://talkbank.org/>

Wörterbücher im Internet

Es gibt eine Reihe von Internet-Tools für die lexikalische Beschreibungen, die für die linguistische Arbeit von Bedeutung sind.

Beispiel: Wörterbuch zum DWDS-Korpus: www.dwds.de

Wörterbuch-Portale

Portal <http://woerterbuchnetz.de> mit vielen Wörterbüchern, u.a.:


- Wörterbuch von Adelung
- Deutsches Wörterbuch von Jacob und Wilhelm Grimm
- Mittelhochdeutsche Wörterbücher von Lexer, von Wilhelm Müller & Friedrich Zarncke
- Diverse Mundarten-Wörterbücher
- Das Goethe-Wörterbuch der Berlin/Brandenburg. Akademie der Wissenschaften.

Die Wörterbücher und Nachschlagewerke (Mit * gekennzeichnete Wörterbücher sind externe Angebote.)

Adelung Grammatisch-Kritisches Wörterbuch der Hochdeutschen Mundart ⁽¹⁾	BMZ Mittelhochdeutsches Wörterbuch von Benecke, Müller, Zarncke ⁽²⁾	DRW Deutsches Rechtswörterbuch* ⁽³⁾
DWB Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm ⁽⁴⁾	ElsWB Wörterbuch der elsässischen Mundarten ⁽⁵⁾	FindeB Findebuch zum mittelhochdeutschen Wortschatz ⁽⁶⁾
GWB Goethe-Wörterbuch ⁽⁷⁾	Hederich Gründliches mythologisches Lexikon von Benjamin Hederich	Kruenitz Oekonomische Encyklopaedie von Johann Georg Krünitz* ⁽⁹⁾
LEI Lessico Etimologico Italiano ⁽¹⁰⁾	LLU Lexikon der Luxemburger Umgangssprache* ⁽¹¹⁾	LWB Luxemburger Wörterbuch* ⁽¹²⁾
Lexer Mittelhochdeutsches Handwörterbuch von Matthias Lexer ⁽¹³⁾	LmL Lexicon musicum Latinum medii aevi ⁽¹⁴⁾	LothWB Wörterbuch der deutsch-lothringischen Mundarten ⁽¹⁵⁾
MHDBDB Mittelhochdeutsche Begriffsdatenbank*	MWB Mittelhochdeutsches Wörterbuch* ⁽¹⁷⁾	Meyers Meyers Großes Konversationslexikon ⁽¹⁸⁾
NLexer Nachträge zum Mittelhochdeutschen Handwörterbuch von Matthias Lexer ⁽¹⁹⁾	NRhWB Nachträge zum Rheinischen Wörterbuch ⁽²⁰⁾	PfWB Pfälzisches Wörterbuch ⁽²¹⁾
REDE Regionalsprache.de*	RhWB Rheinisches Wörterbuch ⁽²³⁾	WLM Wörterbuch der Luxemburgischen Mundart* ⁽²⁴⁾
Wander Deutsches Sprichwörter-Lexicon von Karl Friedrich Wilhelm Wander		

Deutsches Wörterbuch von Jacob und Wilhelm Grimm (1852 – 1960):


<http://woerterbuchnetz.de/DWB/>

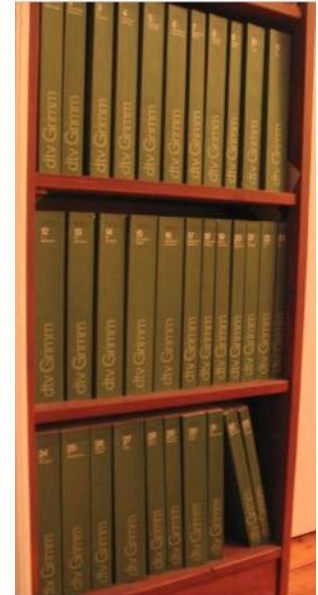
Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm 

Wörterbuch Vorworte Entstehungszeit des DWB **NEU!**

🔍 A bis aalglatt (Bd. 1, Sp. 1 bis 5) ◀

A
A
Ä
Ä
AA, *f.*
AA, *m.*
AAL, *m.*
AALBEERE, *f.*
AALEN
AALFANG, *m.*
AALGLATT
AAIHAUT, *f.*

 **A**, *der edelste, ursprünglichste aller laute, aus Brust und Kehle voll erschallend, den das Kind zuerst und am leichtesten hervorbringen lernt, den mit Recht die Alphabete der meisten Sprachen an ihre Spitze stellen. a hält die Mitte zwischen i und u, in welche beide es geschwächt werden kann, welchen beiden vielfach es sich annähert. Vorgeschichte und Geschichte unserer Sprache verkünden solche Übergänge allenthalben: lat. pater Iupiter Diespiter, goth. fadar, vater; lat. taceo conticeo, goth. þaha, ahd. dagêrn; lat. sapio desipio, goth. safja; lat. habeo cohibeo, goth. haba, ahd. hapêrn; skr. saptan, goth. sibun; skr. navja, litt. naujas, goth. niujis; skr. madhja, goth. midjis; skr.*



Das DWB ist auch über die DWDS-Seite einsehbar, zusammen mit

'DWB (1854-1961)

spiel Detailansicht

Fundstelle: Lfg. 13 (1904), Bd. X,I (1905), Sp. 2275, Z. 20

SPIEL, n.

ludus, lusus. das wort scheint dem germ. des festlandes ... ↓

I. *spiel im allgemeinen.*

1) *spiel bezeichnet im allgemeinen eine ...* ↓

a) *so besonders spiel der kinder: ...* ↓

b) *von thieren: da trippelt' ich (...* ↓

c) *in bezug auf erwachsene menschen ist spiel ...* ↓

d) *specialisiert spiel mit etwas, ...* ↓

2) *spiel wird zunächst nur ...* ↓

a) *von theilen des menschlichen leibes, spiel ...* ↓

b) *spiel der augen, der blicke u. ähnl.: ...* ↓

c) *spiel der lichter, schatten: die zwei flämmchen bewegten ...* ↓

d) *vom winde: heisz kam ...* ↓

1dwb-@rev145

WDG (1961-1977)

Spiel

Spiel, das: -(e)s, -e

1. *zwanglose, frei gewählte, nicht oder nur mittelbar gesellschaftlich oder ökonomisch zielgerichtete Tätigkeit a) die der Unterhaltung, dem Zeitvertrieb, dem Vergnügen, der Entspannung dient und oft keine besondere Anstrengung erfordert: ein lustiges, spannendes, langweiliges, lehrreiches S.; ein S. anregen, anfangen, abbrechen, machen, gewinnen, verlieren, aufgeben, verloren geben; das Kind ist ganz in sein S. vertieft; den Haushalt schafft sie wie im S. (bewältigt sie mühelos, leicht) es sind noch alle im S. (alle machen noch mit, es ist noch niemand ausgeschlossen) b) die nach festgelegten Regeln unter bestimmten Bedingungen durchgeführt wird, bes. sportlicher Kampf von Mannschaften: ein*

Etymologisches Wörterbuch (nach Pfeifer)

Spiel, ferner: [u'spielen', u'spielend', u'abspielen', u'aufspielen', u'zuspielen', u'Anspielung', u'Spieler', u'spielerisch', u'Spielmann', u'Spielraum', u'Spielsachen', u'Spielwaren', u'Spielzeug']

Spiel n. 'nicht auf Nutzen ausgerichtete, vergnügliche, mit Ernst betriebene Tätigkeit, Zeitvertrieb, Vergnügen, Wettkampf'. Die Herkunft des nur kontinentalwestgerm. bezeugten Substantivs (bzw. Verbs, s. unten) ahd. (9. Jh.), mhd. *spil* 'Tanz, Zeitvertrieb, Scherz, Unterhaltung, Vergnügen, Musik, Waffen-, Kampfspiel, Wettkampf', asächs. *spil*, mnd. mnl. *spel*, *spil*, nl. *spel* (schwed. *spel*, dän. *spil*, norw. *spill* sind wie aengl. *spilian*, schwed. *spela*, dän. *spille* Entlehnungen) ist unbekannt. Dazu schwach flektierendes **spielen** Vb., ahd. *spilān* (8. Jh.), mhd. *spiln*, *spilen* 'Scherz treiben, sich vergnügen (mit Leibesübungen, Kampfspiel, Brett- oder Würfelspiel), sich lebhaft bewegen,

Weitere Wörterbuch-Portale:

- <http://www.linguistlist.org/sp/Dict.html>
- <http://www.yourdictionary.com>
- <http://www.mavicanet.com/directory/eng/8403.html>

Das Oxford English Dictionary ist über die HU-Webseite zu bekommen:

- <https://ssl.cms.hu-berlin.de/,DanaInfo=www.oed.com+>

Wortschatz Uni Leipzig

Internet-Adresse:

<http://wortschatz.uni-leipzig.de/>

Ein Wörterbuch-Portal zum heutigen Deutsch:

mit Hinweisen zu aktuellen Wörtern.

Hier finden Sie auch Zugang zu weiteren Wörterbüchern, auch zweisprachigen.

Beispiel Wortsuche: *aufheben*

Wort: aufheben

Anzahl: 619

Häufigkeitsklasse: 13 (d.h. *der* ist ca. 2¹³ mal häufiger als das gesuchte Wort)

Morphologie: auf|he|ben

Grammatikangaben: Wortart: Verb

Partizip II mit haben

lautet ab

transitiv

Präfix: **auf**

Relationen zu anderen Wörtern:

- Synonyme: **abschaffen**, **abschließen**, **aufbewahren**, **aufklauben**, **auflesen**, **au beseitigen**, **bewahren**, **einstellen**, **erheben**, **erübrigen**, **hamstern**, **heben**, **hoch speichern**, **zurückbehalten**, **zurückhalten**, **zurücklegen**
- vergleiche: **annullieren**, **beseitigen**, **streichen**, **verwahren**, **widerrufen**
- ist Synonym von: **absagen**, **abschaffen**, **abzweigen**, **abzweigen**, **annullieren**, **aufnehmen**, **auffragen**, **auf sammeln**, **aufsparen**, **aufsuchen**, **ausgleichen**, **aus einkleben**, **einmotten**, **einschließen**, **einsparen**, **emporheben**, **erheben**, **heber klauben**, **löschen**, **lüften**, **reservieren**, **übriglassen**, **verschließen**, **verstauen**, **zurückbehalten**, **zurücklegen**, **zurücknehmen**, **zusammennehmen**
- wird referenziert von: **annullieren**

Links zu anderen Wörtern:

- Grundform: **aufheben**
- Teilwort von: **Beschränkungen aufheben**, **die Blockade aufheben**, **alle Besch**
- -ung-Form von: **Aufhebung**
- Form(en): **aufgehoben**, **aufzuheben**, **aufheben**, **aufhebt**, **aufhob**, **aufgehoben aufhebend**, **aufhebenden**, **aufhebende**, **aufhebst**

Dornseiff-Bedeutungsgruppen:

- 4.26 Nichts, Null: **aufheben**, **leeren**, **räumen**, **vernichten**
- 5.43 Erhaltung: **aufbewahren**. **aufheben**. **auslagern**. **bewahren**. **einkochen**. **e**

OWID

Lexikon-Portal des Instituts
für Deutsche Sprache (IDS),
Mannheim.

<http://www.owid.de/>

Gegenwärtig:

lexikon,

Neologismen,

Feste Wortverbindungen,

Diskurswörterbuch

Beispiel einer

Kollokationsanzeige

für *aufheben*:



770202 verschiedene Kookkurrenzpartner gespeichert

als Kookkurrenzpartner

alphabetisch nach Kohäsion nach W

Suchmuster (regexp) verwenden

Textwort **aufheben** tritt als Kookkurrenzpartner bei folgende Analysewörtern und -typen auf

Wort	Synsemantika	
	ohne	mit
Sanktion	2364	2364
Immunität ⊗	1797	1797
Waffenembargo ⊗	1075	1075
gegenseitig ⊗	957	957
Embargo ⊗	840	840
Verbot ⊗	753	753
Trennung ⊗	617	617
Importverbot ⊗	609	609
Blockade ⊗	571	571
Beschränkung ⊗	495	495
verhängt ⊗	490	490
Sperre ⊗	483	483
verhängen	472	472
Beschluß ⊗	464	



Paradigmenübersicht

Alphab

Paradigmen Verben

Allgemeine	Typ:	abstreiten	
Repräsen	Assertive	anflunkern	s. schwindeln
Direktive	Informationsverben	anlügen	s. lügen
Kommissive		anschwindeln	s. schwindeln
Expressive		anzweifeln	
Deklarative		argumentieren	
Gesprächs- und themenstrukturierende		beharren auf	s. bestehen auf
		behaupten	
		beipflichten	s. zustimmen
		bejahen	s. zustimmen
		bekräftigen	s. beteuern
		bekunden	s. beteuern
		belügen	s. lügen
	

Canoonet:

Wörterbuchportal mit zahlreichen grammatischen und idiomatischen Informationen (hier ebenfalls für das Wort *aufheben*), siehe <http://www.canoo.net/>

aufheben: Verb, Hilfsverb haben [Rechtschreibung](#) [Wortformen](#) [Wortbildung](#)
Aufheben: Nomen, neutrum [Rechtschreibung](#) [Wortformen](#) [Wortbildung](#)

Bedeutungswörterbuch

Aufheben

Synonyme [Abschaffung](#), [Abschaffen](#), [Aufhebung](#)
Oberbegriff [Beendigung](#)
Unterbegriffe [Deregulierung](#), [Streichung](#), [Annullierung](#)

aufheben

Bedeutung 1

Im Sinn von (jmdn/etwas, was liegt) in die Höhe heben
Beispiel "Er hob das Papier vom Boden auf."

Oberbegriff [heben](#)

Unterbegriffe [aufklauben](#), [hochnehmen](#)

Bedeutung 2

Im Sinn von (etwas) offiziell beenden
Beispiel "Er hob die Tafel auf."

Präteritum Nebensatz				
	Indikativ		Konjunktiv II	
	Person	Verb	Person	Verb
...dass	ich	aufhob	ich	aufhöbe
	du	aufhobst	du	aufhöbest aufhörst
	er/sie/es	aufhob	er/sie/es	aufhöbe
	wir	aufhoben	wir	aufhoben
	ihr	aufhobt	ihr	aufhötet aufhöt
	sie	aufhoben	sie	aufhoben

Grammatiken

Zum Deutschen: Seiten des IDS in Mannheim:

- progr@mm: Propädeutische Grammatik, für Studenten der Linguistik.
<http://hypermedia.ids-mannheim.de/programm/> .
- Grammis: Datenbank auf Grundlage der IDS-Grammatik
<http://hypermedia.ids-mannheim.de/>
- EValbu: Valenzwörterbuch
<http://hypermedia.ids-mannheim.de/evalbu/index.html>

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

abbliegen
abfahren
abfliegen
abgeben
abhängen I
abhängen II
abheben
abheben, sich
abholen
ablehnen
ableiten
ableiten, sich
abmachen

Das elektronische Valenzwörterbuch deutscher Verben

In E-VALBU, dem Valenzwörterbuch des Instituts für Deutsche Sprache (IDS), liegt das Gewicht vor allem auf der inhaltlichen und formalen Erfassung der Umgebung der behandelten Verben (Valenzinformationen). Sie finden aber auch Informationen zu ihrer:

- Bedeutung
- Konjugationsklasse
- Aussprache (Akzent)
- Stilebene
- Passivfähigkeit

ProGr@mm
die Propädeutische Grammatik

Grammatisches Grundwissen
Seminarbausteine
Kontrastiv
Terminologisches Wörterbuch
Grammatisches Wörterbuch
Grammatische Bibliografie
Forum

Zu anderen Sprachen

Portal: <http://linguistlist.org/langres/index.html>

Tonaufnahme, Transkription & Annotation

Untersuchung gesprochener Sprache

Mündliche und schriftliche Sprache sind verschiedene Ausprägungen des Sprachgebrauchs; es gibt auch Zwischenformen:

- Gelesene Sprache, sorgfältiger Vortrag: konzeptionelle Schriftlichkeit
- informelle schriftliche Sprache, z.B. Internet-Foren: konzeptuelle Mündlichkeit

Mündliche Sprache ist primär:

- Historisch sind mündliche Sprachen weit älter
- Spracherwerb in der Erstsprache zunächst mündlich
- Menge der produzierten und verarbeiteten Sprache: in der Regel größer bei mündlicher Sprache

Eigenschaften mündlicher Sprache (im Vergleich zu schriftlicher):

- Planung eingeschränkt; kürzere Satzäußerungen, Abbrüche, Verschränkungen, Korrekturen
- Interaktion: Rückmeldesignale, Sprecherwechsel, Aufgreifen von Konstruktionen
- Spezielle Eigenschaften mündlicher Sprache, zum Beispiel Topik-Drop:
A: *Wer ist denn der Typ dort drüben?*
B: *__ Kenn ich nicht.*

Untersuchung gesprochener Sprache:

- Akustische Aufnahmen, Videoaufnahmen (z.B. Gesten, Interaktion)
- Verschriftlichung, die Aspekte der Mündlichkeit mit erfasst: Transkription

Ton- und Videoaufnahme

Digitale Aufnahmegeräte

PCM-Recorder (Pulse-Code-Modifikation: Repräsentation der Amplitude in regelmäßigem Abstand).

- Computer (auch Handy) mit eingebautem oder (besser) externem Mikrofon
- Spezialisierte Aufnahmegeräte mit Speicherkarten, Beispiele für < 200 Euro:



Olympus
LS 5



Tascam
DR-07



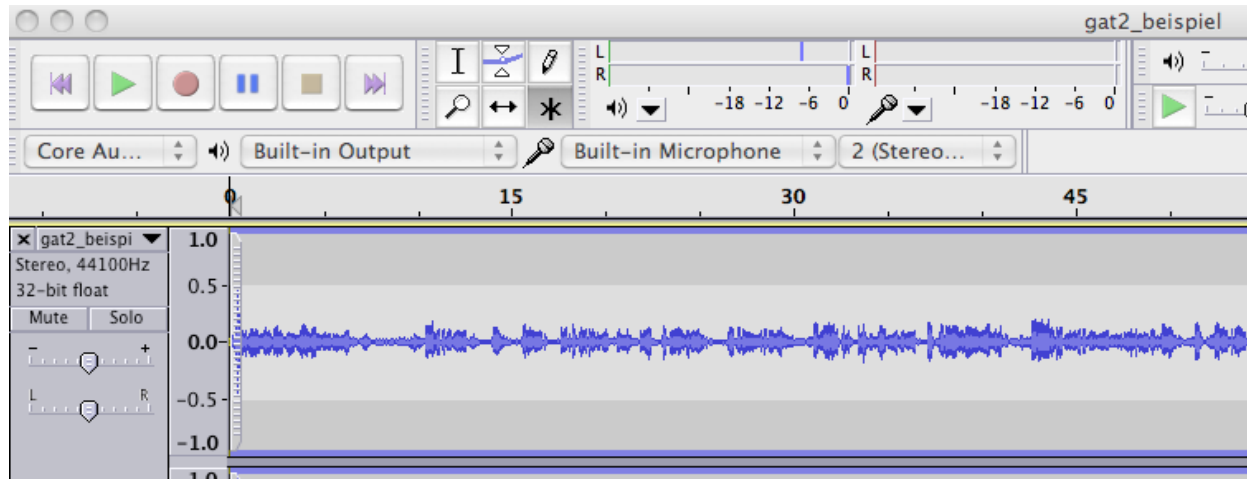
Zoom Q3

Dateiformate

- Nicht-komprimiert: WAV
- Komprimierte Dateien: MP3 (MPEG) (patentiert), Ogg-Vorbis (patentfrei), MP4 (Multimedia), wesentlich kleinere Dateien, aber Informationsverlust (Problem bei phonetischen Untersuchungen)
- Dateien aus Rundfunk/Fernsehen, Bundestagsdebatten usw. liegen i.d.R. komprimiert vor.

Tonbearbeitung

Es gibt verschiedene Programme zur Tonbearbeitung;
hier empfohlen: Audacity, <http://audacity.sourceforge.net/>



Beispiel: Arbeit mit Audacity

- Einlesen einer Tondatei.
- Aufnehmen einer Tondatei "Eins zwei drei..."
- Stereo → Mono, Anpassen Amplitude, Herausschneiden und Wiedereinsetzen
- Export in verschiedene Dateiformate

Transkription

Grundsätzliches

Verschriftlichung von Audio-Dateien nach Kategorien, die für die geplante Untersuchung relevant sind
Anders als bei der glättenden Niederschrift etwa von Interviews in Zeitungen geht es oft darum, Eigenschaften des mündlichen Sprachgebrauchs exakt festzuhalten:

- Verschleifungen, Zögerlaute, Wiederholungen, genaue phonetische Realisierungen, prosodische Effekte
- Dialogorganisation: Sprecherwechsel, Pausen, Parallelsprechen
- Elemente der Gestik, der Mimik oder der Situation.

Es gibt dabei keine “absolute” Transkription, die alles mit maximaler Genauigkeit festhalten würde:

- Die anzustrebende Genauigkeit hängt vom Zweck der Transkription ab.
- Zu einem gewissen Teil auch theoriegeleitet.

Transkriptionskonventionen

- GAT / GAT2 (“Gesprächsanalytisches Transkriptionssystem”, Margret Selting u.a.)
- HIAT (“Halb-Interpretative Arbeits-Transkription”) (Ehlich und Rehbein)
- DIDA (Diskursdatenbank) (IDS Mannheim)
- CHAT (“Codes for the Human Analysis of Transcriptions”), v.a. für Kindersprache
Einführung: Norbert Dittmar. 2002. *Transkription. Ein Leitfaden mit Aufgaben für Studenten*.
Opladen: Leske und Buderich.

Transkriptionssystem GAT / GAT 2

Selting et.al.: “Gesprächsanalytisches Transkriptionssystem (GAT)”,

Selting, Margret e.a. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2).
Gesprächsforschung - Onlinezeitschrift zur verbalen Interaktion 10: 353-402,

siehe <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf>

Vgl. Webseite: <http://agd.ids-mannheim.de/html/gat.shtml>

Vgl. Online-Tutorial: <http://paul.igl.uni-freiburg.de/gat-to/>

Prinzipien

- Die Transkription soll auf einfache Weise verfeinert und ausgebaut werden können.
- Die Transkription soll trotz aller Notation gut lesbar sein, auch für Nicht-Linguisten (Sozialwissenschaftler, Psychologen usw.)
- Die Transkription soll ökonomisch und eindeutig sein.
- Sie soll möglichst ohne Sonderzeichen auskommen.
- Sie sollte Phänomene erfassen, die für die Gesprächsforschung von potentieller Relevanz sind.

Wichtig: Oft ist es aus rechtlichen Gründen nötig, die Personen zu anonymisieren (Konvention: Vornamen bei Du-Beziehungen, Vor- und Nachnamen bei Sie-Beziehungen).

Metadaten (Transkriptionskopf)

Angaben über eine bestimmte Aufnahme und deren Transkription:

- Herkunft, Zugehörigkeit zu einem bestimmten Korpus bzw. Projekt,
- Aufnahmeummer oder Kennwort/Name des Gesprächs;
- Aufnahmetag, Ort der Aufnahme; Dauer der gesamten Aufnahme;
- Name der/des Aufnehmenden; Name der/des Transkribierenden;
- kurze Charakterisierung der Situation, z. B. Interview, informelles Gespräch, Telefongespräch, Radio-Anrufsendung;
- kurze Charakterisierung der Teilnehmerrollen, z.B. informelles Gespräch mit gleichberechtigten Teilnehmern, Ärztin und Patient, Lehrer und Schülerin;
- kurze Charakterisierung der Sprechenden unter Angabe ihrer Decknamen, z. B. Geschlecht, Alter, Beruf. Altersangabe bei Kindern z.B. 4;11 (4 Jahre 11 Monate).
- sonstige Informationen, die relevant sein könnten, z. B. Herkunft, Dialektalisierungsgrad,
- ggf. Hinweis auf durchgängige Kennzeichen wie eine besonders schnelle Sprechgeschwindigkeit, eine extrem hohe Stimme o.ä.;
- kurze Charakterisierung des Gesprächsverlaufs (in Form einer Art Inhaltsangabe, ggf. mit Verweis auf für die Analyse interessante Phänomene);
- ggf. Hinweis auf Bearbeitungsstand der Transkription.

Einige allgemeine Konventionen

- Oft werden Schriften mit konstanter Buchstabenbreite verwendet (Courier).
- Generelle Kleinschreibung; Großbuchstaben dienen der Notation von Akzenten.
- Transkriptionszeilen werden nummeriert: Zeilennummer – 3 Leerzeichen – Sprecherkennzeichen: --- 3 Leerzeichen – Transkripttext.
- Zeilen sollen kurze, sinngemäße Einheiten bilden; Überlappungen werden angezeigt.
- Zum Basistranskript können weitere Zeilen (Übersetzung (in Kursiv), evt. Interlinearübersetzung, Gestik und Mimik etc.) hinzugefügt werden.

Beispiel:

```
01  A:  hier fängt der transkripttext an
02  B:  ja genau
03      (---)
04  A:  wenn du mir ins wort fä[llst
05  B:                                     [ich fall dir ja
06      gar nicht ins wort
07  A:  doch
08      (.)
09      hast du wohl getan
10      du hast (.) mich grade unterbrochen
11  B:  ja
12      tut mir leid
```


Umschrift

GAT empfiehlt eine relativ standard-nahe Umschrift.

Vorteil: leichtere Lesbarkeit als bei phonetischer Umschrift; man kann nach Wörtern besser suchen.

Klitische Elemente sollen als solche wiedergegeben werden,

z.B. *schauma* 'schauen wir', *ich kauf n buch* usw.

Sprecherwechsel

Neue Sprecherbeiträge beginnen auf einer neuen Zeile; bei Überlappungen werden eckige Klammern gesetzt .

01	A:	eigentlich wollt ich jetzt [noch sa'
02	B:	[das tut nichts [zur sache
03	A:	[noch sagen
04		daß mich diese ständige unterbrecherei einfach stört.
10		du hast (.) mich grade unterbrochen
11	B:	ja
12		hab ich das wirklich getan
13		(.)
14		okay
15		tut mir leid
16	A:	schon gut

Pausen

(.) Mikropause, (-), (--), (---) längere Pausen; ab Pausen von einer Sekunde Angabe der Sekunden (2.0), (3.8) etc. Pausen innerhalb eines Turns werden innerhalb der Zeile notiert, sonst als eigene Zeile.

Weitere segmentale Transkriptionskonventionen

- Dehnungen werden durch :, ::, ::: wiedergegeben, z.B. so:::
- Verzögerungssignale durch *äh, öh* etc.
- Glottalverschluss durch ' , z.B. *ich hab geda'*
- Lachen segmental: *hahaha, hehe* oder durch Beschreibung in Doppelklammern: *((lacht))*
- Rezeptionssignale durch *hm, ja, nein, nee*, zweisilbig *hm=hm, ja=a*; reduplizierendes verneinendes Signal mit Glottalverschluss *'hm'hm*.

Prosodie

Die Prosodie ist ein eigenes komplexes System, mit unterschiedlichen Notationskonventionen.

Am weitesten verbreitet ist das ToBI-System, vgl. <http://www.ling.ohio-state.edu/~tobi/>, in seiner deutschen Version <http://www.uni-koeln.de/phil-fak/phonetik/gtobi/index.html>, u.a. Tutorials!

Die GAT-Empfehlungen sagen:

- Der Hauptakzent jeder Phrasierungseinheit wird durch Großbuchstaben der akzentuierten Silbe dargestellt; besonders starker Akzent wird durch Rufezeichen markiert.
- Tonhöhenverläufe:
? hoch steigend, , mittel steigend, - gleichbleibend, ; mittel fallend, . tief fallend.

Weitere Konventionen

- Nichtverbale Handlungen in Doppelklammern: *((schnieft)), ((hustet))*
- Interpretierende Kommentare zur Verbalisierung in eckigen Klammern, mit Angabe der Reichweite:
<<*empört*>> *also SO was.* >

- Unverständliche Passage in einfachen Klammern:
 () nicht verständlich
 (*solche*) vermuteter Wortlaut
 (*solche / sonne*): Alternativen
 (*s/f*)*angen*

Feinere Konventionen

Bei speziellem Interesse können die oben angeführten Konventionen des sogenannten **Basistranskripts** weiter verfeinert werden. Beispiele:

- Hauptakzent und Nebenakzent: *akZENT* vs. *akzEnt*
- Tonhöhen sprünge nach oben/unten: ↑, ↓
- Verändertes Tonhöhenregister: <<h>...> hoch, <<t>...> tief
- steigender, fallender, gleichbleibender, steigend-fallender, fallend-steigender Akzent auf Silbe: *so*
˘SO, *so`SO*, *so⁻SO*, *so^SO*, *soˇSO*
- Dabei können Intonationssignale auch in einer Sonderzeile stehen:
 / hoch steigend, / mittel steigend, - gleichbleibend, \ mittel fallend, _ tief fallend,
 ^ steigend-fallend, v fallend-steigend (für weiteres siehe den Originalartikel)
- Lautstärke: <<f> ...> forte, <<ff> ...> fortissimo, <<p> ...> piano, <<pp> ...> pianissimo,
 <<res> ...> lauter werdend, <<dim> ...> leiser werdend
- Geschwindigkeit: <<all> ...> allegro, <<len> ...> lento,
 <<acc> ...> schneller werdend, <<rall> ...> langsamer werdend
- Einatmen: *.h*, *.hh*, *.hhh*; Ausatmen *h*, *hh*, *hhh*, je nach Dauer, *.hh* und *hh* normale Länge.

➤ Notation von Aspekten der

-- Proxemik

(Zu- und Abgänge,
relative Distanz
von Personen)

-- Kinesik

(Körperbewegungen)

-- Gestik

-- Mimik

-- Blickrichtung

in gesonderten Zeilen,
mit Angabe der Dauern.

Für diese Zwecke kann man
auch Bilder integrieren,
welche die Konversations-
teilnehmer zeigen.

```
01 Holger: bei netzwerk KlapPTS nich,  
02 weil zu kompliziert,  
03 zu viele verschiedene  
04 produkte, .hhh e: en TE wär dafür ideal:,  
|-----|  
| Michael D. kommt herein  
05 aber da ham wer einfach kein voLumen,  
-----|  
| geht zu seinem vorherigen Sitzplatz  
06 un em pe is zu Wenich.  
-----|  
| und setzt sich
```

Beispiel GAT-Analyse: Grobtranskript

Text: <http://www.mediensprache.net/de/medienanalyse/transcription/gat/>

01	S1:	ja:; (.) die VIERziger generation so;=	03
02		=das=s: !WA:HN!sinnig viele die sich da ham [SCHEIden	04
03	S2:	[ja;	05
04	S1:	lasse[n.=	06
05	S2:	[hm,	07
06	S1:	=oder scheiden lassen ÜBERhaupt.	08
07	S2:	hm,	09
08		(--)	10
09	S1:	heute noch-	11
10		(2.1)	12
11		s=is der UMbruch.	13
12	S2:	n besonders GUTES beispiel das warn mal unsere NACHbarn.	14
13		(1.0)	15
14		ähm (1.0)	16
15		DREISsig jahre verheiratet, (--)	17
16		das letzte kind (.) endlich aus m HAUS,	18
17		zum stuDIERN, (--)	19
18		WEGgegangen,=ne,	20
19		nach berLIN, (--)	21
20		und (.) die ältere tochter is AUCH in berlin gewesen? (1.1)	22
21		und (.) der KE:RL,	23
22		das war aber ein penETRANter: !WI!derling.=also (1.0)	24
23		der hat (.) äh sein GARTen wie (.) !PIK! AS (--) gePFLEGT,=	25

24	=ne,	42
25	!KEIN! BLÄTTchen,=	26
26	=und NICHTS,=	27
27	=englischer Rasen, (--)	28
28	un:dh: bei !JE!der gelegenheit hat er die polIZEI	29
29	gerufen,	30
30	und sich mit den NACHbarn ange[legt,=ne, (1.2)	31
31	S1: [phhh hohoho	32
32	S2: un wenn da: einmal: jemand zum abschied geHUPT hat,	33
33	da war der in NULL komma nix draußen;	34
34	und hat da RUMgeschrien;=	35
35	=ich hol die polIZEI: und [so-	36
36	S1: [das GIBS doch wohl nich.	37
37	S2: ja; V:OLLkommen Widerlich.=ne, (--)	38
38	un:dh: (--) dann hatte er do son (.) son KLEInen BA:RT	39
39	hier,=ne, (.)	40
40	und ham wir immer gesagt HIT[ler;=ne,	41
41	S1: [HITler;	42
42	S1: h[m,	

¶

Beispiel GAT-Analyse: Feintranskript

01 S1: ja; (.) die [↑]VIERziger genera[↑]tiOn so;=
02 =das=s: [↑]!WA:HN!sinnig viele die sich da ham [[↑]SCHEiden
03 S2: [ja;
04 lasse[n.=
05 S2: [[↑]hm, ¶
06 S1: =<<dim> oder [↑]schEiden lassen [↑]ÜBERhaupt.>
07 S2: [↑]hm,
08 (--)
09 S1: <<pp> heute noch- >
10 (2.1)
11 <<p> s=is der [↑]UMbruch.>
12 S2: n besonders [↑]GÜtes beispiel das warn mal unsere [↑]NACHbarn.
13 (1.0)
14 ähm (1.0)
15 [↑]DREIßig jahre ver[↑]hEiratet, (--)
16 das letzte kind (.) [↑]Endlich aus m [↑]HAUS,
17 zum stu[↑]DIERN, (--)
18 [↑]WEGgegangen,=[↑]ne, .h
19 nach ber[↑]LIN, (--)
20 und (.) die [↑]Ältere tochter is [↑]AUCH in berlin gewesen? .hh
21 und (.) der [↑]KE:RL,
22 <<t> das war aber eip pene>[↑]TRANter: [↑]!W:I!derling.=also .hh
23 der hat (.) äh sein [↑]GARTEN wie (.) [↑]!PIK! [↑]AS (--)
24 ge [↑]PFLEGT,=[↑]ne,

25 ↑:KEIN: `BLÄTT`chen,=
26 =und `NICHTS,=
27 =`Englischer `RASen, .hh
28 un:dh: bei <<all>↑:JE:der ge`LEgen`heit hat er die poli↑`ZEI
29 ge`rUFen,>
30 <<all> und sich mit den ↑`NACHbarn ange[`legt,=`ne,> (1.2)
31 S1: [<<pp>phhh hohoho>
32 S2: un `wEnn da: `Einmal: `jEmand zum `Abschied ge↑`HUPT `hat,
33 da `wAr der in <<all> `NULL komma nix> `drAußen;
34 und hat da `RUMgeschrien;=
35 =ich hol die poli↑`ZEI: und [so-
36 S1: [<<p> das ↑`GIBS doch wohl nich.>
37 S2: ja; ↑`V:OLLkommen `Widerlich.=`ne, .h
38 un:dh: (--) dann `hAtte er do son (.) son ``KLEInen ``BA:RT
39 hier,=`ne, (.)
40 <<all> und ham wir immer gesagt> `HIT[ler;=`ne,
41 S1: [`HITler;
42 S1: `h[m,

Methoden zur Transkriptionserstellung

In Textverarbeitungsprogrammen: Tabellen

Man kann Transkriptionen nach Konventionen wie der GAT-Konvention ohne spezielle Software herstellen. Es empfiehlt sich dabei natürlich, mit Tabellen zu arbeiten, wobei die Tabellenstriche im Ausdruck unsichtbar sein können.

ZN	SPR	Transkription
01	S1:	ja;; (.) die VIERziger generation so;=
02		=das=s !WA:HN!sinnig viele die sich da ham [SCHEIden
03	S2:	[ja;
04	S1:	lasse[n.=
05	S2:	[hm,
06	S1:	oder scheiden lassen ÜBERhaupt.
07	S2:	hm,

Transkriptions-Werkzeuge

Spezialisierte Werkzeuge zur Erstellung von Transkriptionen

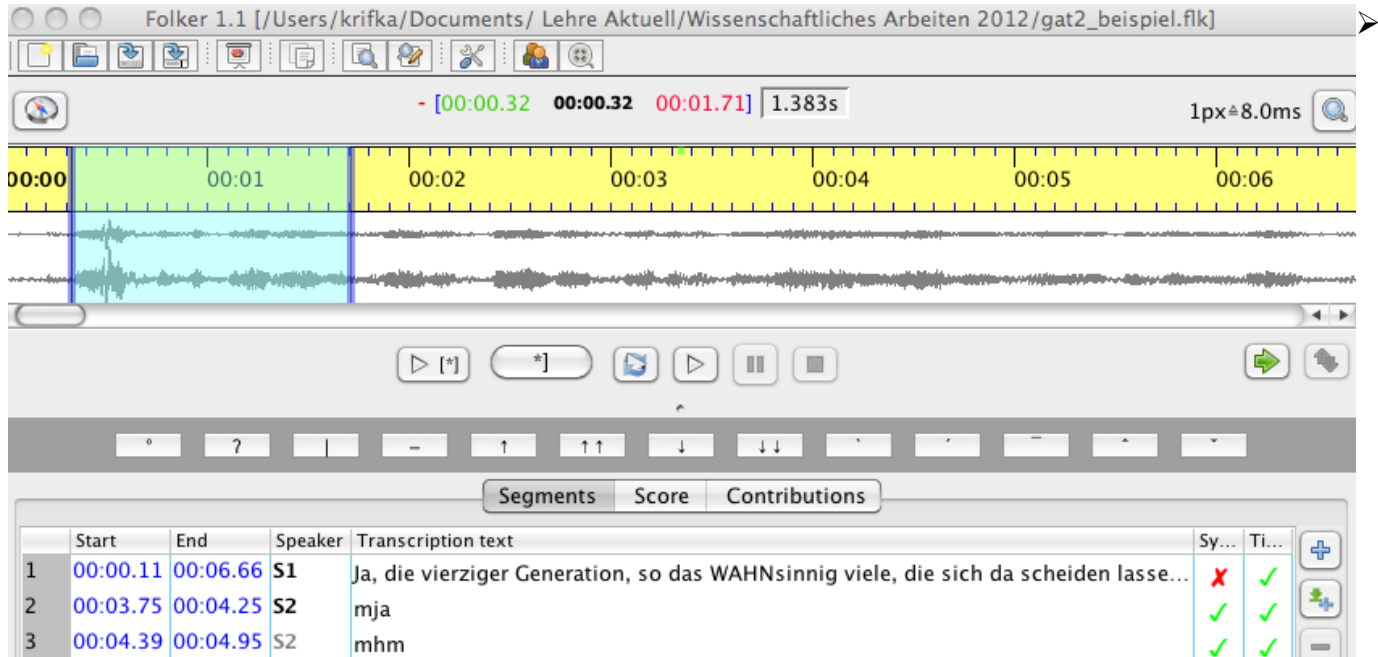
- EXMARALDA: Partitur-Editor, vom SFB Mehrsprachigkeit in Hamburg entwickelt, weit eingesetzt, kostenlos erhältlich: <http://www.exmaralda.org/>
- FOLKER: IDS Mannheim, baut auf EXMARALDA auf, optimisiert für multiple Sprecher, GAT, kostenlos erhältlich: <http://agd.ids-mannheim.de/folker.shtml>
- CLAN: einfacherer Editor, für Spracherwerb / Kindersprache eingesetzt (CHILDES Datenbank), kostenlos erhältlich: <http://childes.psy.cmu.edu>
- ELAN: Editor für Audio- und Video-Annotationen, MPI Psycholinguistik, v.a. ling. Feldforschung, kostenlos erhältlich: <http://tla.mpi.nl/tools/tla-tools/elan/>

Funktionsweise

- Auswahl von Segmenten aus einer Mediadatei; Transkription dieser Segmente
- Genaue Alinierung der Transkriptionen zu den Mediadateien
- Partitur-Editoren: Mehrere Annotationsebenen (z.B. verschiedene Sprecher, Annotation für Gestik)
- Erzeugung von Textdateien zusätzlich zu den Mediadateien, die nicht verändert werden.
- Ausgabe der Textdateien in verschiedenen Formaten, z.B. html oder rtf, zur Einbindung in Dateien von Textverarbeitungsprogrammen.
- Programme zum Durchsuchen von Dateien und zum Aufbau und zur Pflege von Korpora.

Beispiel der Arbeit mit FOLKER

Siehe Handbuch und Schnell-Anleitung

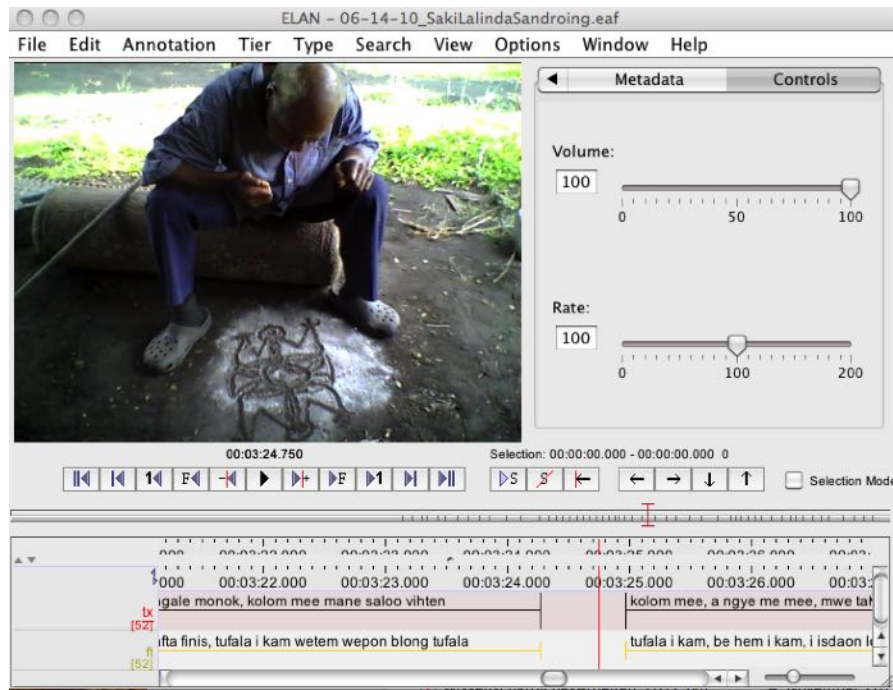


The screenshot displays the FOLKER software interface. At the top, the title bar reads "Folker 1.1 [/Users/krifka/Documents/ Lehre Aktuell/Wissenschaftliches Arbeiten 2012/gat2_beispiel.flk]". Below the title bar is a toolbar with various icons. The main area features a timeline with a yellow background and a green segment from 00:00 to 00:01. A blue vertical line marks the current position at 00:00:11. Below the timeline is an audio waveform. At the bottom, there are playback controls and a table with three tabs: "Segments", "Score", and "Contributions". The "Segments" tab is active, showing a table with the following data:

	Start	End	Speaker	Transcription text	Sy...	Ti...	
1	00:00.11	00:06.66	S1	Ja, die vierziger Generation, so das WAHNSinnig viele, die sich da scheiden lasse...	✗	✓	+
2	00:03.75	00:04.25	S2	mja	✓	✓	+
3	00:04.39	00:04.95	S2	mhm	✓	✓	-

Beispiel der Arbeit mit ELAN

- Werkzeug zur Audio- und Video-Annotation, also z.B. auch für Gesten
- Mehrfache Mediadateien, z.B. Audio mit mehreren Mikrofonen, Video mit mehreren Kameras
- Bevorzugtes Format für The Language Archive,
<http://www.mpi.nl/research/research-projects/the-language-archive>
- Beispiel: Sandzeichnung in Vanuatu
- Beispiel: Annotation Roland Berger, `speak-1380-quest-11468.mov, .wav`



Linguistische Analyse-Werkzeuge

Toolbox

Programm zur Analyse fremder Sprachen:

- Erzeugung eines Wörterbuchs
- Interlinear-Transkriptionen von Texten
- Morphologische Analyse

Nähere Information unter: Field Linguist's Toolbox, <http://www.sil.org/computing/toolbox/>

SIL Fieldworks

Ein neueres Programm, siehe <http://fieldworks.sil.org/>

Korpora für gesprochene Sprache

Für das Deutsche: IDS, Archiv für gesprochenes Deutsch: <http://agd.ids-mannheim.de/index.shtml> ,
insbesondere <http://dsav-oeff.ids-mannheim.de/DSAv/KORPORAI.HTM>

- Deutsche Mundarten (Binnenmundarten, Auslandsmundarten)
- Verbale Interaktionen:
 - 2.1. Sprechen im sozialen Kontext
 - Beratungsgespräche
 - Biographische und Reiseerzählungen
 - Dialogstrukturen
 - Genese von "Eurotexten": Verhandlungen und Interviews in einer EG-Institution
 - Grundstrukturen: Freiburger Korpus
 - Gespräche im Fernsehen: Talkshows, Diskussionen, Interviews
 - Stadtsprache: Mannheim
 - Schlichtungs- und Gerichtsverhandlungen
 - 2.2. Spracherwerb / Sprachentwicklung
 - Kindersprache: Saarbrücker Korpus

Datensicherung

Ein wichtiges Ziel von Korpus- und Dokumentationsprojekten ist die Datensicherung. Dies betrifft nicht nur die physische Datensicherung auf Festplatten, CDs und DVDs, sondern auch die Dateiformate.

Initiativen z.B.

- IDS Mannheim
- MPI für Psycholinguistik Nimwegen (The Language Archive)

XML: Extensible Markup Language

Allgemeines
Format zur
Erfassung und
Annotation von
hierarchisch
strukturierten
Daten.

Beispiel
(aus Eintrag XML,
Wikipedia):

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<verzeichnis>
  <titel>Wikipedia Städteverzeichnis</titel>
  <eintrag>
    <stichwort>Genf</stichwort>
    <eintragstext>Genf ist der Sitz von ...</eintragstext>
  </eintrag>
  <eintrag>
    <stichwort>Köln</stichwort>
    <eintragstext>Köln ist eine Stadt, die ...</eintragstext>
  </eintrag>
</verzeichnis>
```

- XML-Dateien sind hierarchisch aufgebaut.
- Die einzelnen Knoten haben Namen (Bezeichnungen in spitzen Klammern); das Ende von Abschnitten ist durch Namen mit vorgeschaltetem Schrägstrich markiert.
- XML-Dateien werden von spezifischen Programmen verarbeitet oder können durch allgemeine Textprogramme gelesen und dargestellt werden, auch durch Web-Browser (z.B. Firefox).

Die Text Encoding Initiative (TEI)

siehe <http://www.tei-c.org/index.xml> und den Wikipedia-Eintrag zu TEI:
http://de.wikipedia.org/wiki/Text_Encoding_Initiative

Annotation mit ELAN

Grundlegendes

- Werkzeug zur Audio- und Video-Annotation, also z.B. auch für Gesten
- Mehrfache Mediadateien möglich, z.B. Audio mit mehreren Mikrofonen, Video mit mehreren Kameras
- Bevorzugtes Format für The Language Archive, <http://www.mpi.nl/research/research-projects/the-language-archive>
-
- Beispiel: Sandzeichnung in Vanuatu
- Beispiel: Annotation Roland Berger, speak-1380-quest-11468.mov, .wav

Vorbereitende Arbeitsschritte

Vorbereiten der Aufnahme

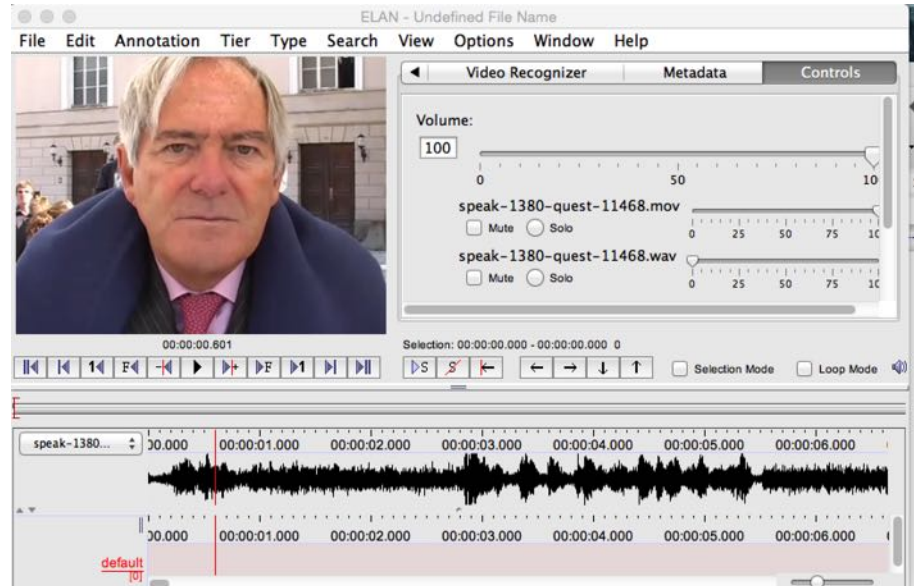
- Video-Aufnahme erzeugt eine Video-Datei (z.B. mov)
- Erzeugung einer Tondatei mit Audacity:
 - Einlesen der Video-Datei (es wird nur der Ton angezeigt)
 - Wenn nötig, Schneiden der Video-Datei
 - Wenn nötig, Erhöhung der Lautstärke / Anpassung der Amplitude
 - Export in eine wav-Datei (Audio-Datei)

Vorbereitung der ELAN-Dateien

- Aufruf von ELAN
- File → New → Add Media File
- Angabe der mov-Datei und der wav-Datei
- Beispiel: Roland Berger, speak-1380-quest-11468.mov, .wav

Gutes Informationssystem:

- Help
- Weiteres Informationsmaterial siehe ELAN-Seite



Einrichtung der Haupt-Annotationsebene

- Tier → Add new tier → „B“, Participant: „R. Berger“, „Add“ + Close
- Einrichtungen von weiteren Annotationsebenen, z.B. für weitere Partizipanten möglich, hier „A“ für „Announcer“
- Löschen von „default“ möglich
- Verschiedenen Ebenen können verschiedene Farben, Schriften zugeordnet werden

Alinierte Annotationen, z.B. für Übersetzungen

- Type → Add new linguistic type
- „translation“; Stereotype: „Symbolic Association“
- Add + Close
- Tier → Add new tier → „tr“, Parent: „B“, type: „translation“
- Add + Close
- Weitere Alinierungen möglich, z.B. auch zeitliche Unterteilung (Time division)



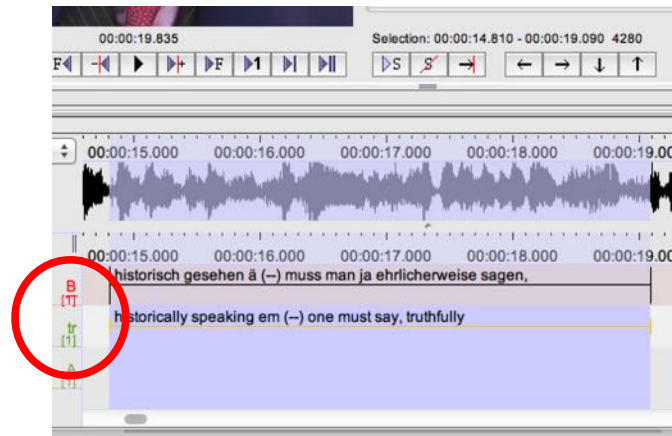
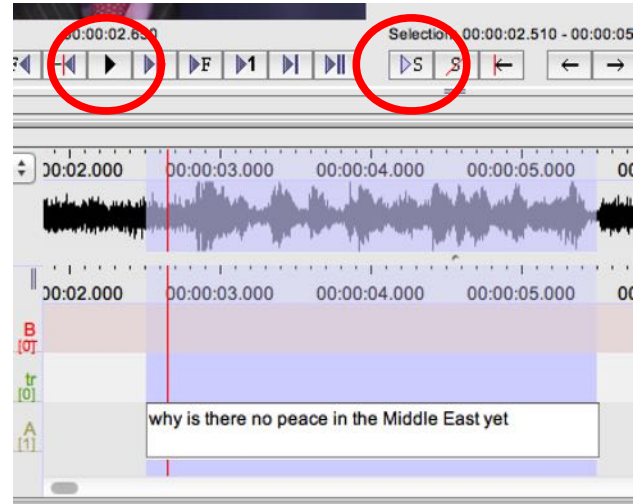
Annotation

Grundlegendes

- Hören mit ► (Rück-, Vorspulen etc.)
- Markieren eines Segments mithilfe der Amplitude der Audio-Datei (Verschiebung der Grenzen: Shift + Klick)
- Wiederholtes Hören des Segments mit S►
- Doppelklick in die entsprechende Annotationsebene, Transkription
- Sicherung mit Command + Return (!)
- Verschiebung der Grenzen möglich: Anklicken, ALT, mit Cursor an den Rand fahren

Annotation in Ebenen mit Parent-Tier:

- Annotation der Parent-Ebene
- Dann Annotation der zugeordneten Ebene, z.B. mit einer Übersetzung



Exportoptionen

Es gibt zahlreiche Exportmöglichkeiten für die Annotationen, in verschiedene Formate

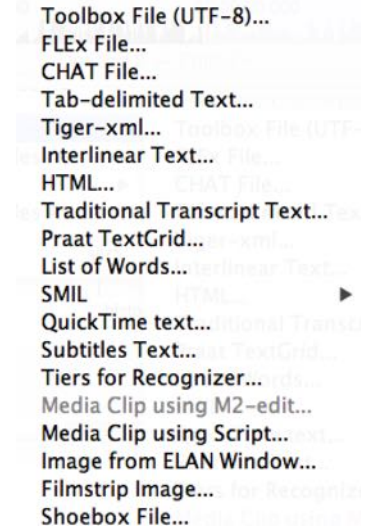
Beispiel: Traditionelle Annotation:

- A why is there no peace in the Middle East yet
(9.31)
- B historisch gesehen ä (--) muss man ja ehrlicherweise sagen,
(0.18)
- B dass der Konflikt im mittleren Ostern eigentlich entstanden ist durch eine politisch unmögliche Entscheidung

- **Bemerke:** Zeiten werden automatisch eingefügt;
Vorteilhaft: möglichst kleine Abschnitte annotieren!

Beispiel: Tab-delimited text, dann Tabellenkalkulationsprogramm oder Tabelle in LibreOffice, Sortieren nach Anfangszeit (dritte Spalte, in Millisekunden)

A	Announcer	2510	5500	2990	why is there no peace in the Middle East yet
B	R. Berger	14810	19090	4280	historisch gesehen ä (--) muss man ja ehrlicherweise sagen,
tr	R. Berger	14810	19090	4280	historically speaking em (--) one must say, truthfully
B	R. Berger	19272	36090	16818	dass der Konflikt im mittleren Ostern eigentlich entstanden ist durch eine politisch unmögliche Entscheidung
tr	R. Berger	19272	36090	16818	that the conflict in the Middle East, basically, originated in a politically impossible decision



Phonetische Software

Für die Untersuchung gesprochener Sprache gibt es ein frei zugängliches Software-Paket, das Analysen erlaubt, wie sie bis in die 90er Jahre hinein nur in gut ausgestatteten Phonetiklaboren möglich waren:

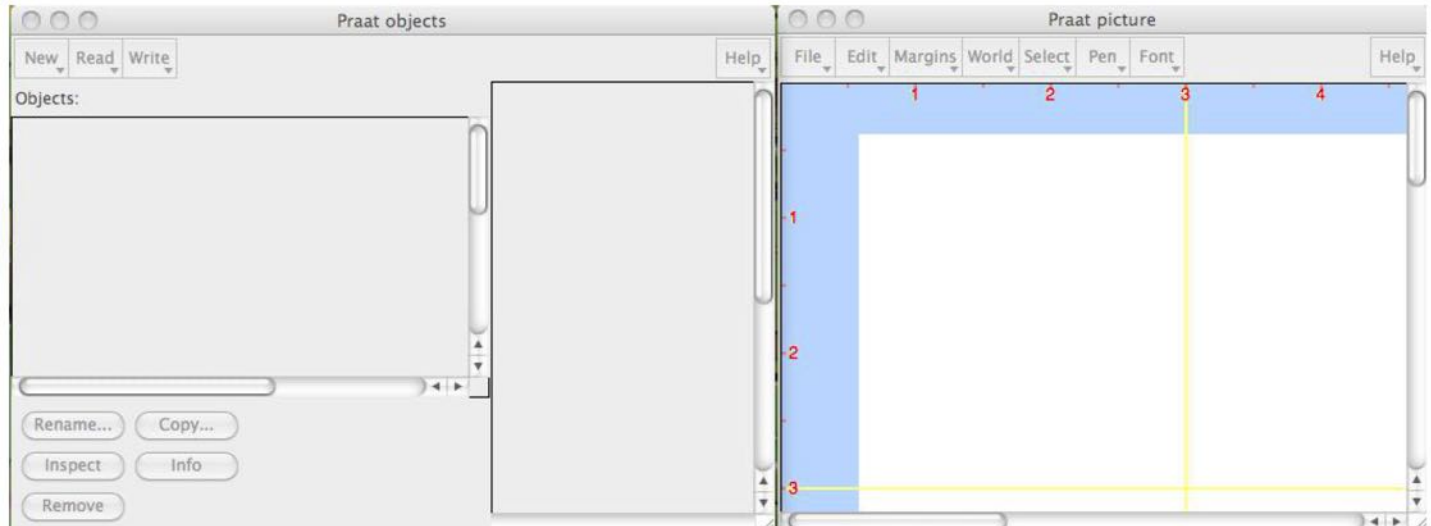
Praat, entwickelt von Paul Boursma und David Weenink

Die Software kann heruntergeladen werden von <http://www.fon.hum.uva.nl/praat>

Erste Schritte mit Praat

Beim Aufruf des Programms erscheinen zwei Fenster:
Praat Objects und Praat Pictures.

Es gibt eine detaillierte Hilfe-Funktion, zugleich eine ausführliche Einführung in die Verwendung von Praat.



Einlesen von akustischen Dateien

Akustische Daten können über verschiedene Dateiformate eingelesen werden, insbesondere das nicht-komprimierte Format WAV und das komprimierte Format MP3. (Das Format MP3 kann allerdings nicht erzeugt werden, da es urheberrechtlich geschützt ist).

Einlesen unter PRAAT OBJECTS: READ ⇒ READ FROM FILE.

Die Datei erscheint zur Weiterverarbeitung im PRAAT OBJECTS-Fenster als ein OBJECT.

Längere Dateien als LONGSOUNDFILE einlesen!

Aufnehmen von akustischen Daten

Akustische Daten können auch direkt über ein Mikrofon aufgenommen werden. (Aufnehmen unter PRAAT OBJECTS: NEW ⇒ RECORD MONO / STEREO SOUND).

In dem neuen Fenster SOUNDRECORDER kann man festlegen:

- Die SAMPLE RATE (in Hz), die festlegt, wie genau die akustische Welle abgetastet wird (wobei größere Abtastraten zu einem größeren Datenumfang führen).
- RECORD beginnt die Aufnahme.
- STOP beendet die Aufnahme.
- SAVE TO LIST speichert die Aufnahme zur Weiterverarbeitung.
- CLOSE schließt das Fenster SOUND RECORDING.

Die Datei erscheint dann als ein Objekt im PRAAT OBJECTS-Fenster.

Darstellung von akustischen Dateien

Durch Auswählen eines Objekts und EDIT erhält man das SOUND EDITOR-Fenster.

Hier sieht man zwei Felder:

- Oben eine grafische Darstellung der Schallwelle (schwarz).
Die Amplitude der Welle (vertikale Auslenkung) gibt die Lautstärkeintensität wieder.



- Unten (je nach Einstellung) ein Feld, in dem Analysen der Audio-Datei vorgenommen werden.
Die Einstellungen kann man in dem Fenster auswählen.

Die Größe des Fensters ist veränderbar (mit Cursors in Ecke rechts unten ziehen).

Man kann in die grafische Darstellung hineinzoomen und die Welle genau betrachten (comm-I/O).

Das folgende Beispiel ist eine Realisierung von [sa], insgesamt 0,2 Sekunden.

Beachte am Beginn die überlagerten Schwingungen (Rauschen), die für Frikative wie das [s] typisch sind, dann die für Vokale typischen periodischen Schwingungen.



Abspielen von akustischen Dateien

Zum Abspielen der gesamten Datei auf die unterste Leiste klicken.

Zur Aufnahme des sichtbaren Stückes auf die obere Leiste klicken.

Man kann den Cursor auf beliebige Abschnitte setzen, um ab dieser Stelle abzuspielen.

Man kann Abschnitte der Datei markieren (Cursor bei gedrückter Taste ziehen).

So kann man sich einzelne Abschnitte gezielt anhören (z.B. für Transkriptionen).

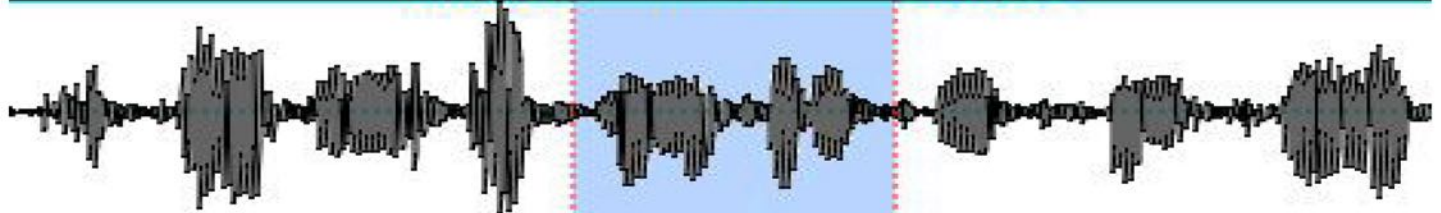
1. Sound PraatSatz

mant Pulses

2.808731

0.602137

3.410868



0.602137

Visible part 6.045147 seconds

Total duration 6.045147 seconds

Schneiden von akustischen Dateien

Man kann markierte Abschnitte in eigene Sound-Dateien abspeichern und auf diese Weise die Datei in kleinere Dateien zerlegen.

Zum Beispiel kann man einen Text in Dateien zerlegen, die jeweils einen einzigen Satz enthalten. (FILE ⇒ EXTRACT SELECTED SOUND, wobei man angeben soll, ob die Zeit von 0 an gezählt werden soll oder ob die Zeit der Gesamtdatei beibehalten werden soll).

Die neue Datei erscheint entweder als neues Objekt im Objekt-Fenster, oder man kann sie direkt als Datei abspeichern).

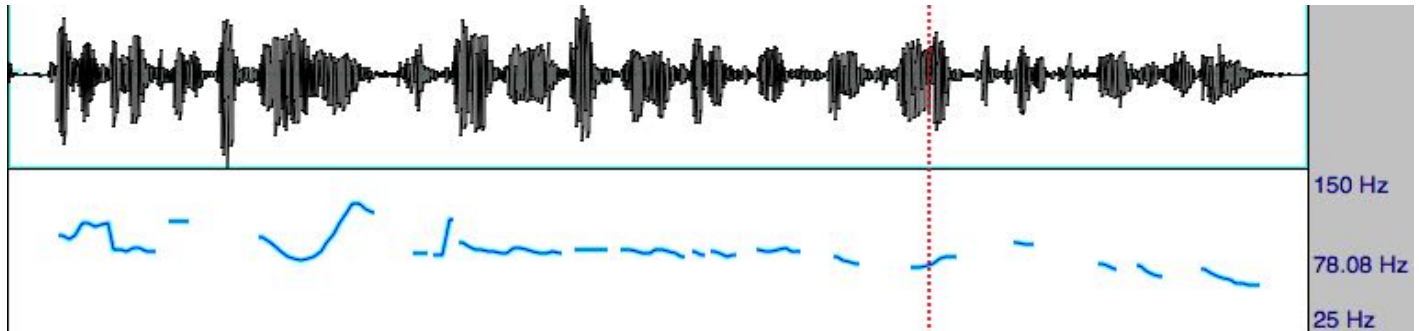
Man kann ferner durch EDIT ⇒ CUT/COPY/PASTE Teile von Sound-Dateien kopieren, ausschneiden und woanders wieder einsetzen, ganz ähnlich wie bei Textdateien.

Phonetische Analyse

Die akustischen Ereignisse können auf verschiedene Weise analysiert werden. Dies wird hier kurz dargestellt; auf die theoretischen Grundlagen kann hier allerdings nicht im Detail eingegangen werden (siehe hierzu Grundkurs).

Stimmton (Pitch)

PITCH ⇒ SHOW PITCH: Grundfrequenz der Stimmlippen (der Stimmton) (in Blau).

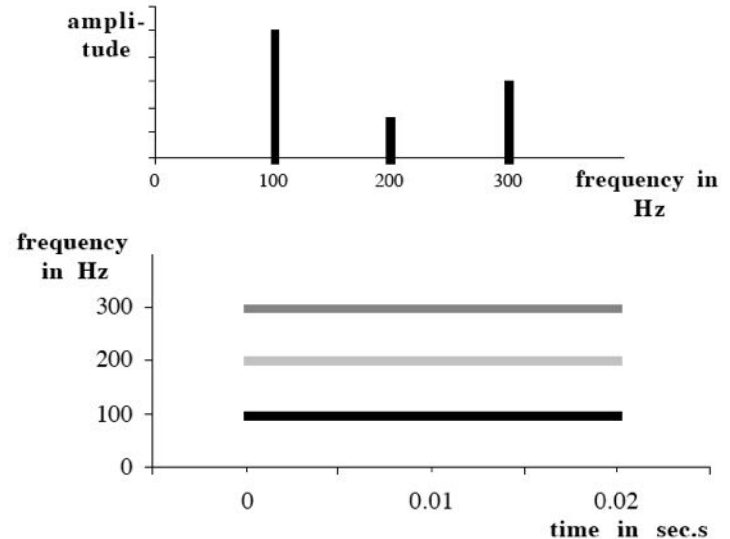
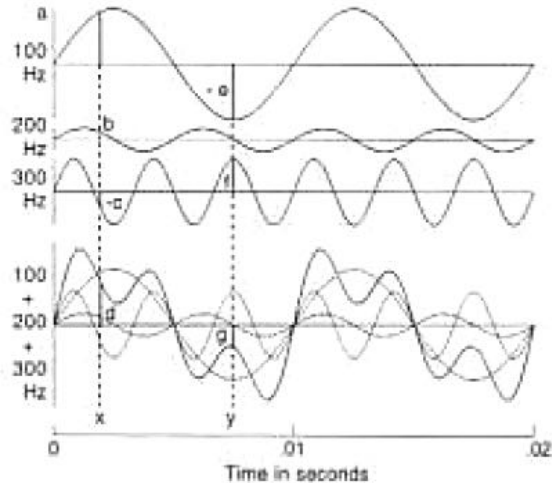


Rechts: Anzeige der Frequenz in Hertz an der Cursor-Position.

Stimmton-Analyse ist wichtig zur Darstellung der Intonation (z.B. Position und Art des Satzakzents, des satzfinalen Akzents usw).

Spektrogramm

Zerlegung einer Welle in Sinuskurven. (Satz von Fourier).



Darstellung nach Ladefoged 1996:

- Periodische Schwingung als Überlagerung von drei Sinuskurven;
- Darstellung durch Spektrum (Amplitude auf der y-Achse, Frequenz auf der y-Achse),
- Darstellung durch Spektrogramm (Frequenz: y-Achse, zeitlicher Verlauf: x-Achse, Amplitude durch Graustufe).

Die verschiedenen Vokale und die Übergänge zwischen Konsonant und Vokal zeichnen sich vor allem durch ihre Formanten aus, wobei hier nur vor allem ersten beiden Formanten (F1, F2) eine Rolle spielen.

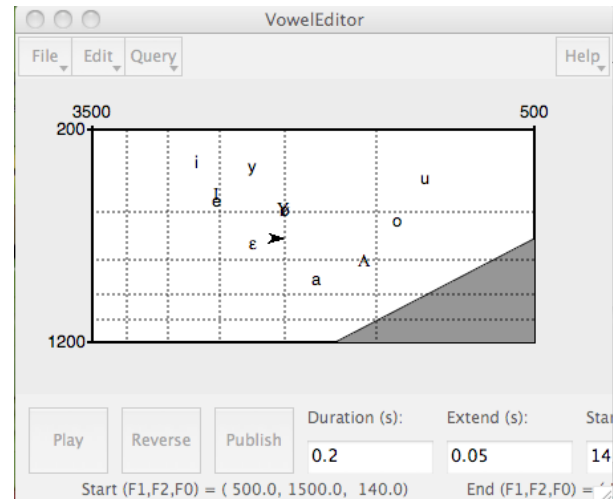
Der Stimmton selbst wird als F0 bezeichnet.

Geräusche bestehen aus einer Mischung von vielen Sinuskurven, d.h. sie haben ein kontinuierliches Spektrum.

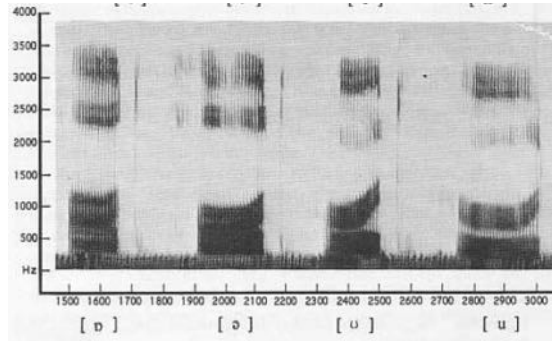
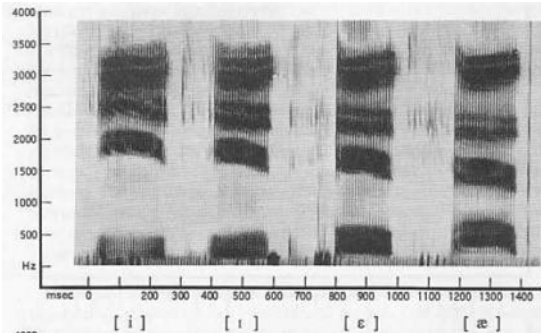
Formanten Vokale

- Je höher die Zungenposition, desto tiefer der erste Formant.
- Je vorderer die Zungenposition, desto höher der zweite Formant.

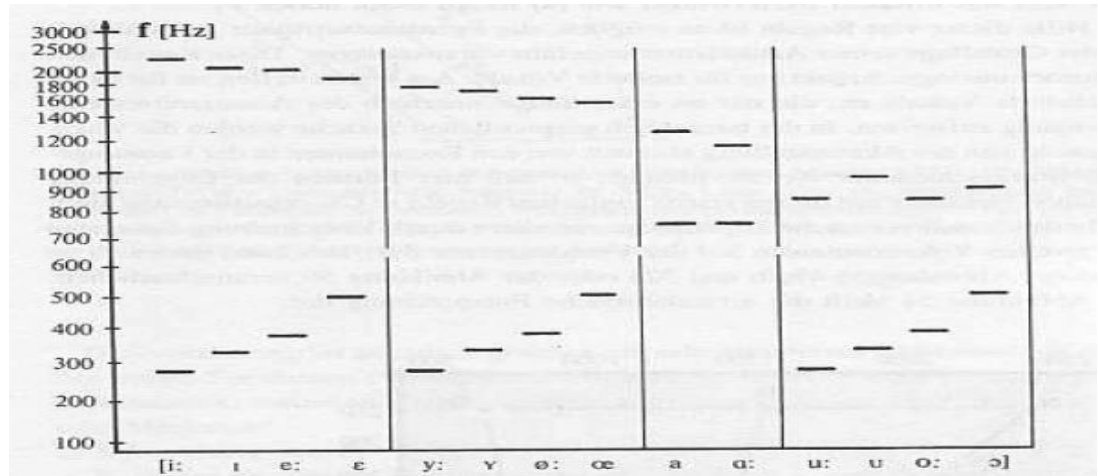
Mit Praat kann man diese Regeln anhand von künstlich erzeugten Vokalen überprüfen (siehe NEW ⇒ SOUND ⇒ CREATE SOUND FROM VOWEL EDITOR).



Vokale im Spektrogramm:



Beispiel: Formanten deutscher Vokale, Pétursson & Neppert 1996; hier: synthetisch erzeugte Vokale.



Konsonanten im Spektrogramm

Verschlusslaute werden im Spektrogramm nicht dargestellt, v.a. bei stimmlosen Verschlusslauten (weiße Bänder); die Natur von stimmlosen Konsonanten kann man nur aus ihrer Umgebung erkennen. Typischerweise hoher Geräuschanteil nach dem Verschluss bei stimmlosen Verschlusslauten.

Bilabiale Konsonanten: Der 2. und 3. Formant sind relativ tief.

Velar: Der zweite Formant ist relativ hoch.

Frikative: Geräusch, v.a. in hohen Frequenzen kontinuierliche Bänder.

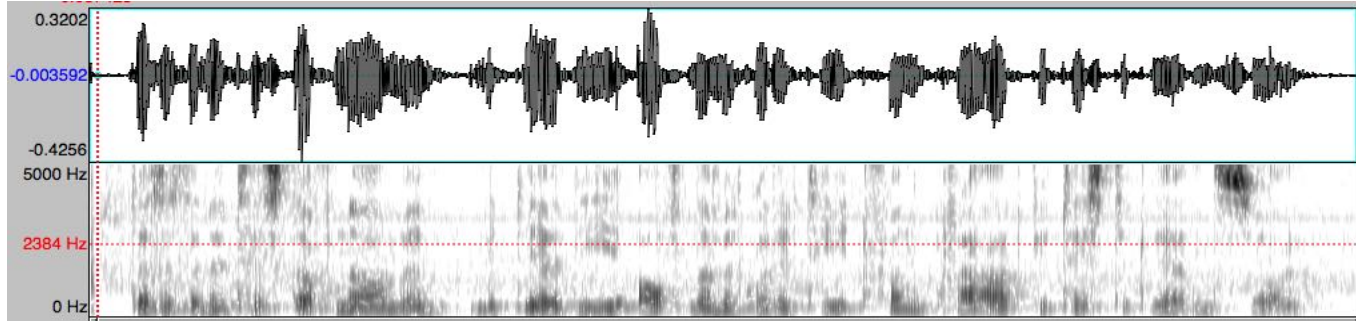
Nasale: Formanten ähnlich wie bei Vokalen, aber mit nasalen Formanten bei 250, 2500 und 3250 Hz.

Laterale: Formanten ähnlich wie bei Vokalen, aber mit Formanten um 250, 1200 und 2400 Hz; höhere Formanten sind reduziert.

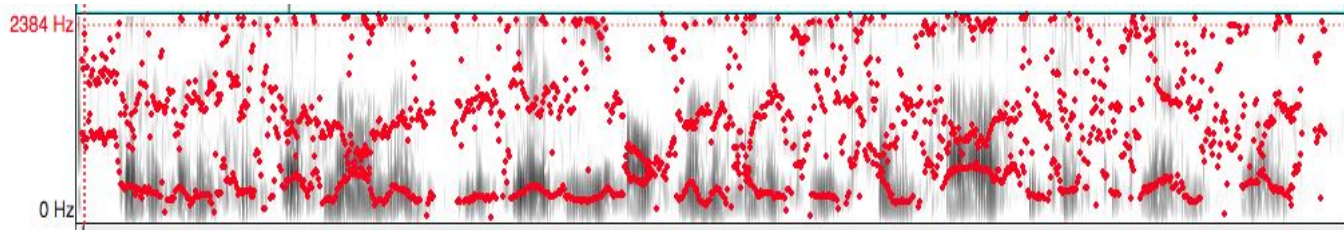
Approximanten: Formanten ähnlich wie bei Vokalen, die sich aber über den Verlauf des Vokals verändern.

Spektrogramm im Praat

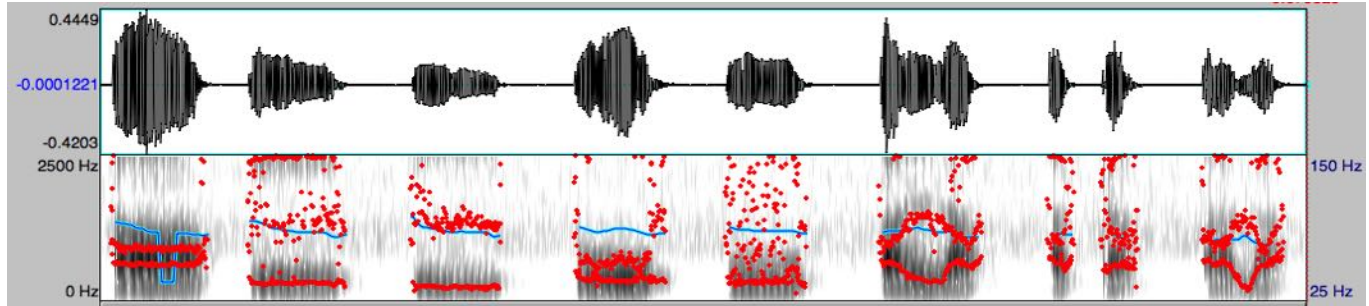
Dies kann über SPECTROGRAM \Rightarrow SHOW SPECTROGRAM gezeigt werden.



Standardmäßig wird das Spektrum bis 5000 Hz gezeigt, dies kann man jedoch auch verändern. Man kann auch die Formanten anzeigen lassen (FORMANTS \Rightarrow SHOW FORMANTS); hier ist das mit dem Spektrum für den Bereich bis 2500 Hz durchgeführt. Die Formanten werden rot angezeigt.

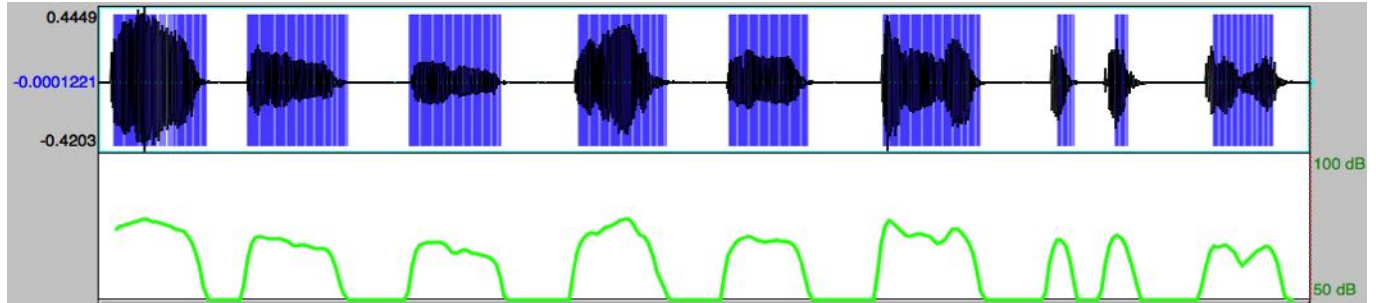


Hier eine Anzeige des Spektrogramms der fünf Vokale a, e, i, o, u, ferner [ala], [ata], [ada]. Es ist zugleich auch die F0-Frequenz (Pitch) eingezeichnet, die weitgehend gleich bleibt.



Intensität, Pulses

Weitere Anzeigemöglichkeiten: Die Intensität der Schallwelle und die “Pulses”, d.h. die Vibrationen der Stimmlippe.



Annotation

Praat enthält einen Partitur-Editor zur Annotation des akustischen Materials.

Anlegen einer Annotation

Wenn eine Datei im Objekt-Fenster angewählt ist: ANNOTATE ⇒ TO TEXT GRID.

Man kann die Zahl und die Bezeichnungen der Ebenen (“tiers”) auswählen.

Hier wählen wir drei Ebenen: Phoneme, Wörter, Akzente.

Es erscheint ein neues Objekt im Objekt-Feld. Wir wählen sowohl das originale Objekt als auch das zugehörige Grid-Objekt aus und gehen auf Edit. Wir sehen sowohl das Sound-Objekt als auch die Annotationsebenen, die wir verwenden können. Diese kann man durch Klicks annotieren.

Aufgabe: Annotieren der orthographischen Wörter (2. Ebene).

Es empfiehlt sich, über VIEW ⇒ ZOOM IN den Ausschnitt zu vergrößern.

Einführung und Entfernen von Grenzen

Man kann Grenzen in den Ebenen einführen, indem man den Cursor auf die intendierte Stelle setzt. Es erscheinen Grenzen in den Ebenen mit kleinen Kreisen; wenn man in den kleinen Kreis klickt, wird eine Grenze eingeführt.

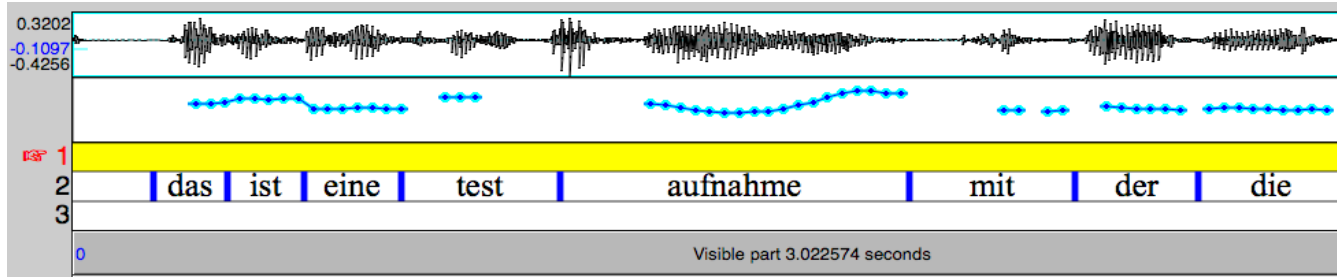
Grenzen können wieder entfernt werden: BOUNDARY ⇒ REMOVE.

Verschieben von Grenzen durch Klicken mit Taste + Ziehen.

Annotation: Wörter

Wir wollen die Wörter unseres Ton-Beispiels annotieren. Dabei gehen wir wie folgt vor:

- Wir gehen auf den Beginn des ersten Wortes und führen dort eine Grenze in der Ebene “Wörter” ein.
- Wir markieren den Abschnitt bis zum Ende des ersten Wortes. Hierfür ist es günstig, den Endpunkt des markierten Abschnitts mit der Tastatur zu markieren:
UMSCHALTEN + AUF/AB bzw. CTRL + AUF/AB (siehe Cursorbewegungen unter SELECT).
Abschnitte zwischen Äußerungen sind an der Schallwelle erkennbar.
- Wenn wir das Ende des Abschnitts erreicht haben, führen wir eine neue Grenze ein.
- Dies wiederholen wir, bis die ganze Datei durch-analysiert ist.



Speichern der Annotationsdatei:: WRITE ⇒ WRITE TO SHORT TEXT FILE.

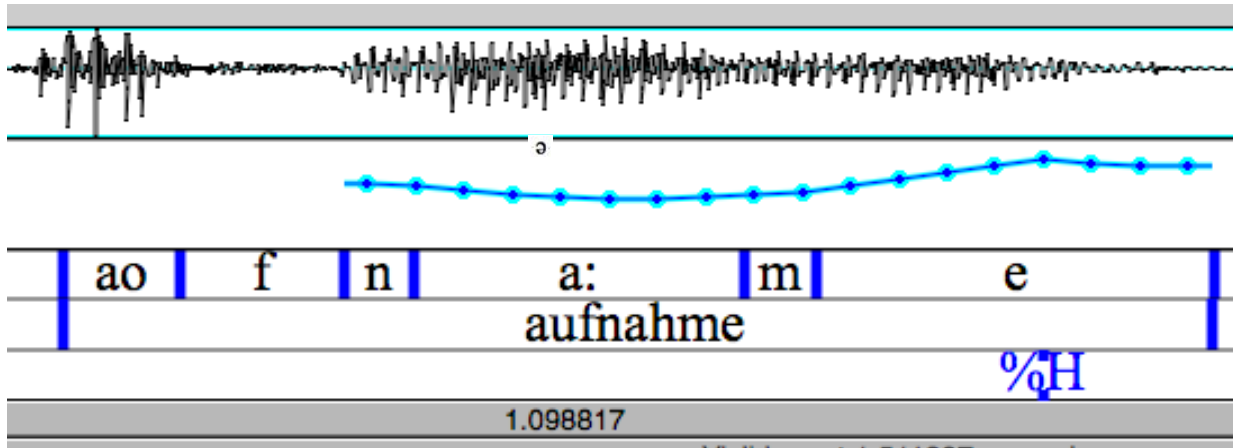
Klicken auf Intervall zeigt die Länge; Werkzeug zur Messung von Zeitdauern.

Die entstandene Annotationsdatei muss aus dem Objekts-Fenster heraus gespeichert werden:
WRITE ⇒ WRITE TO SHORT TEXT FILE.

Wenn man auf ein Intervall klickt, sieht man die Länge dieses Intervalls, man kann also Längenmessungen von akustischen Einheiten vornehmen.

Annotation: Laute, Akzente

Wie Wörter, so kann man auch einzelne Phoneme annotieren. Auch die Annotation von Akzenten ist möglich (z.B. %H für hohen Grenzton); hierfür eignen sich punktuelle Annotationsebenen. Die Verwendung phonetischer Umschrift ist möglich

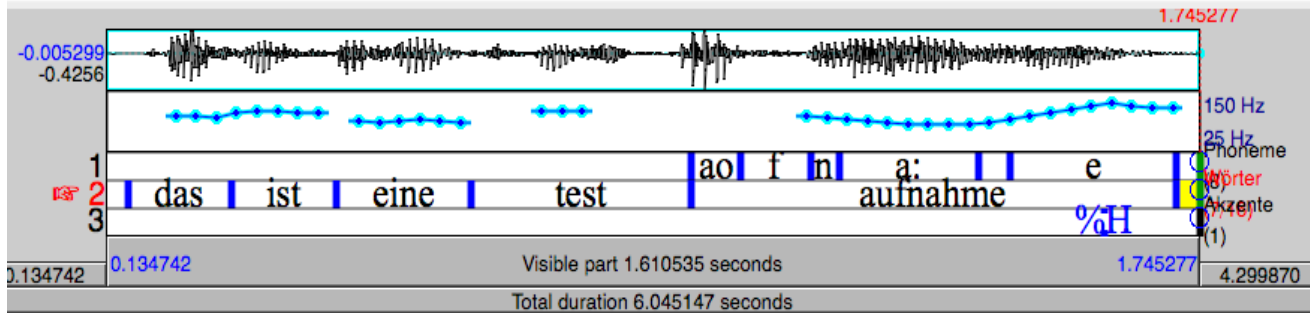


Weiterverarbeiten von Praat-Objekten

Die obigen Beispiele aus Praat wurden direkt vom Bildschirm kopiert.

Ein besseres Verfahren zum Ausdrucken oder zur Einfügung in Textverarbeitungsdateien ist es, Objekte über das Picture-Fenster zu senden.

Beispiel: Markieren eines Abschnitts im Editor; VIEW ⇒ ZOOM TO SELECTION

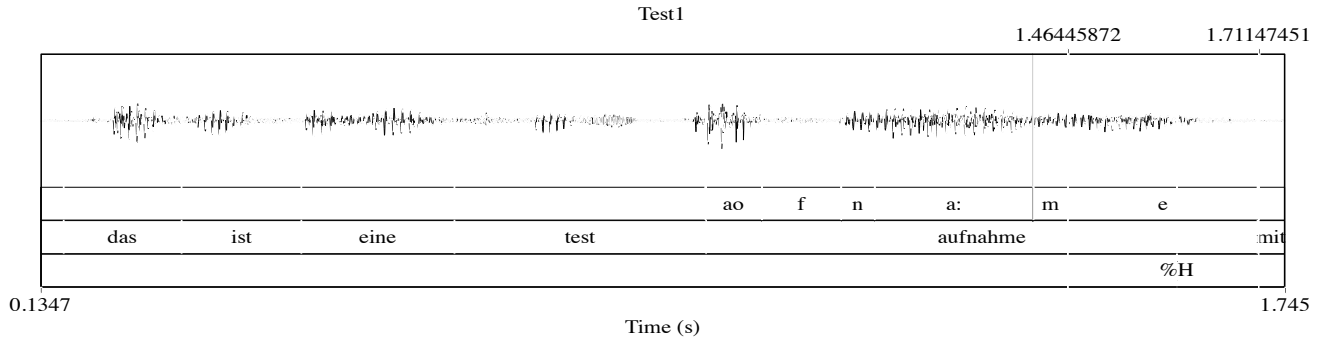


Im Praat Picture Fenster wird ein Feld mit der intendierten Größe gezogen.

Im Editor-Fenster wird angewählt: EDIT ⇒ DRAW VISIBLE SOUND AND TEXT GRID

Man sieht dann die Schallwelle und die Transkription.

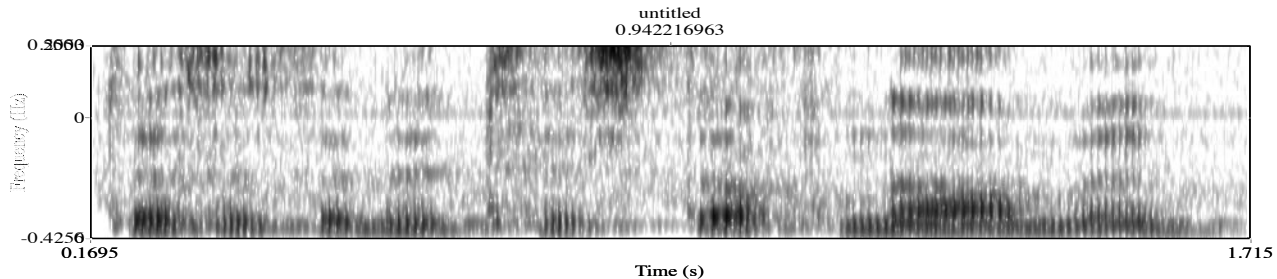
Diese können z.B. über das Clipboard in eine Textdatei eingefügt werden:



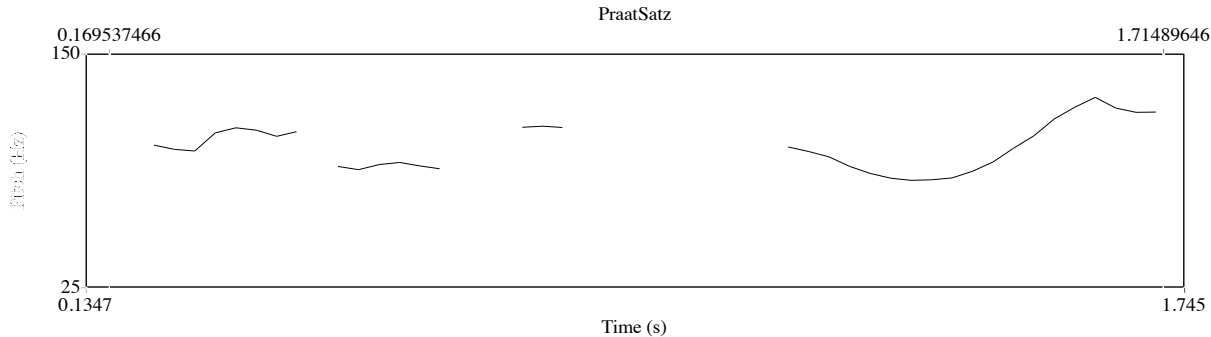
Darstellung von Spektrogrammen und Pitchverläufen in eigenen Fenstern:.

Erster Schritt: In EDITOR-Fenster: SPECTROGRAM ⇒ EXTRACT VISIBLE SPECTROGRAM (hier muss man sicherstellen, dass das sichtbare Spektrogramm mit der Annotation übereinstimmt).

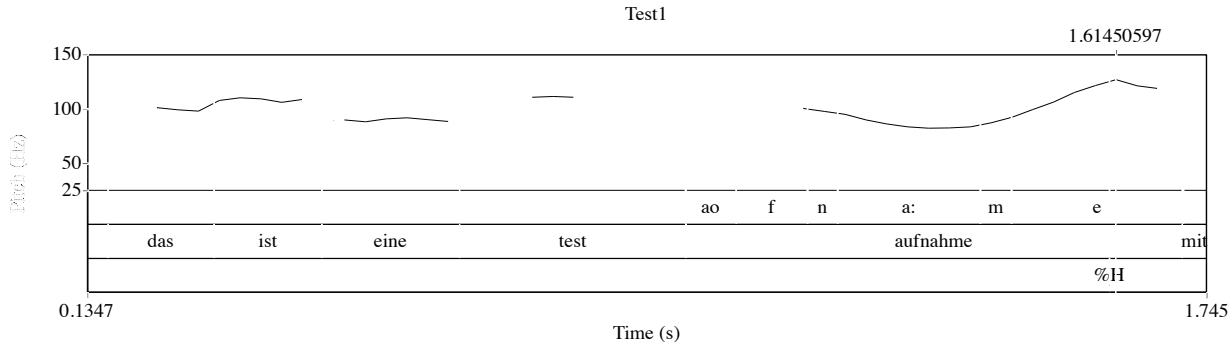
Zweiter Schritt: Anklicken des neu entstandenen Objekts im Objekt-Fenster; PAINT



Ähnlich kann man auch mit der Grundfrequenz verfahren:



In dem EDITOR-Fenster kann man jedoch auch unter PITCH die Option DRAW VISIBLE PITCH CONTOUR AND TEXT GRID anwählen; man erhält dann folgende Darstellung:



Empirische Untersuchungen

vgl. Ruth Albert, Nicole Marx (2010),
Empirisches Arbeiten in Linguistik und Sprachlehrforschung,
Tübingen: Narr.



Daten: Beobachtung, Befragung, Experiment

Die Beobachtung

Daten, die unabhängig von der wissenschaftlichen Fragestellung entstanden sind und die nach bestimmten Kriterien gesammelt werden.

Beispiel: Linguistische Korpora.

Der Beobachter sollte so wenig wie möglich Einfluss auf das beobachtete Material nehmen. Dies ist nicht immer möglich: **Beobachterparadox**.

Beispiel: Bei Audio- und Videoaufnahmen verhalten sich Sprecher anders, weil sie sich beobachtet wissen (Befangenheit, größere Deutlichkeit, Rücksicht auf Image).

Minimieren des Einflusses des Beobachters, z.B., indem man die beobachteten Personen an den Aufnahmeprozess gewöhnt.

Aber: Verdeckte Aufnahmen sind aus ethischen und rechtlichen Gründen nicht zulässig.

Die Befragung

Explizite Befragungen von Personen über das eigene sprachliche Verhalten oder über die Einschätzung des sprachlichen Verhaltens von anderen.

Wir können Personen befragen,

- ob sie bestimmte Sätze grammatisch wohlgeformt finden,
- ob ein Satz in einer bestimmten Situation wahr, falsch oder unangemessen ist,
- oder wie sie den Sprecher eines Satzes hinsichtlich Herkunft oder sozialen Status einschätzen.

Problem: Falsche Selbsteinschätzung. Personen können – auch ungewollt – falsche Auskunft geben.

Gründe:

- Viele Phänomene sind der Selbstbeobachtung nicht zugänglich, z.B. bei der Sprachverarbeitung.
Beispiel: Überzeugung, dass *Rad / Rat* oder *Lerche / Lärche* verschieden ausgesprochen werden.
- Sprecher wollen ein sozial angemessenes Bild von sich geben und vermeiden bestimmte Antworten wegen der Gesichtswahrung.
Beispiel: *Die Konjunktur springt nicht an, weil die Leute kaufen nichts.*
“Man muss sagen: ‘weil die Leute nichts kaufen’. So wie hier würde ich das nie sagen, weil das ist ungrammatisch!“

Abhilfe:

- anonymisierte Befragungen;
- damit verliert man aber eventuell wichtige Daten über die befragte Person.

Das Experiment

Im sprachwissenschaftlichen Experiment

- beobachtet man das sprachliche Verhalten oder andere Eigenschaften von Versuchspersonen,
- die bestimmten, sorgfältig kontrollierten Umständen ausgesetzt sind.

Diese Umstände werden so ausgewählt oder konstruiert, dass die Reaktion von Versuchspersonen zwischen zwei Hypothesen H und H' unterscheidet:

- Nach der Hypothese H erwarten wir ein beobachtbares Verhalten B,
- nach der Hypothese H' erwarten wir ein davon verschiedenes beobachtbares Verhalten B';
- wenn wir nun B beobachten, sollte Hypothese H angenommen und H' verworfen werden.

Dieses Bild wird aber kompliziert durch:

- Es ist oft schwierig, vorauszusagen, welches beobachtbare Verhalten aus einer Hypothese folgt, u.a. weil es von verschiedenen Faktoren bestimmt wird.
- Versuchspersonen können das Verhalten B bzw. B' in unterschiedlichem Grade zeigen, z.B. 80% der VPn zeigen B, 20% zeigen B' – wie soll man dann entscheiden?

Wir werden uns der letzten Frage im Statistik-Teil widmen.

Bemerke:

- Hypothesenbildungen gibt es natürlich auch bei der **Befragung**.
- Bei der **Beobachtung** (z.B. Korpuslinguistik) gibt es sie ebenfalls, allerdings ist es oft nur schwer möglich, in der Masse der beobachteten Phänomene solche herauszufinden, die zwischen zwei Hypothesen entscheiden.

Gütekriterien

Empirische Untersuchungen sind in der Regel hypothesengeleitet (wobei sich manchmal die Hypothesen erst nach der Untersuchung ergeben und dann eine neue Untersuchung erfordern – vgl. hierzu später den Statistik-Teil).

Eine entscheidende Frage:

Wie gut stützt / widerlegt eine empirische Untersuchung die Hypothese?

Es gibt hier unterschiedliche **Gütekriterien**:

- Verlässlichkeit (Reliabilität)
- Geltungsbereich
- Gültigkeit (Validität)

Verlässlichkeit (Reliabilität)

Frage: Wie gut misst das Messverfahren dasjenige, was es misst?

Ein Messverfahren gilt als verlässlich, wenn Wiederholungen dasselbe Resultat liefern.

Daraus ergeben sich auch die Möglichkeiten, die Verlässlichkeit zu testen:

- **Testwiederholung.** Derselbe Test wird noch einmal durchgeführt.
Der erste Test darf allerdings den zweiten nicht beeinflussen (Lerneffekt!)
- **Paralleltest:** Man untersucht dieselben Versuchspersonen mit einer zweiten Version des Tests.
Auch hier müssen Lerneffekte ausgeschaltet werden.
- **Testhalbierung** (“split half”): Man halbiert die Ergebnisse eines Tests nach dem Zufallsprinzip und betrachtet die beiden Hälften als unabhängige Tests.
Wichtig: Man sollte nicht z.B. die zeitlich erste Hälfte und die zweite Hälfte wählen, da es während des Tests selbst Lerneffekte geben kann.
- **Konsistenzprüfung.** Der Test wird in die Einzelaufgaben zerlegt und die Konsistenz der Ergebnisse wird statistisch berechnet.

Geltungsbereich

Frage: Sind die anhand einer kleinen Gruppe von Personen gewonnenen Daten für die Grundgesamtheit relevant?

Was ist überhaupt die Grundgesamtheit?

- Die Menge aller Deutschsprecher?
Dialektsprecher, Norddeutsche, Süddeutsche?
Auch diejenigen, die Deutsch als Zweitsprache sprechen?
Kinder? Sehr alte Personen? Bildungshintergrund?

Die Frage stellt sich auch bei Korpusuntersuchungen: Wie aussagekräftig sind Beobachtungen an einem bestimmten Korpus für die Sprachverwendung allgemein?

Untersuchungsverfahren:

- Definition der intendierten Grundgesamtheit
- Zufällige Auswahl einer Stichprobe aus der Grundgesamtheit
- Wiederholte Auswahl von Stichproben und Tests.

Gültigkeit (Validität)

Sprachliche Fähigkeiten sind in der Regel nicht direkt messbar:

Wir messen Phänomene, die mit den sprachlichen Fähigkeiten zusammenhängen, z.B. die Häufigkeit der Verwendung syntaktischer Konstruktionen oder die Zeit, in der ein Wort erkannt wird.

Die **Gültigkeit** befasst sich mit der Frage:

Wie weit korreliert das, was wir messen, mit dem, was wir eigentlich untersuchen wollen?

Dies ist das schwierigste Gütekriterium; empirische Untersuchungen sind hier am ehesten angreifbar:

Es hängt oft von theoretischen Voraussetzungen ab, ob die Resultate eines Messverfahrens mit den Phänomenen, die wir untersuchen wollen, tatsächlich in Verbindung stehen.

Beispiel: Intelligenztests – „Intelligenz ist, was Intelligenztests messen.“

Ein Hinweis auf Gültigkeit ist, dass unterschiedliche Testverfahren für das zugrundeliegende Phänomen, die möglichst auch theoretisch unterschiedlich motiviert sind, ähnliche Ergebnisse liefern.

Beobachtung

Grundsätzliches

Die Beobachtung von nicht elizitierten Daten ist besonders aufwändig:

- Wir wissen oft nicht, ob sich in den beobachtbaren Daten überhaupt hypothesen-relevante Daten befinden, und diese sind oft schwer zu identifizieren.
- Beobachtungen müssen dokumentiert werden; dies wirft logistische und legale Probleme auf (z.B. Tonaufnahmen).
Die Beobachtungen interferieren oft mit anderen nötigen Aufgaben (Beispiel: Spracherwerb) oder laufen Gefahr, Persönlichkeitsrechte zu verletzen.
- Beobachtungen sind oft unvollständig und fehlerhaft (Beispiel: Flugzeugunfälle und Aufnahmen in der *black box*).
- Offene Beobachtungen führen zur Beeinflussung des Beobachteten durch den Beobachter.

Das letzte Problem kann z.T. umgangen werden durch die Beobachtung von öffentlich zugänglichen Materialien:
Fernseh-Interviews, Talkshows (Youtube!), Chats, veröffentlichte Texte.

Datenklassifikation

Bei der Beobachtung (wie auch im Experiment) müssen stets Phänomene klassifiziert werden.

Beispiele:

- Sprachbegleitende Gesten,
- Sprechakt-Typen in Konversationen,
- inhaltliche Beziehungen zwischen Textteilen.

Hierbei kann es zu unterschiedlichen Klassifikationen zwischen verschiedenen Bearbeitern kommen (“inter-rater reliability”).

Es ist wichtig, die Kriterien möglichst klar festzulegen, Beispiele und Gegenbeispiele zu geben und insbesondere auf die Klassifikationen von Phänomenen einzugehen, die “zwischen” den Kategorien liegen.

Beispiel: Sprechakt-Inventar für eine Klassifikation von Sprechakten

Diegritz / Fürst (1999), Klassifikation von Sprechakten im Schulunterricht.

Es wird ein Inventar von ca. 60 Sprechakten aufgestellt, nach denen Kommunikationsakte im Schulunterricht klassifiziert werden.

Es stellt sich dabei das Problem der richtigen Zuordnung von Sprechakten.

Dies wird durch eine detaillierte Beschreibung gelöst.

Hier zwei Beispiele:

- Sprechakttyp Rat
- Sprechakttyp Überlegung anzeigen

Sprechakttyp: Rat

Handlungsvoraussetzung: S hat zur Kenntnis genommen, dass A bei der Lösung eines Problems Hilfe braucht. A hat S um RAT gefragt oder nicht.

Beziehung zwischen S und A: Rollenstruktur komplementär, S ist in der höheren Position, Sprechhandlung nicht reversibel. Umgangsqualität: förderlich, hilfreich.

Paraphrasierung: S versucht A bei der Lösung eines Problems indirekt zu helfen, indem er A Information übermittelt, von denen er annimmt, dass sie zur Lösung des Problems beitragen.

Mögliche (typische) perlokutive Effekte: A nimmt RAT an; A fordert weitere Information von S, A lehnt RAT ab.

Beispiele:

Ich würde mal im Duden nachschlagen.

Schaut doch mal, wie es Gruppe drei gemacht hat.

Mögliche Verwechslungsgefahr: Ähnlichkeit mit VORSCHLAG. Unterschied: Prinzipiell bleibt beim Vorschlag offen, ob das empfohlene Verhalten für die betreffenden Personen vorteilhaft ist oder nicht. Bei VORSCHLAG ist die Rollenstruktur symmetrisch und reversibel. Das bedeutet, dass die Lehrkraft den Sprechakttype VORSCHLAG gar nicht ausführen kann. Ähnlichkeiten mit ANWEIS; Unterschied: ANWEIS ist für A verbindlich.

Explizit performative Formel: *Ich rate dir / euch (hiermit), ...*

Sprechakttyp: Überlegung anzeigen

Handlungsvoraussetzung: Jemand hat S eine Frage gestellt.

Beziehung S/A: Rollenstruktur symmetrisch, Sprechhandlung reversibel, Umgangsqualität sachlich-neutral.

Paraphrasierung: S versucht eine gestellte Frage zu beantworten, kann aber nicht sofort das notwendige Wissen abrufen. S signalisiert A, dass er sich bemüht, die geforderte Information zu finden. S zeigt A dadurch auch an, dass er nicht bei seinem mentalen Akt gestört werden möchte und dass A abwarten soll, bis der Denkprozess abgeschlossen ist.

Mögliche (typische) perlokutive Effekte: A wartet ab, bis S die Antwort gefunden hat; A wird ungeduldig und fragt erneut nach.

Beispiele:

Ähm...

Also...

Sp wiederholt Frage

Deliberatives Wegblicken

Mögliche Verwechslungsgefahr mit HÖRERRÜCKMELD und GLIEDERUNG SIGNALISIER.

Explizit performative Formel: *Ich zeige dir/euch (hiermit) an, dass ich überlege.*

Befragung

Grundsätzliches

In der Befragung kann man gezielt Daten einholen, diese Daten sind allerdings “gefiltert” in dem Sinne, dass sie die sprachlich dargestellte Meinung von Personen darstellen.

Unterscheidung **mündliche** vs. **schriftliche** Befragungen.

Insbesondere bei mündlichen Befragungen ist mit **Interviewer-Effekten** zu rechnen:

Die Persönlichkeit, der Fragestil des Interviewers hat Einfluss auf das Resultat der Befragung.

Gegenmittel: Gelenkte Interviews, bei denen die Fragen vorher genau festgelegt werden.

Auch die Formulierung von schriftlichen Fragen kann unabsichtliche oder beabsichtigte Wirkungen auf die Art der Beantwortung von Fragen haben.

Schriftliche Befragungen haben den Vorteil, dass man sie nicht in Anwesenheit der befragten Person durchführen muss. Fragebögen können per Post oder per E-Mail verschickt werden, oder können über einen Web-Browser ausgefüllt werden.

Allerdings ergibt sich hier das Problem der **Selbst-Selektion**: In der Regel werden nicht alle Personen, die dazu gebeten werden, den Fragebogen auch ausfüllen.

Das hat Einfluss auf die Stichprobe, die man zieht, und kann daher das Ergebnis der Befragung verfälschen, indem sie ihren Geltungsbereich einschränkt.

Die Stichprobe

Eine Stichprobe soll Aussagen über die Grundgesamtheit liefern.

Damit sie dies tun kann, muss sie bestimmte Anforderungen erfüllen.

Bei der **Auswahl** muss man vermeiden, dass bestimmte Arten von Individuen der Grundgesamtheit überrepräsentiert sind.

Möglichkeiten:

➤ **Zufalls-Stichprobe.**

Oft nicht leicht durchzuführen – nach welchen Kriterien soll ausgewählt werden?

Z.B. Auswahl nach Telefonnummern im Telefonbuch erfasst keine Personen ohne Telefonnummer oder mit nicht-veröffentlichter Telefonnummer.

➤ **Quotenverfahren.**

Man geht von bestimmten bekannten Eigenschaften der Grundgesamtheit aus (Zahl der Männer vs. Frauen, Altersklassen, Bildungsgrad der Bevölkerung usw.) und versucht, diese in der Stichprobe zu repräsentieren.

In Abschlussarbeiten darf man die “Kosten” der Auswahl der Stichprobe nicht unterschätzen:

➤ In der Regel wird es sich um eine Konvenienz-Stichprobe handeln:

Man fragt Freunde, Familie, Mit-Studierende.

➤ Das sollte in der Arbeit aber angesprochen und problematisiert werden, und der Geltungsbereich sollte entsprechend eingeschränkt werden.

Größe der Stichprobe

Um Aussagen über die Grundgesamtheit zu liefern, muss eine Stichprobe genügend groß sein.

Wie groß – das hängt von der Fragestellung und von den verwendeten statistischen Tests ab.

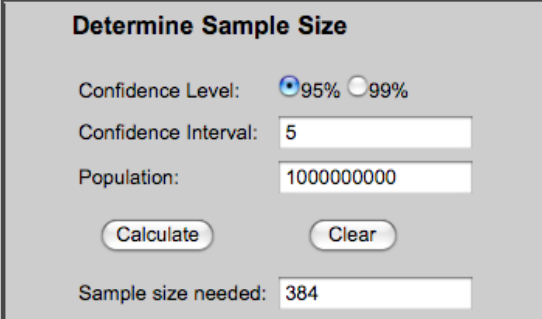
Wenn die Daten relativ klar ausgeprägt sind genügt eine kleinere Stichprobe
(Beispiel: Verwendung des Präteritums vs. Perfekts als Erzähltempus im nord- vs. süddeutschen Raum)

Die notwendige Größe von Stichproben kann errechnet werden durch einen **Sample Size Calculator**, z.B. <http://www.surveysystem.com/sscalc.htm>:

Beispiel:

- Sie wollen herausfinden, von welchem Anteil von Sprechern ein Satz als grammatisch wohlgeformt angesehen wird.
- Sie wollen diese Aussage zu 95% Sicherheit treffen (das ist das **Konfidenzniveau**),
- und es genügt Ihnen, wenn sie diesen Anteil zu $\pm 5\%$ kennen (das **Konfidenzintervall**), d.h. 80-85% der Sprecher finden den Satz grammatisch.
- Ergebnis: Sie brauchen ein Sample von 384 Personen (wenn wir eine Sprecherzahl von 100 Millionen annehmen, das spielt aber ab einer bestimmten Größe keine große Rolle). Die Personen des Samples sollten natürlich zufällig ausgewählt sein.

In Abschlussarbeiten wird selten (bei Befragungen) eine Stichprobengröße erreicht, mit der man statistisch signifikante Resultate erwarten kann.



The image shows a web-based calculator titled "Determine Sample Size". It has a light gray background and a dark border. At the top, the title "Determine Sample Size" is in bold black text. Below the title, there are four rows of input fields and two buttons. The first row is "Confidence Level:" with two radio buttons: "95%" (which is selected) and "99%". The second row is "Confidence Interval:" with a text input field containing the number "5". The third row is "Population:" with a text input field containing "1000000000". Below these fields are two buttons: "Calculate" and "Clear". At the bottom, there is a row for "Sample size needed:" with a text input field containing "384".

Befragungsarten

Es ist zu unterscheiden zwischen

➤ **Offenen Interviews:**

- Vorteil: Der Interviewer kann flexibel reagieren.
Phänomene, an die gar nicht gedacht wurden, können ins Blickfeld rücken.
- Nachteil: Kategorisierung der Auswertung schwierig und manchmal nicht zuverlässig.
Diese Form eignet sich eher für qualitative als für quantitative Forschung.

➤ **Festgelegten Fragen**, oft in einem Fragebogen.

- Vorteil: Bessere Auswertungsmöglichkeit,
Verminderung des Interviewer-Effekts.
- Nachteil: Es werden nur bestimmte Reaktionsmöglichkeiten erfasst.

Manchmal ist es günstig, erst offene Interviews durchzuführen und daraufhin einen Fragebogen zu entwickeln.

Fragebogen

Grundsätzliche technische Möglichkeiten:

- Papiergebundener Fragebogen („paper-and-pencil“);
Antworten müssen i.d.R. nicht sequentiell gegeben werden
- Digitaler Fragebogen, auch über das Internet;
Antworten i.d.R. sequentiell festgelegt,
erlaubt Fragen, die je nach vorheriger Antwort verzweigen,
erlaubt raschere Auswertung,
erlaubt z.T. auch Beobachtung des Zeitverlaufs.

Möglichkeiten für die Erstellung von gebührenfreien Internet-Fragebogen (survey software):

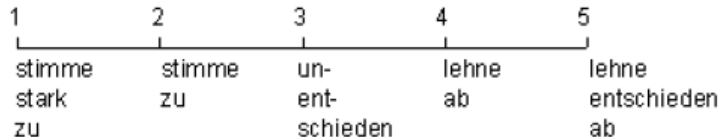
- <http://www.unipark.com/de/>
- www.fluidsurveys.com
- www.surveymonkeys.com
- www.freeonlinesurveys.com

Fragebogen (Fortsetzung)

Häufig bestehen Fragebögen aus Fragen oder Aussagen, zu denen Zustimmung oder Ablehnungsreaktionen verlangt werden.

Hier gibt es verschiedene Antwortmöglichkeiten:

- Ja/Nein: Nur zwei Alternativen (dichotome Fragen)
- Mehrere Alternativen, mit verbalen oder numerischen Antworten:
 - Ja/Nein/Weiß nicht.
 - Bei Grammatikbeurteilung: richtig – möglich – seltsam – kaum möglich – falsch
 - Bei Häufigkeitsbeurteilungen: immer / häufig / gelegentlich / selten / nie
 - sehr einverstanden o--o--o--o--o--o überhaupt nicht einverstanden (sog. Likert-Skala)



Bei abgestuften Antworten: drei, fünf oder sieben Kategorien anbieten (eine ungerade Zahl), es sei denn, man will eine positive / negative Antwort erzwingen.

Alternativ hierzu ist das Verfahren der Größenabschätzung (**Magnitude estimation**):

- Die Länge eines Striches markiert z.B. die Akzeptabilität eines Vergleichssatzes;
- die Akzeptabilität von Testsätzen wird durch entsprechende Strichlängen markiert.

Fragebogen (Fortsetzung)

Es gibt weitere Fragebogentechniken, die in der Sprachwissenschaft eingesetzt werden, z.B. Satzkomplettierungsaufgaben oder Aufgaben, sprachliche Reaktionen in einem Kontext anzugeben.

➤ **Lückentests** (z.B. Auxiliarselektion)

|*Peter hereingekommen, sich hingesetzt und sofort angefangen.*

➤ **Komplettierungstext Syntax** (z.B. Selektion von Komplementierern *ob* vs. *dass*)

Peter zweifelt,

Peter bezweifelt,

➤ **Komplettierungstext Diskurs** (Diskursergänzung)

z.B. Fortsetzung zur Identifizierung impliziter Kausalität: Fortsetzung *er/sie*

Martin hat sich bei Katharina beschwert, weil

Martin hat Katharina beeindruckt, weil

z.B. Höflichkeit bei bestimmten sprachlichen Handlungen):

Sie haben sich von einem guten Freund ein (nicht besonders wertvolles) Buch ausgeliehen und können es nicht mehr finden. Sie wollen sich bei Ihrem Freund entschuldigen. Was sagen Sie?

.....
.....

Diskurskontext systematisch nach Forschungsfragen variieren (z.B.: guter Freund vs. Bekannter vs. fremde Person; weniger wertvolles Buch vs. wertvolles Buch).

Bedenkenswert: Nicht völlig authentisch, schriftliche Produktion ≠ mündliche Produktion.

Reaktionen auf bestimmte Handlungen, z.B. Richtungsangaben:

Lege den Ball vor den Stuhl!

(Aufgabe: Ankreuzen).



Beim Aufbau des Fragebogens ist zu beachten:

- Die Beantwortung früherer Fragen kann Einfluss haben auf die Beantwortung späterer Fragen, u.a. Lern-Effekt. Um dem entgegenzutreten, kann man die Reihenfolge zufällig variieren.
- Heikle Fragen werden leichter beantwortet, wenn sie später gestellt werden (dies ist auch bei Interviews wichtig).
- Bei heiklen Fragen kann man auf die Ehrlichkeit der Antwort Bezug nehmen. Beispiel:
Fast alle Menschen lassen ihrem Ärger ab und zu Luft. Wann haben Sie das letzte Mal geflucht? Beschreiben Sie die Situation, und beschreiben Sie, was sie gesagt und getan haben.

Metadaten und Anonymisierung

Wahrscheinlich werden Sie verschiedene Gruppen von Personen untersuchen (z.B. 10-jährige vs. 15-jährige, Bilinguale vs. späte Fremdsprachenlerner).

Sie müssen die notwendigen Metadaten hierzu aufnehmen, und darüber hinaus weitere Daten, die relevant sein könnten (z.B. Geschlecht, Bildungsgrad, regionaler Hintergrund).

Es können Faktoren eine Rolle spielen, an die Sie nicht denken, z.B. Rechts/Linkshändigkeit der Probanden, Tageszeit und Jahreszeit der Befragung.

In aller Regel müssen Sie den Versuchspersonen zusichern, die Daten anonym zu behandeln. Professionelle Befragungen sichern das in einem eigenen Vertrag mit jeder Versuchsperson, der die Pflichten der Versuchsperson, die Anonymisierung und die Datensicherung genau beschreibt.

Beispiel Fragebogenerhebung

Fragebogenerhebung zu Verständlichkeit von Sprichwörtern, Veronika Türcke, U. Eichstätt;
Forschungsfrage: Allgemeinbildung oder speziell sprachliches Wissen?

Einführender Text

Liebe Teilnehmerin, lieber Teilnehmer,
vielen Dank, dass Sie sich bereit erklärt haben, an meiner Umfrage teilzunehmen und so maßgeblich zum Erfolg meiner Arbeit beizutragen.

Die Daten zum Bildungsstand, Geschlecht und Alter der Teilnehmer zu Beginn der Befragung dient lediglich der statistischen Auswertung der Ergebnisse im Zuge meiner Arbeit. Eine weitergehende Speicherung oder Verwendung der Daten findet nicht statt.

Sämtliche gemachten Angaben werden nur als statistischer Mittelwert in die Arbeit eingehen, sodass eine nachträgliche Verknüpfung der Antworten mit Ihrer Person durch Dritte nicht möglich ist.

Der Ihnen vorliegende Fragebogen beinhaltet 26 Fragen, jeweils 13 als offene Frage und 13 mit Multiple Choice. Die Beantwortung dauert in etwa 15-30 Minuten. Füllen Sie bitte sowohl das Feld „Was bedeutet das Sprichwort“ als auch das Feld „Woher kommt das Sprichwort“ aus. Zu den offenen Fragen ist ein Beispiel gegeben. Im Zweiten Teil kreuzen Sie bitte jeweils die Ihnen am wahrscheinlichsten erscheinende Möglichkeit an. Es ist immer nur eine Möglichkeit richtig.

Sie sollten sich zwar bemühen, eine richtige Antwort zu geben, aber ich möchte Sie inständig bitten, die Antworten nicht im Wörterbuch oder bei Google nachzuschlagen. Jede auf diese Art erzielte richtige Antwort wirkt sich negativ auf die Gültigkeit der von mir erhobenen Daten aus.

Den fertig ausgefüllten Fragebogen senden Sie bitte an die unten genannte Email-Adresse. Sollten Sie Fragen zur Ausfüllung des Fragebogens haben, können Sie mich unter xxx erreichen.

Vielen Dank nochmals, dass Sie mir bei meiner Arbeit helfen

Fragebogen

Statistische Angaben

- Geschlecht: männlich weiblich
- Alter: bis 25 26-40 40-60 über 60
- Schulabschluss: keiner Quali mittlere Reife
 Abitur Hochschulabschluss
- Fremdsprachen: Englisch Französisch Spanisch
 Italienisch Latein andere
falls andere, welche? _____
- Studieren Sie eine Sprache, oder haben sie eine Sprache studiert?
 Ja Nein
falls „ja“, welche? _____
- Auf einer Skala von 1-6 mit 1=sehr wenig, wie sehr interessieren Sie sich für Sprache?
 1 2 3 4 5 6

Offene Fragen

Nr.	Spruchwort	Was bedeutet das Sprichwort	Woher kommt es?
Bsp.	„von der Muse geküsst sein“	Künstlerische Inspiration haben.	Die Musen waren die griechischen Schutzgöttinnen der Künste.
1.	Die neuen Hartz Gesetze bedeuten einen gewaltigen Aderlass für alle Arbeitslosen.		

Experiment

Durch Experimente können Sie – wie bei Fragebögen – auf die anfallenden Daten so Einfluss nehmen, dass sie für Ihre Hypothesen relevant sind.

Anders als es in der Regel bei Fragebögen untersuchen Sie aber nicht die bewusste Einstellung (bzw. die Mitteilung dieser Einstellung) von Versuchspersonen, sondern deren unmittelbares Verhalten.

Das untersuchbare Verhalten steht allerdings oft nur indirekt mit den sprachlichen Fähigkeiten in Verbindung, an denen Sie eigentlich interessiert sind.

Abhängige und unabhängige Variable

Die logische Struktur eines Experiments (Fragebögen sind ein Sonderfall von Experimenten):

- Eine **Hypothese** sagt voraus: Es gibt eine systematische **Abhängigkeit** zwischen der Ausprägung von zwei (oder mehreren) Eigenschaften.

Beispiele: Tageszeit und Fähigkeit, einen gehörten Satz wörtlich wiederzugeben;
Regionale Herkunft und Akzeptanz der Verlaufsform *Er ist den Tee am Trinken*.
Alkoholgehalt im Blut und Fähigkeit, Zungenbrecher akkurat nachzusprechen

- Mindestens eine dieser Eigenschaftsausprägungen ist experimentell manipulierbar.

Dies ist die **unabhängige Variable**. Sie wird im Experiment variiert.

Beispiel: Versuchspersonen werden zu unterschiedlichen Tageszeiten aufgefordert, Sätze nachzusprechen;

Versuchspersonen aus unterschiedlichen Regionen werden dazu aufgefordert, Sätze in der oben angegebenen Verlaufsform zu beurteilen.

- Im Experiment wird die **abhängige Variable** beobachtet.

Beispiel: Wie gut sprechen Versuchspersonen Sätze nach;
wie beurteilen sie Sätze mit der oben angegebenen Verlaufsform.

- Die Beobachtungen werden nach der unabhängigen Variable und der abhängigen Variable notiert.

Es wird mit statistischen Verfahren untersucht, ob tatsächlich die erwarteten Korrelationen bestehen, und in welcher Stärke.

Und es wird untersucht, ob diese eventuell nur ein zufälliges Ergebnis darstellen (z.B. bei zu kleiner oder schlecht gewählter Stichprobe), oder ob sie auch für die Grundgesamtheit gültig sind.

Bei der Manipulation der unabhängigen Variablen muss man sich überlegen, wie diese variiert wird.

Beispiele:

- Untersucht man z.B. bei der Verlaufsform nur Sprecher aus Köln und aus München, oder deckt man eine größere regionale Spannbreite ab?
- Untersucht man bei dem Tageszeitexperiment zu jeder Stunde, oder nur um 8:00 Uhr / 20:00 Uhr

Die unabhängige Variable nennt man auch **Faktor**.

In den Beispielen gibt es einen Faktor.

Wir sprechen daher von einem **einfaktoriellem Design**.

Manchmal werden zwei oder mehr unabhängige Variablen untersucht:

Wir sprechen dann von einem **mehrfaktorielles Design**.

Beispiel:

- Tageszeit und Schlaftyp² der Versuchsperson – Nachsprechen eines Satzes
- Herkunft, Alter und Bildungsgrad der VPn – Akzeptanz der Verlaufsform
- Herkunft und Verbtyp (z.B. *trinken, lesen, sehen*) – Akzeptanz der Verlaufsform

In der Regel sind bei mehrfaktoriellem Design auch mehr Versuchspersonen nötig, um für die Grundgesamtheit gültige Aussagen zu treffen.

Zumindest für Abschlussarbeiten, bei denen diese Ressourcen nicht zur Verfügung stehen, ist davon eher abzuraten.

² Hierfür gibt es die technischen Begriffe *Lerche* und *Eule*.

Labor- vs. Feldexperimente

Im **Laborexperiment** löst man ein Phänomen aus der natürlichen Umgebung heraus und untersucht es davon isoliert.

- Vorteil: man kann die unabhängige Variable leichter manipulieren und die Umstände, in denen das Experiment durchgeführt wird, besser kontrollieren und konstanthalten.
- Problem: Die Aussagekraft eines Experiments kann beeinträchtigt sein, weil die Umstände unnatürlich sind.

Im **Feldexperiment** untersucht man ein Phänomen in seiner natürlichen Umgebung. Zum Beispiel kann man an Wegbeschreibungen interessiert sein, sich als Tourist ausgeben und Passanten nach dem Weg zu bestimmten Orten fragen.

- Problem: Man hat weniger Kontrolle über die Umstände des Experiments.

Bei Feldexperimenten muss man oft mit **Kontrollgruppen** arbeiten, um die unabhängige Variable zu manipulieren.

Beispiel: Untersuchung des Erfolgs einer Vokabellernmethode; in einer Schulklasse wird diese Methode eingesetzt, in einer anderen (der Kontrollgruppe) nicht.

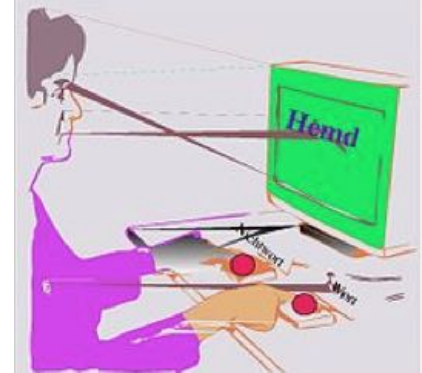
Typen von Experimenten: Reaktionszeitmessung

Priming-Experimente

Grundlage: Die Verarbeitung eines Reizes kann die Verarbeitung eines anderen modifizieren.

Beispiel: Die Geschwindigkeit, in der eine lexikalische Entscheidungsaufgabe getroffen wird (ist α ein Wort oder nicht)? wird beeinflusst von einem unmittelbar vorher gegebenen Reiz (einem Wort β):

Bluse – Hemd, Haus – Hemd, fremd – Hemd



Speeded grammaticality judgement

Die Beurteilung der Grammatikalität eines Satzes muss so schnell wie möglich abgegeben werden.

Annahme: Die Verarbeitungstiefe beeinflusst die Geschwindigkeit und Korrektheit der Beurteilung.

Beispiel: Aktiv/Passiv-Sätze.

Self-paced reading

Ein Text wird Wort für Wort eingeblendet, der Leser drückt eine Taste, um zum nächsten Wort zu gelangen. Gemessen wird die Geschwindigkeit des Drückens, die mit Eigenschaften der vermittelten sprachlichen Information korreliert wird.

Typen von Experimenten: Bewegungsregistrierung

Blickbewegungsmessung (eye tracking)

Blickbewegungsregistrierung

Durch die Erfassung der Blickbewegung (eye tracking) können sprachliche Verarbeitungsprozesse gemessen werden.



Blickerfassung beim Lesen

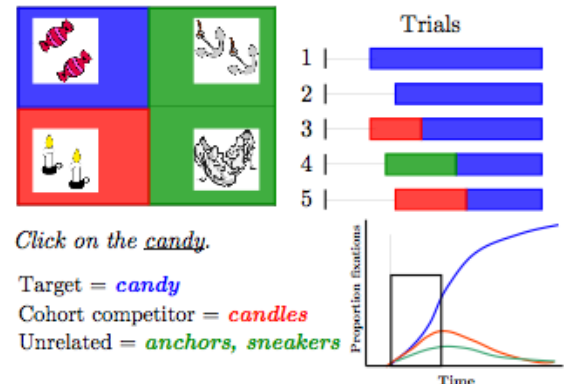
Beobachtung beim Lesen
(Sakkaden, Verlangsamung, Zurückspringen)

Blickerfassung bei der Sprachverarbeitung („Visual World Paradigm“)

Beobachtung beim Verarbeiten von sprachlichen Aufgaben.

DANS, RÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den "synkretiska dansen", en sammansmältning av olika kulturellers dans, har jag i mitt fallföreläse under hösten fört på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar gör sina röster höra genom sång, musik, skrik, skraal och gestaltade känslor och uttryck med hjälp av kroppsspråk och dans.



Empirische Untersuchungen: Blickbewegungsregistrierung

Handbewegungsmessung (Mouse tracking)

Erfassung der Computermaus-Bewegung

Beispiel: Beurteilung, welche Sätze zu Situationen passen,
hier: Verständnis der Disjunktion (*oder*) vs. Konjunktion (*und*)
(Sauerland e.a. 2015)

Versuchsmaterial

GUT

SCHLECHT

Verhalten der Versuchspersonen
(Kinder; japanisch)

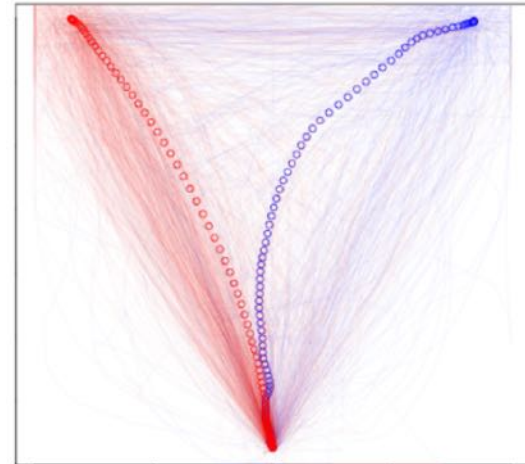
66% GUT

SCHLECHT 34%

Es gibt einen Gorilla oder ein Schaf



X



Typen von Experimenten: Physiologische Experimente

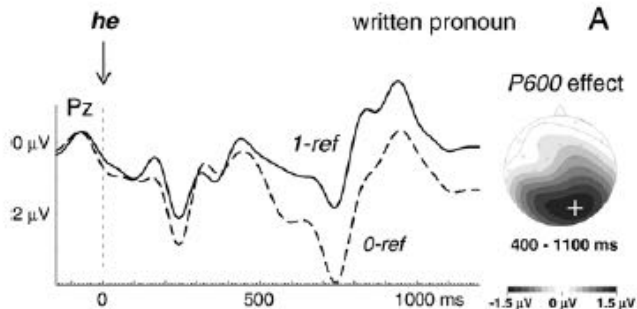
Ereignis-korrelierte Potentiale (EKP, ERP)

Abnahme der durch Nervenzellen erzeugten elektrischen Gehirnaktivität durch Aufzeichnung der Spannungsschwankungen auf der Kopfhaut (EEG);
sehr gute zeitliche, weniger gute räumliche Auflösung

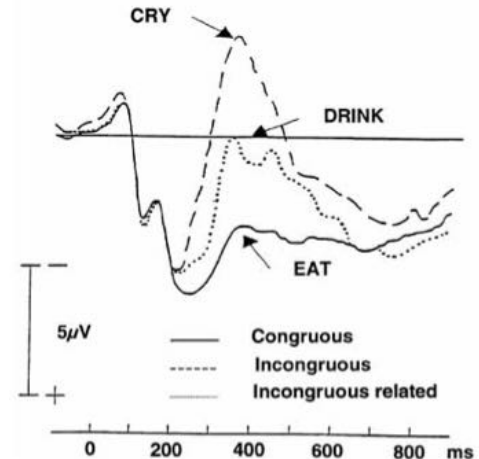
- N400-Effekt, semantische Integration (Beispiel: Kutas & Hillyard 1984)
- P600-Effekt, syntaktische Integration (Beispiel: Berkum e.a. 2007, Pronomenauflösung)

1-ref: David shot at Linda as...

0-ref: Anna shot at Linda as...



THE PIZZA WAS TOO HOT TO...



Beschreibende Statistik

Die nächsten Abschnitte beziehen sich auf statistische Methoden; vgl. hierzu die neueren Lehrbücher

- Stefan Th. Gries, *Statistik für Sprachwissenschaftler*, Vandenhoeck & Ruprecht 2008.
- Keith Johnson, *Quantitative Methods in Linguistics*, Blackwell 2008.
- Claudia Meindl, *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. 2011.
- Andrew Fields u.a., *Discovering Statistics using R*. 2012.

Im vorliegenden Teil geht es um einige wesentliche Grundzüge der **deskriptiven (beschreibenden) Statistik**, d.h. um die Beschreibung und Darstellung von Daten.

Grundgesamtheiten und Stichproben

Wir beschreiben eine Menge von sog. **statistischen Einheiten**, wie z.B.

- die Menge der Wörter in einem Text
- die Reaktionen von Sprechern auf eine Liste von Fragen
- die Menge der grammatisch wohlgeformten Sätze einer Sprache
- die Eigenschaften von Grammatiken von Sprachen.

Die Menge der statistischen Einheiten nennt man **Grundgesamtheit** ("population").

Diese kann **endlich** sein (z.B. die Menge der Sprecher einer Sprache)

oder **unendlich** (z.B. die Menge der wohlgeformten Sätze einer Sprache)

Oft können nicht alle statistischen Einheiten untersucht werden, weil die Grundgesamtheit zu groß ist oder nicht alle statistischen Einheiten bekannt sind. Wir untersuchen daher Teilgesamtheiten, sog.

Stichproben (engl. „samples“) und schließen damit auf die Grundgesamtheit.

Beispiele von Stichproben:

- Die durchschnittliche Silbenzahl von Wortvorkommnissen der deutschen Sprache wird anhand von Wortvorkommnissen in einer Stichprobe von Texten abgeschätzt.
- Es wird abgeschätzt, wie oft Subjekte vs. Objekte in informellen Konversationen animat sind (d.h. sich auf belebte Entitäten beziehen), anhand eines Korpus gesprochener Sprache
- Die Reaktionen von Sprechern des Deutschen in einem psycholinguistischen Experiment werden anhand einer Stichprobe von 50 Versuchspersonen abgeschätzt
- Grammatische Eigenschaften wie z.B. die Wortstellung im einfachen Satz in den Sprachen der Welt werden anhand einer Stichprobe von 100 Sprachen abgeschätzt.

Damit anhand von Stichproben auf die Grundgesamtheit zurückgeschlossen werden kann, muss die Stichprobe groß genug und hinreichend zufällig ausgewählt werden.

(Wenn wir nur bilinguale Sprecher des Deutschen wählen oder nur Gesetzestexte, oder wenn wir nur europäische Sprachen wählen, werden wir möglicherweise zu falschen Rückschlüssen kommen).

Merkmale und ihre Ausprägungen

Was beobachten wir an den statistischen Einheiten?

Wir beobachten an ihnen die **Ausprägung** bestimmter **Merkmale**.

Beispiele; die Ausprägungen der Merkmale stehen in Klammern:

- Die Länge der Wörter (z.B. 1, 2, 3, 4, ... Grapheme, Phoneme, Silben, oder Morpheme):
Schweinsgalopp ist 14 Grapheme, 10 Phoneme, 3 Silben, 2 Morpheme lang).
- Die syntaktische Form der Sätze (z.B. Verb-Erst, Verb-Zweit, Verb-Letzt)
- Die Reaktionsarten von Sprechern auf Fragen (z.B. *Ja/Nein/Weiss nicht*)
- Die Reaktionszeiten von Antworten auf eine Frage (z.B. 100ms, 200ms, 300ms, ...)
- Die Grundwortstellung von Subjekt, Objekt und Verb in transitiven Sätzen
in den Grammatiken von Sprachen (z.B. SVO, SOV, VSO, OVS, OSV, VOS).

Die Ausprägungen von Merkmalen hängen von bestimmten beobachtbaren Größen bestimmt, den **Einflussgrößen** oder **Faktoren**.

Beispiele:

- Textsorte
- Entstehungszeit des Textes
- Alter der Sprecher
- Sprachfamilie, zu der eine Sprache gehört

Daneben muss man auch mit dem Einfluss von nicht beobachteten Größen rechnen, den **Störgrößen**.

Merkmalstypen: Diskret vs. Stetig

Die Ausprägungen von Merkmalen können von verschiedenen Typen sein, was wesentlich ihre Darstellung bestimmt.

Diskrete Merkmale

haben klar voneinander unterscheidbare Ausprägungen. Beispiele:

- Reaktionen auf eine Frage in einem Experiment (Ja/Nein/Weiss nicht)
- Grundwortstellung einer Sprache (SVO, SOV, VSO, OVS, OSV, VOS)
- Syntaktische Form der Sätze des Deutschen (V1, V2, VL)
- Zahl der Phoneme in den Wörtern eines Korpus (1, 2, 3, 4, ...).

Stetige Merkmale

Hier können die Ausprägungen beliebig nah beieinander liegen. Beispiele:

- Reaktionszeiten auf eine Frage (800ms, 801ms, 800,5ms, 800,7ms usw.)
- Mögliche Länge der Voice Onset Time (VOT, der Zeit zwischen Öffnen des Verschlusses und Einsetzen der Stimmlippen-Bewegung (20ms, 21ms, 20,5 ms usw.)

Da die Messungen für stetige Merkmale in der Regel nicht beliebig präzise gemacht werden können, werden stetige Merkmale oft auch als diskret behandelt.

Man kann z.B. bei Reaktionszeiten unterscheiden: 800-899ms, 900-999ms, usw.

Nominal-, Ordinal- und Kardinalskalen

Unterscheidung von Merkmale nach der Struktur ihren Ausprägungen – sog. **Skalentypen**.

Nominalskalen

Keine bestimmte Ordnung zwischen den einzelnen Ausprägungen.

- Grundwortstellung einer Sprache (SVO, SOV, VSO, OVS, OSV, VOS)
- Syntaktische Form der Sätze des Deutschen (V1, V2, VL)

Ordinalskalen

Die Ausprägungen sind geordnet;

es ist aber nicht sinnvoll, von Abständen zwischen Ausprägungen sprechen.

- Reaktionen auf eine Frage in einem Experiment (Ja – Weiss nicht – Nein)
- Einschätzung der Akzeptabilität einer Wortkette durch Sprecher einer Sprache:
(5-Punkte-Skala: Normal – Akzeptabel – Grenzwertig – Schlecht – Sehr schlecht)

Nominalskalen u. Ordinalskalen mit endlich vielen Ausprägungen: **qualitative** Skalen.

Kardinalskalen

Die Ausprägungen sind geordnet; man kann von Abständen zwischen Ausprägungen zu sprechen.

- Zahl der Phoneme in den Wörtern einer Sprache (1, 2, 3, ...)
- Reaktionszeiten auf Fragen in einem Experiment (800 ms, 900 ms, 1000 ms, ...)

Statistikprogramme

Je nach der Art der Ausprägungen von Merkmalen können diese verschieden grafisch dargestellt werden. Dies erlaubt eine bessere Einschätzung als die Präsentation von Daten in Tabellen.

Tabellenkalkulationsprogramme

- Microsoft Excel, Teil von Microsoft Office
- Tabellenkalkulation von LibreOffice: LibreOffice Calc.
Siehe hierzu: <http://www.comfsm.fm/~dleeling/statistics/notes000.html>
Bequeme Dateneingabe; Statistikfunktionen; Grafische Darstellung statistischer Funktion.

Statistikprogramme

- R: Ein frei verfügbares, plattform-unabhängiges Statistikpaket, das wesentlich mehr erlaubt als Excel. Link: <http://www.r-project.org/>, mit Handbuch *An Introduction to R*.
Es gibt verschiedene Graphical User-Schnittstellen, z.B. Deducer: <http://www.deducer.org/>
- IBM SPSS: Ein kommerzielles Programm in vielen Varianten, das für Studenten (mit zeitlich beschränkter Lizenz) preisgünstig angeboten wird, für andere Anwender aber teuer ist.
Link: <http://www-01.ibm.com/software/de/analytics/spss/>
- PSPP: Ein freies Programm, das SPSS nachgebildet ist und mit den SPSS-Dateien kompatibel ist:
<http://www.gnu.org/software/pspp/>

Darüber hinaus gibt es noch zahlreiche weitere Statistikprogramme und auch Online-Ressourcen.
Hier verwendet: **LibreOffice Calc** (Excel ist ähnlich) und **R** (bitte von der Seite oben installieren!)

Darstellung von qualitativen Daten

Beispiel: Häufigkeit der Grundwortstellung von Subjekt, Verb und Objekt in transitiven Sätzen, nach dem *World Atlas of Language Structure*, siehe <http://wals.info/>.

Stellung	SOV	SVO	VSO	VOS	OVS	OSV	keine dominante St.
Anzahl Sprachen	565	488	95	25	11	4	189

Es handelt sich hierbei um eine Nominalskala:

Es gibt keine bestimmte Ordnung zwischen den einzelnen Ausprägungen.

Wir wollen diese Daten zur besseren Veranschaulichung in einem **Säulendiagramm** (Balkendiagramm, “bar chart”) darstellen.

Darstellung qualitativer Daten in LibreOffice

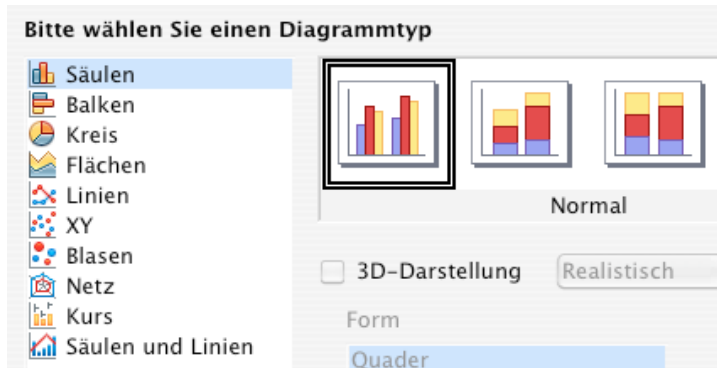
Wir tragen die Daten in einer Zeile eines Arbeitsblattes in der Tabellenkalkulation ein:

	A	B	C	D	E	F	G
1	SOV	SVO	VSO	VOS	OVS	OSV	NichtDom
2	565	488	95	25	11	4	189
3							

Wir markieren die Daten und drücken auf die Schaltfläche für Grafik

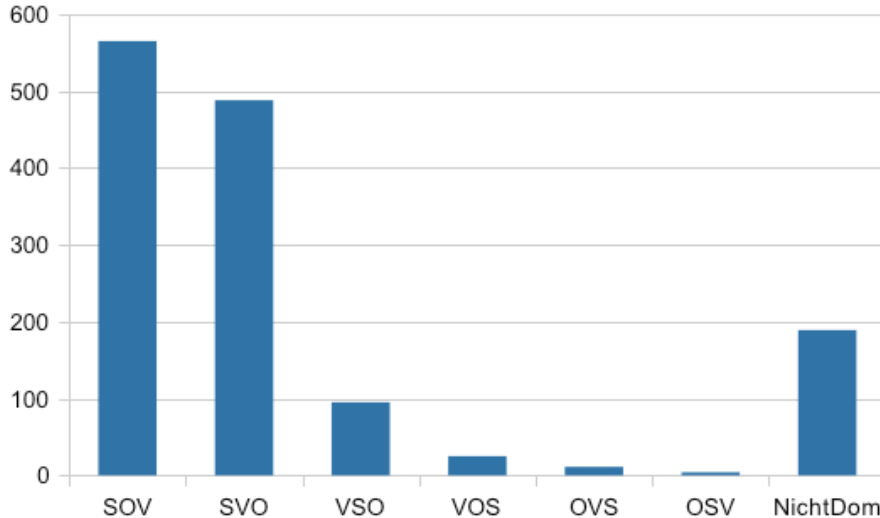


Dabei haben wir viele mögliche Darstellungen zur Auswahl.

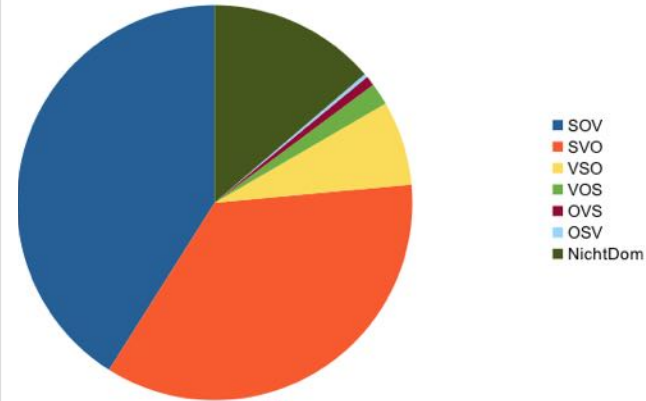


Es sollten möglichst einfache Darstellungen, keine 3D-Darstellungen gewählt werden, hier z.B. Säulendiagramm oder Kreisdiagramm.

Säulendiagramm
(absoluter Größenvergleich)



Kreis („Torten“) diagramm
(relativer Größenvergleich)

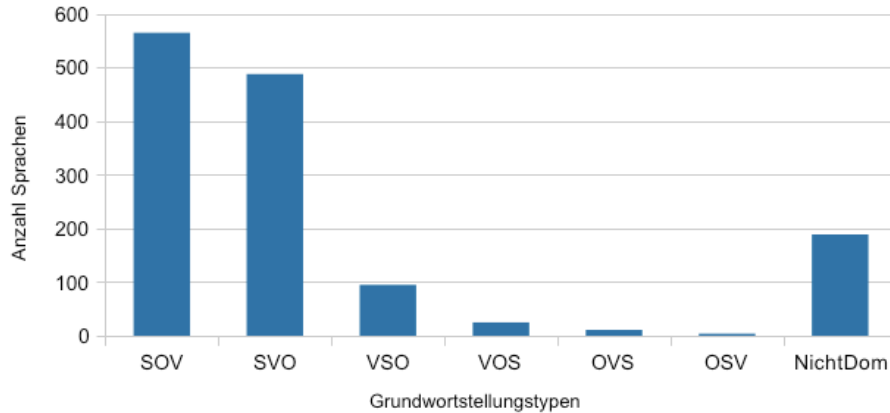


Für die Einfügung in Artikel müssen Sie die Grafiken bearbeiten, da die vorgegebenen Grundeinstellungen meist nicht optimal sind.

Es gibt zahlreiche Optionen.

Hier Beispiel: Bearbeitung des Säulendiagramms.

Häufigkeiten von Grundwortstellungstypen



Einfügen von Beschriftungen

Diagramm-Assistent

Schritte

1. Diagrammtyp
2. Datenbereich
3. Datenreihen
4. Diagrammelemente

Wählen Sie Titel, Legenden- und Gittereinstellungen

Titel: en von Grundwortstellungstypen Legende anzeigen

Untertitel:

X-Achse: Grundwortstellungstypen Links

Y-Achse: Anzahl Sprachen Rechts

Z-Achse: Oben

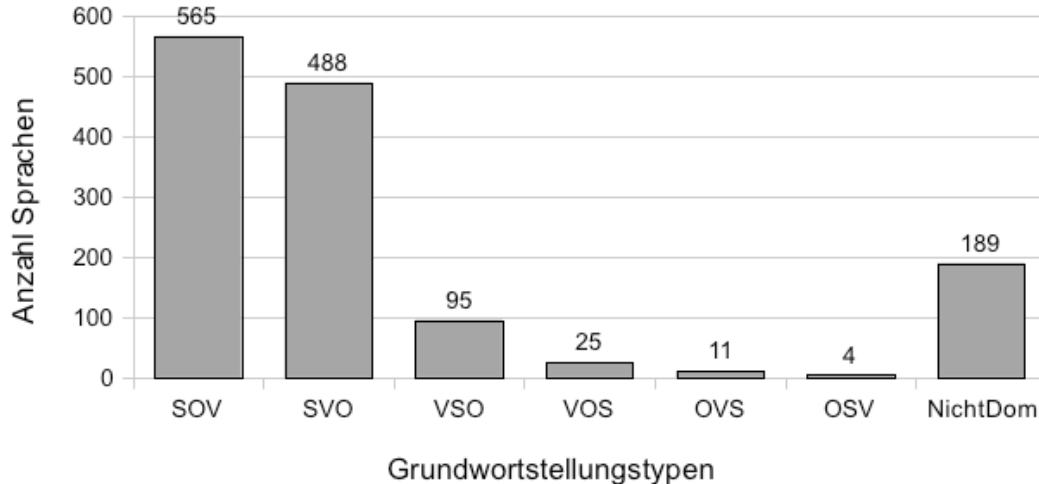
Unten

Gitter anzeigen

X-Achse Y-Achse Z-Achse

Nachdem das Diagramm in das Tabellenkalkulationsblatt eingefügt wurde, kann es weiter durch Anklicken (ctrl oder Rechtsklick) von Teilen verändert werden – ausprobieren!

Häufigkeiten von Grundwortstellungstypen



Säulen grau
(besserer Druck)
Legenden
vergrößert
Abstand zwischen
Säulen verkleinert
Absolute Werte
über den Säulen

Anschließend kann man die Grafik kopieren und in ein Textdokument einfügen.

Darstellung qualitativer Daten in R

Wir rufen das Programm R auf und erhalten ein Konsole-Fenster mit einem Prompt `>`, bei dem wir Befehle eintippen können.

Diese Befehle werden sofort ausgewertet, z.B. mathematische Befehle: $(0,3 \times 8,7)^2$

```
> (0.3 * 8.7)^2
[1] 6.8121
>
```

Wichtige Konventionen:

- Zahlen werden als Zahlen notiert, z.B. **0.3** (englische Notation mit Dezimalpunkt, nicht 0,3).
- Variablen sind Zeichenketten (außer Zahlen), z.B. **Grundwortstellung**, **x**, **x7**
- Zeichen und Folgen von Zeichen werden in doppelten Anführungszeichen angegeben, z.B. **"A"**, **"7"**, **"SVO"**, **"Hinz und Kunz"**
- Argumente von Funktionen stehen in Klammern, z.B. bei der Konkatenationsfunktion **c**
- Zuweisungen von Werten zu Variablen geschehen durch den Operator **<-** (alternative auch **=**)

Beispiel: Bildung einer Folge (eines Vektors) von Triplets von Buchstaben mit Komposition **c**, Zuweisung auf Variable *Wortstellungen*, Aufruf dieser Variable und Ausgabe der Werte:

```
> Wortstellungen <- c("SOV", "SVO", "VSO", "VOS", "OSV", "OVS", "NichtDom")
> Wortstellungen
[1] "SOV"      "SVO"      "VSO"      "VOS"      "OSV"      "OVS"      "NichtDom"
```

Zugriff auf Vektorwerte:

```
> Wortstellungen[1:4]
[1] "SOV" "SVO" "VSO" "VOS" , Ausgabe der Werte an 1. bis 4. Stelle von „Wortstellungen“.
```

Eingabe von Werten

Mit der Sprache R können wir wie folgt vorgehen (die Angaben sind jeweils bei dem R-Prompt einzutippen; wir können die Daten in beliebiger Reihenfolge angeben).

```
> Grundwortstellung <- c("SVO"=435, "VOS"=26, "NA"=172, "SOV"=497,  
"OSV"=4, "VSO"=85, "OVS"=9)↵
```

Damit werden die angegebenen Werte zu einem sog. **Vektor** zusammengefügt und mit dem Zeichen "<->" einer **Variablen** mit dem Namen Grundwortstellung zugeordnet. (Statt "<->" können wir auch "=" schreiben).

Es gibt natürlich auch andere Eingabemöglichkeiten – dazu später.

Den Inhalt dieses Vektors kann man sich anschauen:

```
> Grundwortstellung↵
```

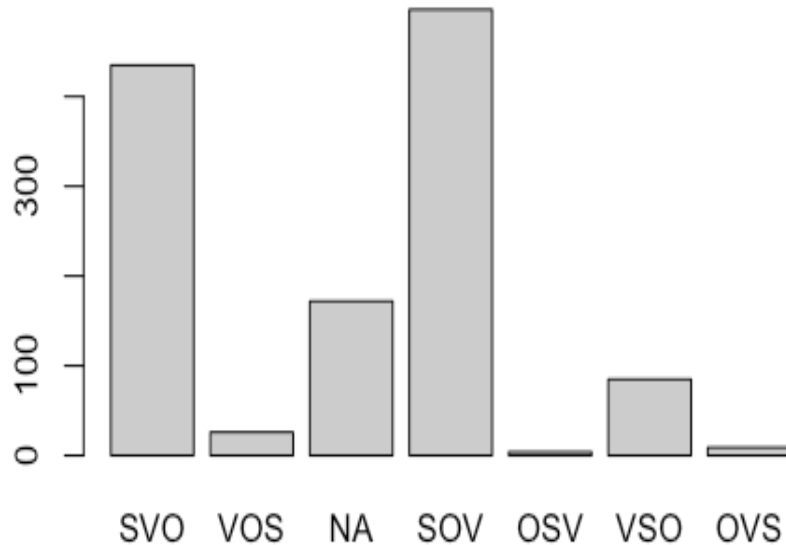
Wir erhalten eine tabellarische Darstellung der Daten:

SVO	VOS	NA	SOV	OSV	VSO	OVS
435	26	172	497	4	85	9

Wir können diese Daten in einem Säulendiagramm darstellen:

```
> barplot(Grundwortstellung)
```

In einem eigenen Fenster wird folgendes Diagramm angezeigt:



Wir können zur besseren Darstellung die Daten nach absteigender Größe sortieren:

```
> sort(Grundwortstellung)↵
```

```
OSV OVS VOS VSO NA SVO SOV  
4 9 26 85 172 435 497
```

Das ist aber nicht was wir wollen – die Sortierreihenfolge ist aufsteigend.

Es gibt eine Hilfsfunktion für alle Befehle: `?sort` Wir werden über ein Hilfe-Fenster darüber informiert, dass wir eingeben müssen:

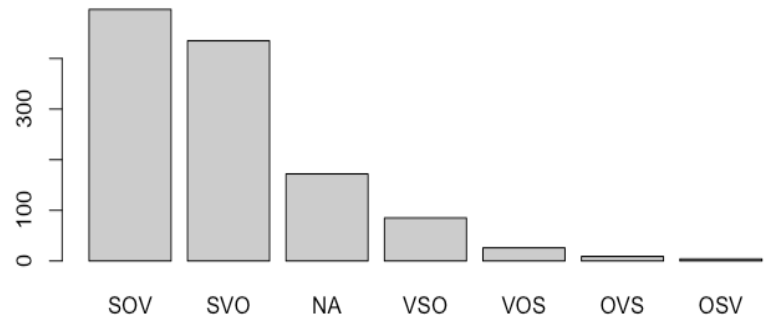
```
> sort(Grundwortstellung, decreasing=TRUE)↵
```

Wir können den Vektor `Grundwortstellung` ein für alle Mal so sortieren:

```
> Grundwortstellung = sort(Grundwortstellung, decreasing=TRUE)↵
```

Wir erzeugen nun ein Säulendiagramm und erhalten die folgende Darstellung

```
> barplot(Grundwortstellung)
```



Hinweis:

R merkt sich frühere Eingaben; sie können diese mit den Pfeiltasten wieder sichtbar machen und ggf. modifizieren.

Es gibt viele verschiedene Möglichkeiten, diese Darstellung zu verändern (durch Beschriftung der Achsen, verschiedene Farben für die Kolumnen usw.); man erhält hierüber über `?barplot` Auskunft.

Aufgabe:

Bitte führen Sie die R-Befehle wie beschrieben durch. Erzeugen Sie ein Histogramm von Grundwortstellung mit der Überschrift "Häufigkeit von Grundwortstellungstypen", der Beschriftung "Grundwortstellungstypen" für die X-Achse, "Anzahl von Sprachen" für die Y-Achse, und einer y-Achse, die Werte von 0 bis 500 aufweist.

(Hinweis: Überschriften mit `main=...`, Label der X-Achse mit `xlab=...`, Label der Y-Achse mit `ylab=...`, wobei Ausdrücke jeweils in Anführungszeichen angegeben werden. Bereich der y-Achse mit `ylim=...`, wobei der Bereich mit `range(n,m)` angegeben wird und n hierbei die unterste, m die oberste Zahl ist.)

Darstellung von quantitativen Daten

Bei quantitativen Daten spielt die Ordnung zwischen den Ausprägungen eine wesentliche Rolle;

typischerweise handelt es sich bei den Ausprägungen um numerische Daten.

Beispiel:

Zeichenlänge von Wörtern (Tokens) in dem englischsprachigen PAROLE-Korpus.

Die Daten zeigen:

Längere Wörter kommen seltener vor; wir haben ein relatives Maximum bei 3 (z.B. der definite Artikel: *the*)

und auch bei 1

(der indefinite Artikel *a*, das Personalpronomen *I* und mögliche Artefakte, z.b. in Abkürzungen wie *e.g.*).

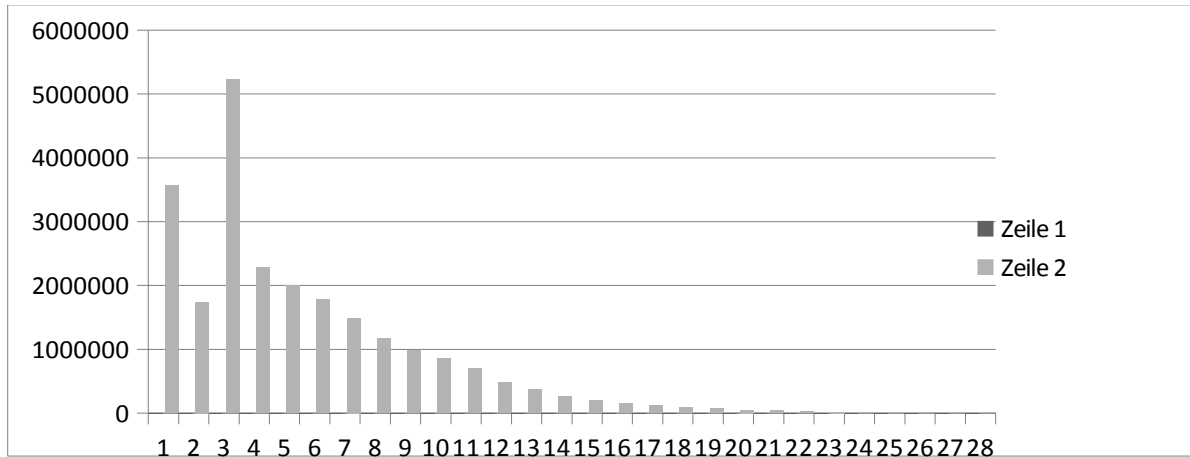
1	3564430	11	693824	21	35344
2	1742674	12	484361	22	26020
3	5226684	13	373814	23	17548
4	2282737	14	257697	24	13768
5	2004860	15	192147	25	9134
6	1775106	16	147870	26	5727
7	1489365	17	115291	27	3928
8	1175359	18	90033	28	2717
9	985117	19	67285	29	1744
10	861835	20	50051	30	1266

Darstellung von quantitativen Daten mit LibreOffice

Wir geben wiederum die Daten in ein Arbeitsblatt ein:

	A	B	C	D	E	F
1	1	2	3	4	5	6
2	3564430	1742674	5226684	2282737	2004860	1775106
3						

Wir können daraus wieder ein Säulendiagramm gewinnen:

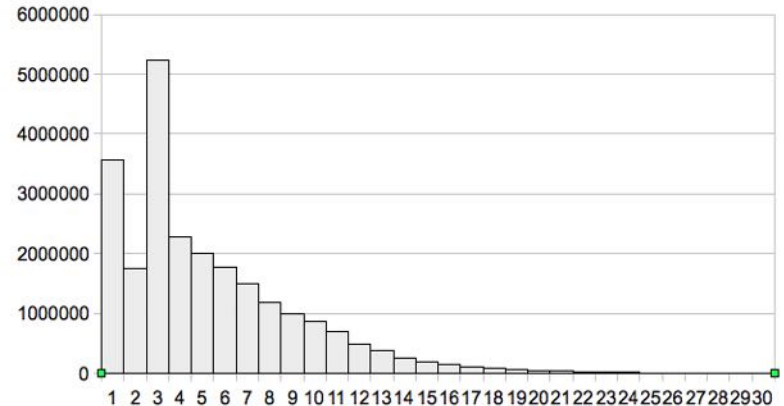


Da es sich um **quantitative** Daten handelt, sollte das Säulendiagramm in ein **Histogramm** umgewandelt werden.

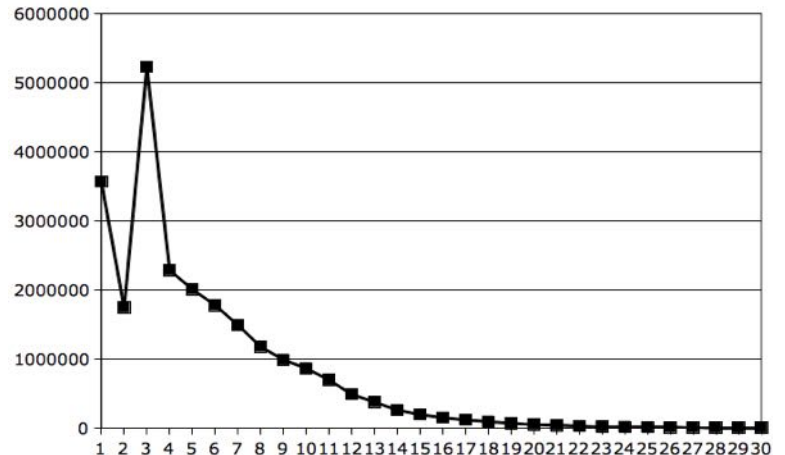
Konvention beim Histogramm:

- Keine Zwischenräume zwischen den Säulen;
- Die Fläche der “Säulen” ist eigentlich relevant, d.h. die Breite der Säulen kann auch unterschiedlich sein.

Klicken auf die Säulen erlaubt es, sie entsprechend zu formatieren.



Bei quantitativen Daten sind auch **Liniendiagramme** sinnvoll.



Darstellung von quantitativen Daten mit R

Wir können diese Daten am R-Prompt eingeben.

Mit der `scan`-Funktion weisen wir der Variablen `w1` die eingegebenen Werte zu:

```
> w1 <- scan()  
1: 3564430  
2: 1742674  
3: ...
```

Wir schließen die Reihe ab, indem wir am Ende (bei der Zahl 31) ↵ eingeben.

Die hier verwendete `scan`-Funktion kann auch dazu verwendet werden, um Daten aus einer Datei einzulesen.

Wir geben mit einem Texteditor die folgende Reihe von Zahlen in einer Zeile ein und speichern sie unter dem Namen `wortlaenge.txt` ab.

```
3564430; 1742674; ... 1744; 1266
```

Wir können nun mit einem der beiden Befehle diese Datei einlesen:

```
> t1 <- scan("Dateipfad+Dateiname", sep=";")  
> t1 <- scan(file.choose(), sep=";")
```

Im ersten Fall muss man den Dateinamen und `-pfad` angeben, im zweiten Fall öffnet sich ein Dialogfeld, und man kann danach suchen (bei Windows und OSX).

Die Angabe `sep=";"` sagt, wie die einzelnen Werte voneinander separiert sind, hier durch das Semikolon.

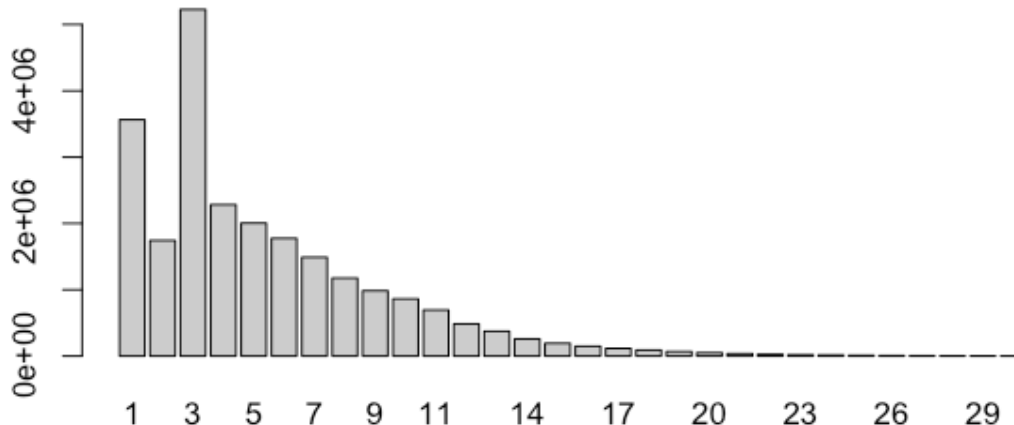
Darstellung im Säulendiagramm / Histogramm

Wir können den auf die Variable `t1` eingelesenen Vektor mit einem Säulendiagramm darstellen (die Angabe `names.arg=1:30` besagt, dass als Säulennamen der Vektor mit den Zahlen von 1 bis 30 verwendet werden soll).

```
> barplot(t1, names.arg=1:30)
```

Hier steht `1:30` für eine Liste von Zahlen von 1 bis 30.

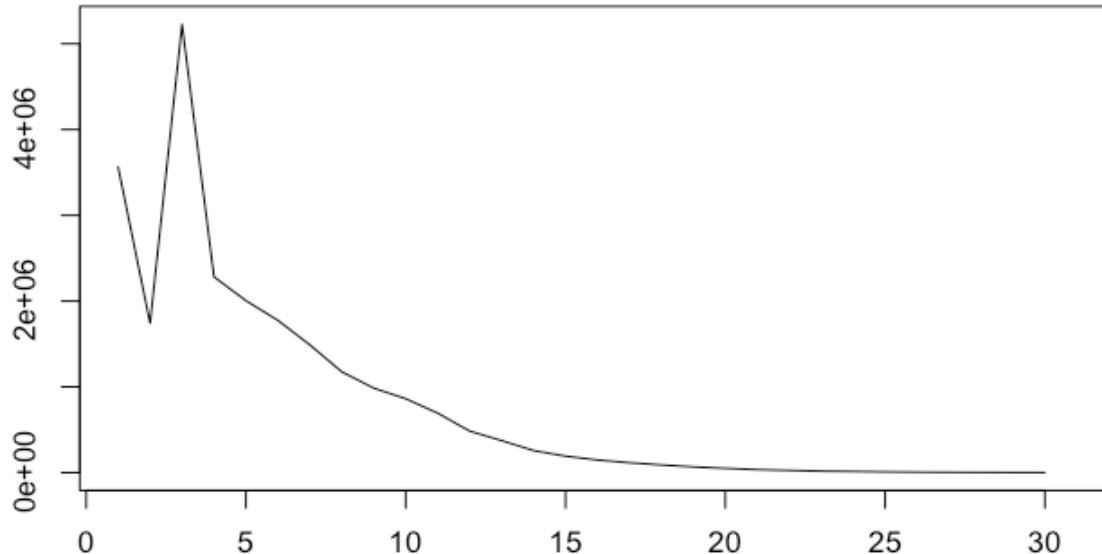
Anzahlen werden in der Exponentenschreibweise angegeben (`4e+06` steht für $4 \cdot 10^6$, also 4 Mio.)



Darstellung im Liniendiagramm

Ein Liniendiagramm kann wie folgt erzeugt werden:

```
> plot(w1, type="l") ↵ (es handelt sich um den Buchstaben l)
```



Für weitere Diagramm-Möglichkeiten siehe `?plot`.

Mittelwerte und Streuung

Nehmen wir an, wir testen die Behaltungsleistung von Vokabeln bei Fremdsprachlernern.

Wir haben dreißig Versuchspersonen, die je zwanzig Vokabeln lernen, welche nach einer Stunde abgefragt werden.

Ergebnisse (tabellarische Darstellung, Beispiel für LibreOffice-Tabellenkalkulation):

VP	Vokabeln	VP	Vokabeln	VP	Vokabeln
1	9	11	14	21	14
2	12	12	17	22	14
3	19	13	12	23	9
4	13	14	15	24	10
5	16	15	13	25	12
6	14	16	17	26	3
7	10	17	14	27	10
8	11	18	11	28	14
9	12	19	16	29	11
10	12	20	16	30	11

	A	B
1	VP	Behaltensleistung
2	1	9
3	2	12
4	3	19
5	4	13
6	5	16
7	6	14
8	7	10
9	8	11
10	9	12
11	10	12
12	11	14
13	12	17

Wir können die Verteilung nun auf verschiedene Weise charakterisieren

Das arithmetische Mittel (Mittelwert, engl. Average)

Auch arithmetischer Mittelwert (μ) genannt;

dies ist die Summe aller behaltene Vokabeln (381) geteilt durch die Menge der Versuchspersonen (30), hier 12,7.

Wenn wir mit $X = (x_1, x_2, x_3, \dots, x_n)$ die Werte für alle $n = 30$ Versuchspersonen angeben, können wir schreiben: $\mu_x =$

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

, vereinfacht: $\mu_x = \Sigma X / n$.

Im folgenden nehmen wir an, dass die Zahl der behaltene Vokabeln in den Feldern B2 bis B31 angezeigt ist.

➤ = SUMME(B2:B31) (die Summe der Felder B2 bis B31), = 381

➤ = MITTELWERT(B2:B31) (der Mittelwert; 381/30), = 12.7

31	30	11	
32		381	Summe
33		12,7	Mittelwert
34			

Neben dem arithmetischen Mittel gibt es weitere Mittelwerte, die eine Menge von Zahlenwerten oft besser charakterisieren.

Minimum, Maximum, Streuung, Median, Quartile, Modalwert

Minimum und **Maximum** sind die minimalen und die maximalen Werte:

➤ $\text{MIN}(B2:B31) = 3$, $\text{MAX}(B2:B31) = 19$

Streuung: die Differenz von Minimum und Maximum, $\text{MAX} - \text{MIN}$, hier: 16.

Median: 50% der Versuchspersonen haben höchstens so viel erzielt und 50% mindestens so viel. Führt das zu zwei Werten (gerade Anzahl von Versuchspersonen), hier 12 und 13: deren arithmetisches Mittel, hier 12,5.

➤ $\text{MEDIAN}(B2:B31) = \text{hier} = 12,5$.

Quartile: Das 1. oder 25%-Quartil: 25% haben höchstens so viel erzielt und 75% mindestens so viel. Das 2. Quartil: 75% haben höchstens so viel erzielt und 25% mindestens so viel. Bei Bedarf wird wieder gemittelt.

Abfrage bei LibreOffice: Minimum 0, 25%-Quartil 1, 50%-Quartil (= Median) 2, 75%-Quartil 3, Maximum 4.

➤ $\text{QUARTILE}(B2:B31, 1) = 11$, $\text{QUARTILE}(B2:B31, 3) = 14$

Der Median lässt sich auch auf Ordinaldaten anwenden und ist weniger beeinflussbar durch Extremwerte ("Ausreißer") als das arithmetische Mittel.

Modalwert: Der am häufigsten auftretende Wert; hier 14 (tritt 6 mal auf).
Bei mehr als einem Modalwert: Bimodale Verteilung.

Vokabeln	Rang	
3	1	
9	2	
9	3	
10	4	
10	5	
10	6	
11	7	
11	8	1. Quartil = 11
11	9	
11	10	
12	11	
12	12	
12	13	
12	14	
12	15	
		Median = 12,5
13	16	
13	17	
14	18	
14	19	
14	20	
14	21	
14	22	
14	23	3. Quartil = 14
15	24	
16	25	
16	26	
16	27	
17	28	
17	29	
19	30	

Die Standardabweichung

Die Streuung gibt ein grobes Maß darüber, wie dicht die Werte beieinander liegen; ein einzelner "Ausreißer" wie die Versuchsperson 3 affiziert die Streuung stark.

Ein besseres Maß ist die **Standardabweichung** σ , die **durchschnittliche Abweichung vom Mittelwert**.

Berechnung der Standardabweichung:

- Berechne jede Abweichung vom Mittelwert: x_1, x_2, x_3, \dots
- Quadriere diese Werte, damit werden sie stets positiv: $x_1^2, x_2^2, x_3^2, \dots$
- Nimm die Summe der quadrierten Abweichungen:

$$X^2 = x_1^2 + x_2^2 + x_3^2, \dots$$
- Teile sie durch die Anzahl der Beobachtungen N (bei Stichproben*: durch $N - 1$).
 Dies ist die **Varianz**.
- Ziehe daraus die Quadratwurzel,
 das macht die Quadrierung im 2. Schritt rückgängig;
 das ist die **Standardabweichung**.

Mit LibreOffice kann man die Standardabweichung direkt berechnen:

- STABW(B2..B31) = 3,1200132626

Ein hoher Wert deutet an: Die Werte streuen stark.

Z.B.: Wenn 15 VPn sich 12 Vokabeln und 15 VPn 13 Vokabeln gemerkt hätten, dann hätten wir die Standardabweichung = 0,50

(* Warum bei Stichproben durch $N - 1$? Weil bei Kenntnis von $N - 1$ Werten und der Summe der letzte Wert ebenfalls bekannt ist, es gibt nur $N - 1$ **Freiheitsgrade**).

D	E
Abweichung vom Mittelwert	Quadrat davon
3,7	13,69
0,7	0,49
-6,3	39,69
-0,3	0,09
-3,3	10,89
-1,3	1,69
2,7	7,29
1,7	2,89
0,7	0,49
0,7	0,49
-1,3	1,69
-4,3	18,49
0,7	0,49
-2,3	5,29
-0,3	0,09
-4,3	18,49
-1,3	1,69
1,7	2,89
-3,3	10,89
-3,3	10,89
-1,3	1,69
-1,3	1,69
3,7	13,69
2,7	7,29
0,7	0,49
9,7	94,09
2,7	7,29
-1,3	1,69
1,7	2,89
1,7	2,89
Summe	282,3
/ (N - 1)	9,7344828
Wurzel	3,120013


Häufigkeitsdiagramme und Normalverteilung

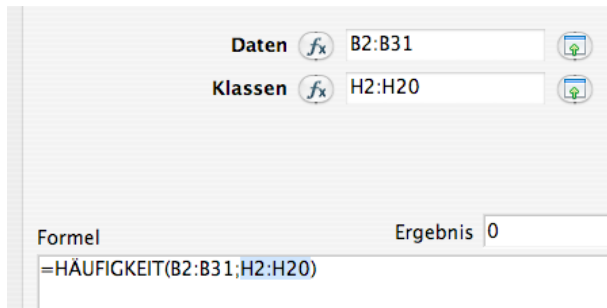
Häufigkeitsdiagramm

Die Daten kann man in einem **Häufigkeitsdiagramm** darstellen, das festhält, wie oft eine bestimmte Anzahl von Vokabeln behalten wurde.

Wir unterscheiden zwischen einfacher und kumulativer Häufigkeit.

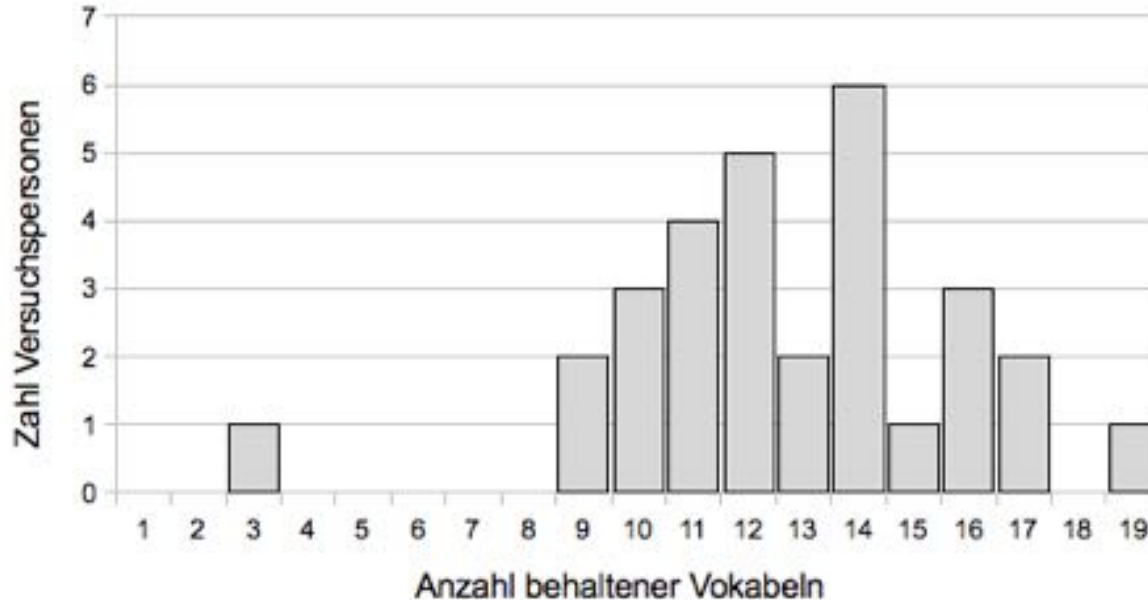
Ein Häufigkeitsdiagramm kann man in LibreOffice wie folgt erzeugen:

- Tragen Sie in einer Spalte, hier H2:H20, die Schwellenwerte für die Klassen ein, hier 1, 2, 3, ... 20.
- Markieren Sie die danebenliegende Spalte.
- Gehen Sie zum Funktionsassistenten , rufen Sie dort die Funktion Häufigkeit auf, tragen Sie dort den Bereich für Daten und Klassen ein.
- Setzen Sie einen Haken bei "Matrix"



H	I
Klassen	Häufigkeiten
1	0
2	0
3	1
4	0
5	0
6	0
7	0
8	0
9	2
10	3
11	4
12	5
13	2
14	6
15	1
16	3
17	2
18	0
19	1
20	0

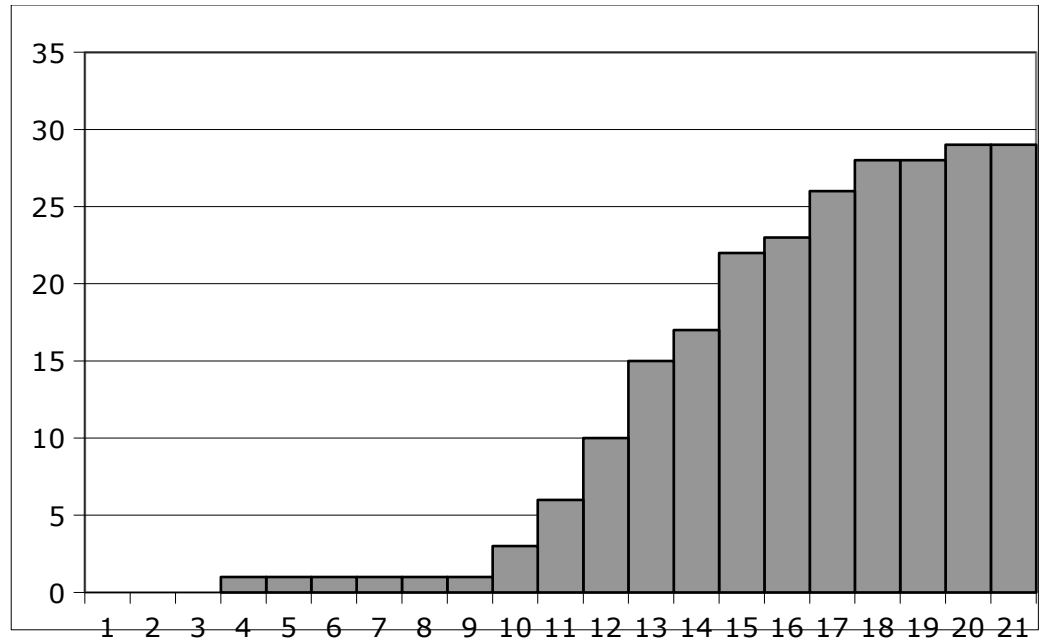
Diese Häufigkeiten können in einem Säulendiagramm dargestellt werden; wir finden hier eine Annäherung an die Normalverteilung, mit Extremwerten („Ausreißern“)



Zur Erinnerung: Mittelwert = 12,7,
MIN = 3, MAX = 19, Median = 12,5, 25%-Quartil = 11, 75%-Quartil = 14
Standardabweichung 3,12, Modalwert = 14.

Kumulative Häufigkeit

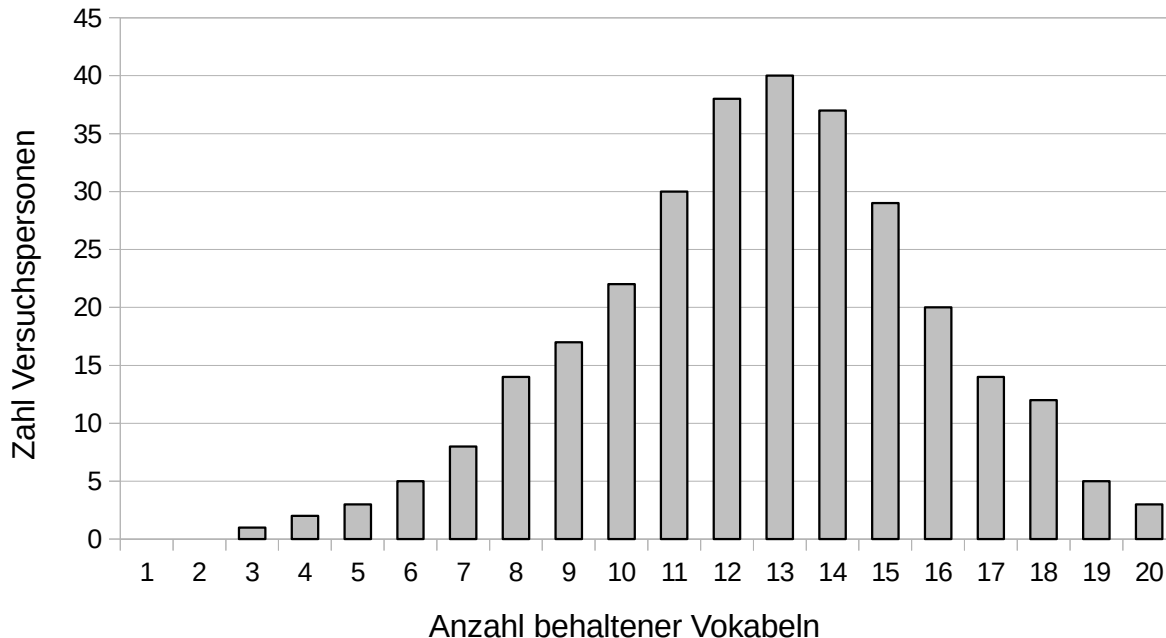
Ein kumulatives Häufigkeitsdiagramm gibt an, wie viele Merkmalsträger Merkmalsausprägungen haben, die gleich oder kleiner als eine gewählte Merkmalsausprägung sind.



Die Größe der Stichprobe

Wir finden sehr häufig:

- Wenn die Größe der Stichprobe erhöht wird, dann erhält das Häufigkeitsdiagramm eine eher „geglättete“ Kontur.
- Typisches Beispiel bei Verzehnfachung der Versuchspersonen auf $N = 300$:



Die Normalverteilung

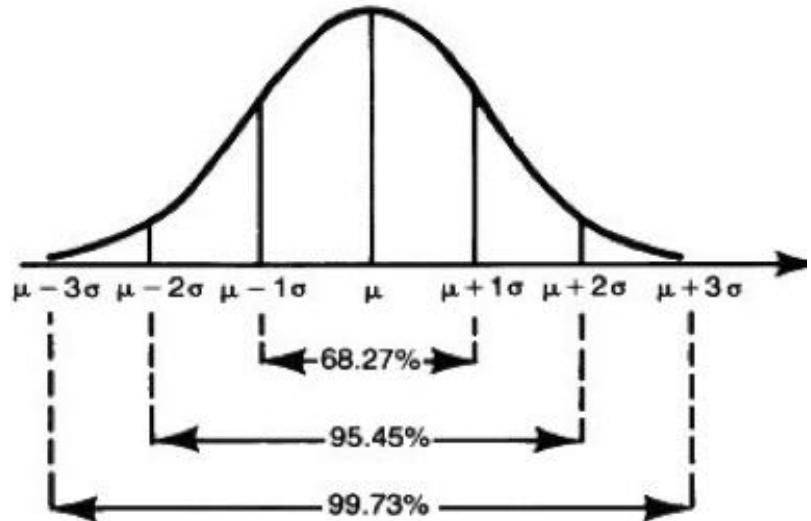
Wir haben hier ein Beispiel dafür, dass bei Vergrößerung der Stichprobe eine **Normalverteilung** entsteht.

Wenn das so ist, dann ist die Standardabweichung wie folgt zu interpretieren:

Ca. 68% der Werte liegen innerhalb der Standardabweichung,

In unserem Beispiel:

- 68% der Werte liegen innerhalb von $12,7 \pm 3,12$.



Normalverteilung und Standard-Abweichung

Messen in Standardabweichungen: z-Scores

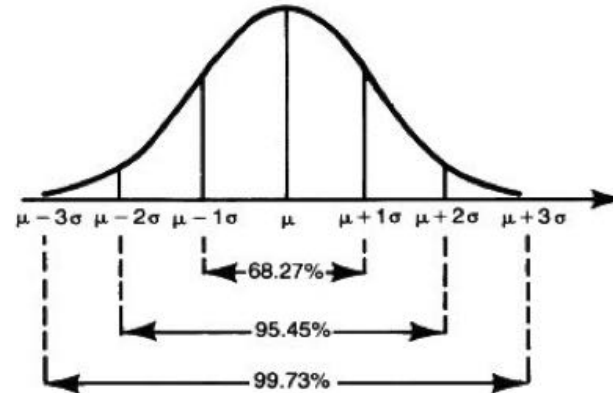
Bei Normalverteilungen können wir den Wert einer einzelnen Messung im Verhältnis zu dem arithmetischen Mittel der Werte angeben, in sog. z-Scores / z-Werten.

Der z-Score eines Messwerts sind seine Abweichungen vom Mittelwert, gemessen in Standardabweichungen.

Berechnung des z-Wertes eines Messwertes x:

$$z = (x - \mu) / \sigma$$

Behaltensleistung	z-Scores
9	-1,19
12	-0,22
19	2,02
13	0,1
16	1,06
14	0,42
10	-0,87
11	-0,54
12	-0,22
12	-0,22
12	-0,22
14	0,42
17	1,38
12	-0,22

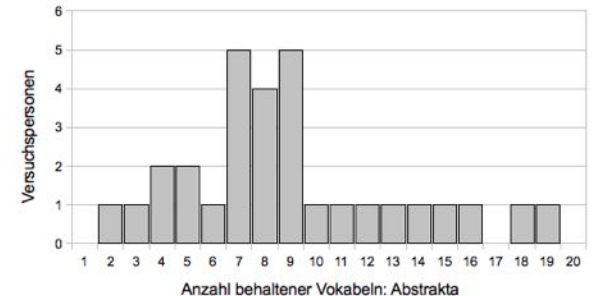
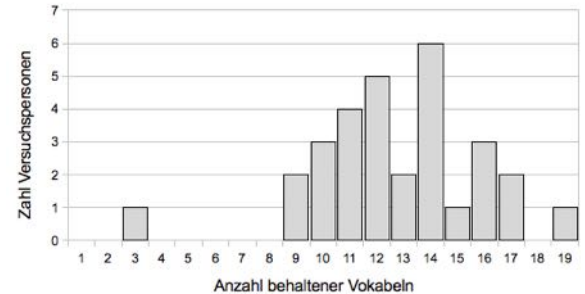


Normalverteilung und Standard-Abweichung

Standardabweichung in Säulendiagrammen

Nehmen wir an, die erste Vokabeluntersuchung geschah mit Wörtern, die konkrete Dinge bezeichnen. Eine zweite mit Vokabeln, die abstrakte Dinge bezeichnen, führte zu dem folgenden Resultat:

VP	Vokabeln	VP	Vokabeln	VP	Vokabeln
1	5	11	9	21	9
2	8	12	19	22	4
3	16	13	7	23	6
4	7	14	18	24	2
5	10	15	9	25	9
6	12	16	15	26	4
7	7	17	8	27	7
8	3	18	7	28	13
9	8	19	11	29	8
10	5	20	14	30	9



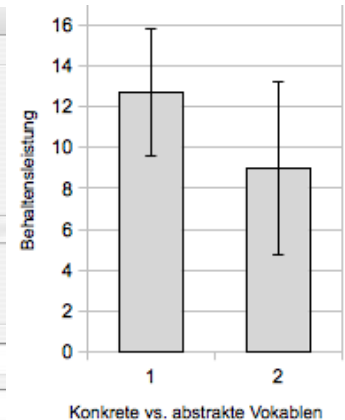
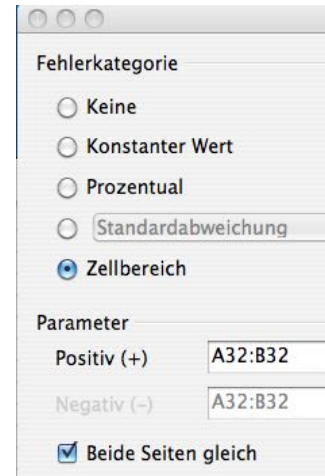
Wir erhalten nun als Mittelwert 8,96 und als Standard-Abweichung 4,24. (Bitte nachrechnen!) Mittelwerte können im Säulendiagramm verglichen werden; Standardabweichung: Strich.

Darstellung in LibreOffice

- Berechne Mittelwerte und die Standardabweichung im Arbeitsblatt.
- Stelle Mittelwerte im Säulendiagramm dar.
- Klicke auf einen Balken, um die Balken zu aktivieren; wähle → Y-FEHLERBALKEN
- Gehe auf ZELLBEREICH; gebe dort die Zellen der Standardabweichung ein (wichtig: NICHT die Option „Standardabweichung“ benutzen – bezieht sich nicht auf Ihre Daten!)
- Eventuell muss man die Y-Achse verlängern, um die Balken richtig darzustellen: durch Klicken auf die Achse und Eingabe entsprechender Werte kann man das tun.
- An dieser Stelle können Sie das Diagramm auch beschriften, die Elemente formatieren usw.

	A	B	C	D
29	11	8		
30	11	9		
31	12,7	8,97	Mittelwert	
32	3,12	4,24	Standardabweichung	
33				
34				

Faustregel: Wenn die Standardabweichungen sich (anders als im Beispiel) nicht überlappen, kann man davon ausgehen, dass der Unterschied signifikant ist, d.h. dass er auch in der Grundgesamtheit besteht.



Darstellung in R: Boxplot-Diagramm

Wir lesen zunächst die zwei Datensätze (insgesamt 60 Daten) auf die Variable `vokabeln` ein, wobei die einzelnen Werte durch Semikolon separiert sind.

```
> vokabeln = scan(file.choose())
```

Hier: Datei `worterinnerung.txt`

Dann wird diese Reihe durch den `dim`-Befehl als ein Array von zwei mal dreißig Elementen umformatiert:

```
> dim(vokabeln) = c(30,2)
```

Der Anfang dieses Arrays wird gezeigt.

Erste Datenspalte: `vokabeln[,1]`,

zweite Datenspalte: `vokabeln[,2]`.

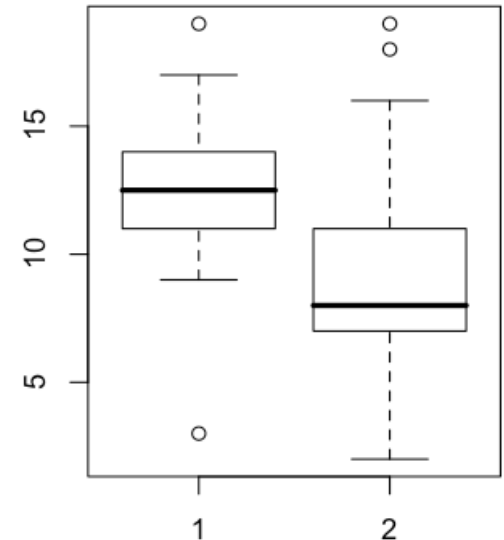
Eine informative Darstellung: **Boxplot-Diagramme.**

```
> boxplot(vokabeln[,1], vokabeln[,2])
```

Boxplots zeigen bei jedem Datensatz an:

- Den Median
- Unteres und oberes Quartil (25%, 75%)
- "Ausreißer"-Daten, durch Kreise markiert.
- Bereich der Nicht-Ausreißer-Daten.

```
      [,1] [,2]  
[1,]    9    5  
[2,]   12    8  
[3,]   19   16  
[4,]   13    7  
[5,]   16   10  
[6,]   14   12  
[7,]   10    7  
[8,]   11    3  
[9,]   15   14  
[10,]  12   11
```



Weiteres zu Boxplot-Diagrammen

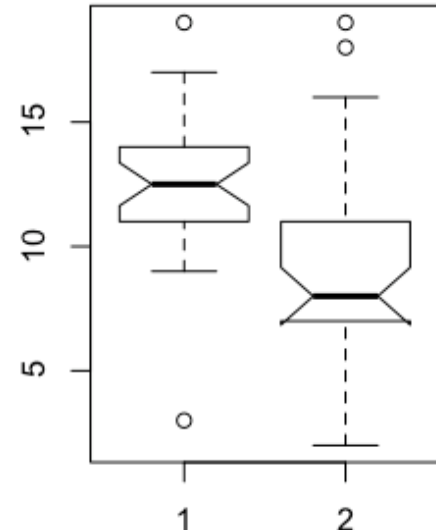
Vorteil von Boxplot-Diagrammen vs. Säulendiagrammen mit Anzeige der Standardabweichung:

- Boxplot-Diagramme machen keine Annahmen einer Normalverteilung in der Grundgesamtheit (die Standardabweichung ist v.a. bei Normalverteilungen ein sinnvolles Streuungsmaß).
- Boxplot-Diagramme sind informativer; sie informieren z.B. über „schiefe“ Verteilungen, der Median ist weniger von Extremwerten abhängig als das arithmetische Mittel.

Weitere Möglichkeiten von Boxplots:

- Boxplots mit Einschnürung („notch“), zur Abschätzung von Signifikanz:
Wenn die Einschnürungen sich nicht überlappen, ist der Unterschied zwischen Stichproben wahrscheinlich für die Grundwahrscheinlichkeit signifikant.
- Boxplots mit variabler Breite:
Zum Anzeigen von Unterschieden in der Sample-Größe,

Beispiel: Diagramm mit Einschnürung;
 Unterschied ist wahrscheinlich signifikant;
 in R: `boxplot(..., notch=TRUE)`
 Diagramm mit Samplegrößen-Angabe;
 hier: Abstrakte Vokabeln eingeschränkt
 auf die ersten zehn Versuchspersonen



Korrelationen

Darstellung von Korrelationen in Streudiagrammen

Wir stellen uns die Frage: Gibt es eine Korrelation zwischen der Erlernbarkeit der beiden Klassen von Vokabeln?

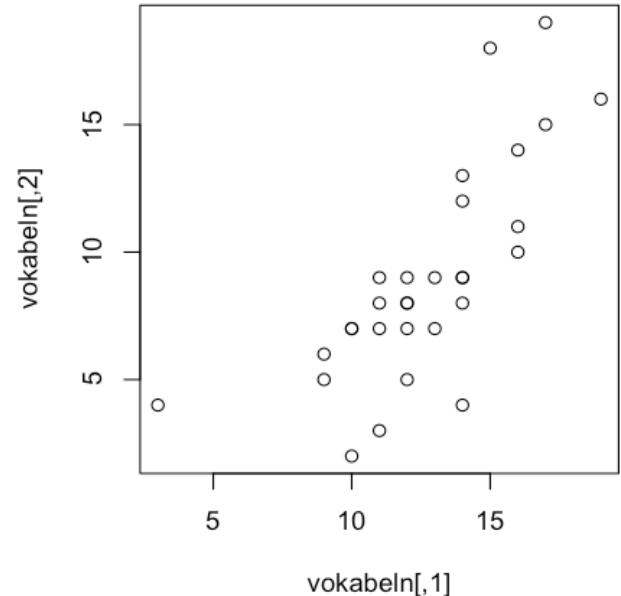
Dies kann man abschätzen, indem man die Zusammenhänge grafisch darstellt.

Mit R genügt hier der einfache Befehl:

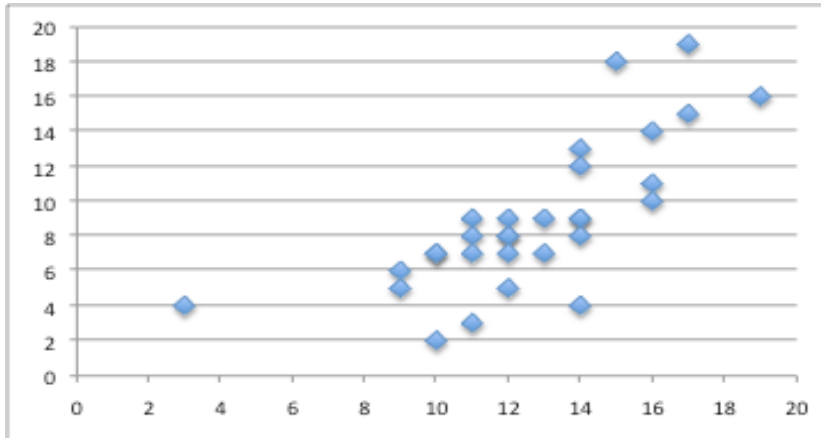
```
> plot(vokabeln)
```

um das linksstehende Streudiagramm zu erzeugen, in welchem für jede Versuchsperson die Leistung in beiden Tests abgetragen ist.

Es besteht ein starker positiver Zusammenhang: Wenn jemand gut bei Vokabeln I ist (horizontale Achse), ist er auch gut bei Vokabeln II (vertikale Achse).



Mit LibreOffice kann man ein solches Streudiagramm erzeugen, indem man die zwei Datenreihen markiert, auf die Schaltfläche für Grafiken drückt und den Diagrammtyp Punkt(X,Y) wählt.



- Positive Korrelation (siehe Beispiel oben): Annäherung durch steigende Gerade.
- Negative Korrelation: Annäherung durch fallende Gerade.
- Keine Korrelation: Annäherung durch Gerade nicht oder kaum möglich.

Man nennt die Gerade auch **Regressionsgerade**. Bei vielen Verteilungen sind andere Regressionskurven angemessen – z.B. exponentiell oder logarithmisch.

Kovarianz und Korrelationskoeffizient

Wir wollen ein Maß für die Korrelation von zwei Datensätzen angeben.

Ein erster Schritt: die **Kovarianz** – ein Maß dafür, wie stark sich zwei Datensätze parallel verändern.

Kovarianz hoch: Wenn der Wert des einen Datensatzes über dem Mittelwert liegt, liegt der Wert des anderen in möglichst gleichem Maße ebenfalls darüber, und umgekehrt.

Berechnungsverfahren für individuellen Datensatz:

$$\triangleright (x_i - \mu_x)(y_i - \mu_y)$$

Hierbei sind x_i und y_i die Behaltensleistungen der Versuchsperson i im ersten bzw. zweiten Vokabeltest, und μ_x und μ_y sind die Mittelwerte der Behaltensleistung aller Versuchspersonen.

Bemerke:

- Weichen x_i und y_i von den Mittelwerten in die gleiche Richtung ab: positiver Wert.
- Weichen x_i und y_i in unterschiedliche Richtungen ab: negativer Wert.
- Je stärker x_i und/oder y_i von den Mittelwerten abweichen, desto größer der Wert.

Wir berechnen die Summe der Werte für alle Versuchspersonen und teilen sie durch die Anzahl $n - 1$ der Versuchspersonen ; dies ist die Kovarianz.

(Oft wird auch der Wert n verwendet, dies ist hier unerheblich).

$$\text{COV}(X,Y) = \frac{\sum (X - \mu_x)(Y - \mu_y)}{n - 1}$$

Die Kovarianz in unserem Beispiel beträgt 9,69.

Berechnung von Kovarianz

Man kann die Kovarianz zwischen zwei Datensätzen auch direkt berechnen durch die Excel-Formel $\text{KOVAR}(X; Y)$, wobei X und Y die Angabe von zwei Datensätzen sind, z.B. $\text{KOVAR}(A1:A30; B1:B30)$.

Problem der Kovarianz als Maßeinheit

Die Kovarianz als Maß hat den Nachteil, dass sie abhängig ist der Einheit, in der gemessen wurde. Wenn etwa die Behaltungsleistung so gemessen würde, dass es für jede gelernte Vokabel 10 Punkte gibt, wäre die Kovarianz 100 mal so groß und betrüge 751,137.

Man kann also durch die Angabe eines Zahlenwertes alleine gar nicht die Korrelation mitteilen.

Der Korrelationskoeffizient

Abhilfe schafft hier der **Korrelationskoeffizient** (hier: der Pearson-Korrelationskoeffizient). Hierzu teilen wir die Kovarianz durch das Produkt der Standardabweichung von X und der Standardabweichung von Y:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

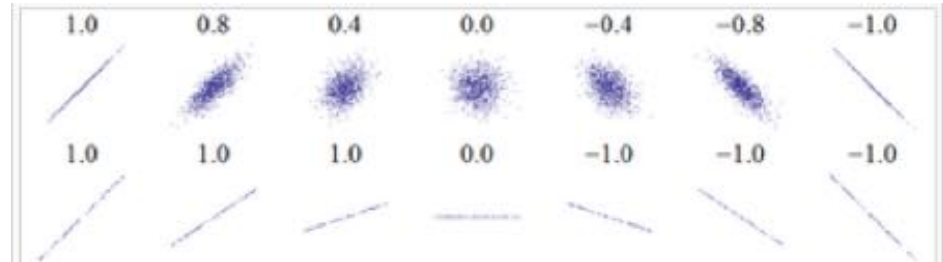
In LibreOffice: PEARSON(A2:A31; B2:B31).

In unserem Beispiel erhalten wir den Wert 0,74. Dies ist ein dimensionsloser Wert,:

- $\rho_{X,Y} = 1$: stärkstmögliche positive lineare Korrelation
- $\rho_{X,Y} = -1$: stärkstmögliche negative lineare Korrelation
- $\rho_{X,Y} = 0$: keine lineare Korrelation

Beispiele für Punktediagramme ("scatterplots") und ihre Korrelationskoeffizienten.

Andere Berechnungsverfahren für exponentielle und logarithmische Korrelationen.



Interpretation von Korrelationen

Es ist wichtig zu sehen, dass mit der Beobachtung einer Korrelation noch kein ursächlicher Zusammenhang **bewiesen** ist.

Ein altes Argument für den Mythos, dass Babies vom Storch gebracht werden, sagt: Die Zahl der Störche ist im 20. Jahrhundert zurückgegangen, die Zahl der Kinder auch, also besteht ein ursächlicher Zusammenhang!

Korrelationen können jedoch zur **Stützung** von Theorien verwendet werden:

Wenn eine Theorie voraussagt, dass die Veränderung einer Größe die Veränderung einer zweiten nach sich zieht und dafür auch einen Grund angibt, dann ist dies ein Argument für diese Theorie und macht die Theorie wahrscheinlicher.

Erklärte Varianz

Die Höhe des Korrelationskoeffizienten gibt an, wie viel der Varianz der abhängigen Variablen durch die unabhängige Variable erklärt werden kann.

➤ ρ^2 ist das Maß, in dem die Wahl der unabhängigen Variable die abhängige bestimmt.

In unserem Beispiel erhielten wir für ρ den Wert 0,74.

Nach Quadrierung erhalten wir den angenäherten Wert 0.55.

Wir können also ca. 50% des Verhaltens einer Versuchsperson in einem Test voraussagen, wenn wir die Versuchsperson und ihre Eigenschaften kennen.

Regressionsgeraden

Oft ist es nützlich, eine Korrelation durch eine mathematische Funktion zu erfassen; diese Funktion kann dann von einer Theorie erklärt werden.

Wenn wir einen **linearen** Zusammenhang annehmen, erhalten wir eine Regressionsgerade.

R stellt für lineare Modelle die Funktion `lm` zur Verfügung, die wie folgt aufgerufen wird:

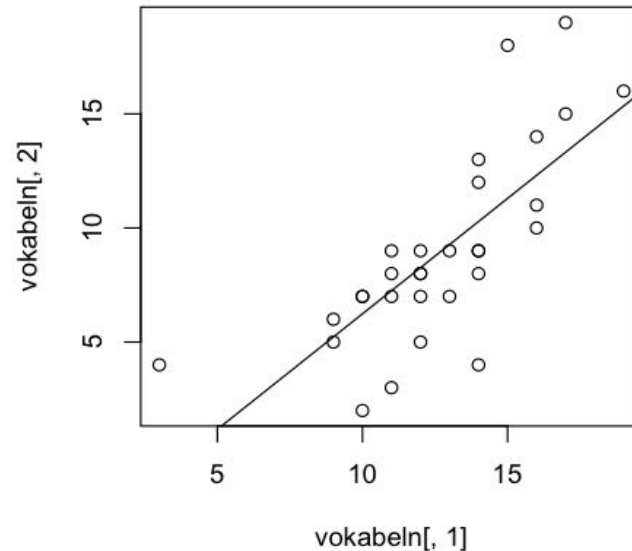
➤ `lm(Variable1 ~ Variable2)` (wobei Variable1 als unabhängige Variable gewertet wird)

Zeichnen einer Regressionsgeraden in R:

➤ `plot([vokabeln[,2]~[vokabeln[,1]])`
`abline(lm(vokabeln[,2]~[vokabeln[,1]))`

Die Regressionsgerade ist die Gerade, für die gilt:

➤ Die Summe der Abstände von jedem Messwertpunkt von der Geraden ist minimiert.



Schließende Statistik

Wir erinnern uns:

- Die **beschreibende** Statistik beschreibt nur die erhobenen Daten (Stichprobe).
- Die **schließende** Statistik überprüft, ob diese diese Daten Rückschlüsse auf die Verhältnisse in der Grundgesamtheit erlauben.

Wir werden hier einige einfache Tests für die Typen von Daten kennenlernen, wie sie häufiger in linguistischen Untersuchungen anfallen.

Der Chi-Quadrat-Test (χ^2 -Test)

Dieser Test eignet sich für **Nominaldaten**, in denen die Merkmalsausprägungen keiner bestimmten Ordnung unterliegen.

Das Testverfahren anhand eines Beispiels

In dem 30-Sprachen-Sample von Greenberg (1966) gibt es folgende Beziehung zwischen den Grundwortstellungen SOV, SVO und VSO und der Voran- bzw. Nachstellung von Adpositionen (Präpositionen: P-NP, Postpositionen: NP-P).

	SOV	SVO	VSO	Σ
P-NP	0	10	6	16
NP-P	11	3	0	14
Σ	11	13	6	30

Σ : Summe

Offensichtliche Tendenz: VSO-Sprachen haben Präpositionen,
SOV-Sprachen haben Postpositionen,
SVO-Sprachen haben einer Tendenz zu Präpositionen.

Frage: Trifft die im Sample beobachtete Tendenz auch für die Grundgesamtheit zu?

Dazu gibt der χ^2 -Test (der **Chi-Quadrat-Test**) eine Schätzung.

Überlegungen:

Wenn es **keine** Korrelation zwischen den Merkmalsausprägungen gäbe (die **Nullhypothese**),

würden wir folgendes Resultat als Mittelwert von vielen Stichproben von jeweils 30 Sprachen erwarten (der **Erwartungswert** unter der Nullhypothese).

	SOV	SVO	VSO	Σ
P-NP	$\frac{16 \times 11}{30}$	$\frac{16 \times 13}{30}$	$\frac{16 \times 6}{30}$	16
NP-P	$\frac{14 \times 11}{30}$	$\frac{14 \times 13}{30}$	$\frac{14 \times 6}{30}$	14
Σ	11	13	6	30

Das ergibt die folgenden gerundeten Werte:

	SOV	SVO	VSO
P-NP	5,87	6,93	3,2
NP-P	5,13	6,07	2,8

Die tatsächlichen Werte der Stichprobe weichen davon folgt ab,
wobei wir jeweils von dem beobachteten Wert grün) den Erwartungswert abziehen:

	SOV	SVO	VSO
P-NP	0 – 5,87	10 – 6,93	6 – 3,2
NP-P	11 – 5,13	3 – 6,07	0 – 2,8

Wir erhalten das folgende Ergebnis für diese Abweichungen.
(Die Zahlen in der oberen und unteren Zeile entsprechen einander, da es nur zwei Zeilen gibt).

	SOV	SVO	VSO
P-NP	- 5,87	3,07	2,8
NP-P	5,87	- 3,07	- 2,8

Um einheitlich positive Zahlen zu erhalten, quadrieren wir die Zahlen in jeder Zelle:

	SOV	SVO	VSO
P-NP	34,46	9,42	7,84
NP-P	34,46	9,42	7,84

Diese Zahlen messen die Abweichung der beobachteten Werte
von den bei der Nullhypothese zu erwartenden.

Wenn wir diese Zahlen als Maß für die Abweichung vom Erwartungswert verwenden wollen, müssen wir noch den Erwartungswert selbst berücksichtigen.

Wir teilen daher die Zellen durch den Erwartungswert...

	SOV	SVO	VSO
P-NP	<u>34,46</u> 5,87	<u>9,42</u> 6,93	<u>7,84</u> 3,2
NP-P	<u>34,46</u> 5,13	<u>9,42</u> 6,07	<u>7,84</u> 2,8

und erhalten die folgende Matrix:

	SOV	SVO	VSO
P-NP	5,87	1,36	2,45
NP-P	6,77	1,55	2,8

Wir summieren nun die Werte und erhalten 20,8.

Dies ist ein Maß für die Abweichung der Stichprobe von dem Wert, der bei der Nullhypothese zu erwarten gewesen wäre.

Je größer er ist, desto größer die Abweichung.

Berechnung des Chi-Quadrat-Wertes

Der χ^2 -Wert wird also wie folgt errechnet:

$$\chi^2 = \sum \frac{(\text{beobachteter} - \text{erwarteter Wert})^2}{\text{erwarteter Wert}}$$

Der Chi-Quadrat-Wert zur Abschätzung der Stichprobenwahrscheinlichkeit

Mit dem erhaltenen Wert verfahren wir wie folgt:

Wir berechnen die **Freiheitsgrade**.

- Bei den Spalten gibt es 2 Freiheitsgrade: Wenn die Werte von zwei Spalten festgelegt sind und die Gesamtsumme der Spalten bekannt ist, dann ist auch der Wert der dritten Spalte bekannt.
- Bei den Zeilen gibt es 1 Freiheitsgrad: Wenn der Wert einer Zeile bekannt ist und die Gesamtsumme der Zeilen, dann ist auch der Wert der anderen Zeile bekannt.
- Wir multiplizieren diese beiden Werte und erhalten **2 Freiheitsgrade**.

Die Vorstellung der Freiheitsgrade bezieht sich also darauf:

Wie viel von der spezifischen Verteilung muss man wissen, sodass der Rest festgelegt ist?

Bei einer 3x2-Matrix genügen hierfür 2 Zellen, wie man selbst ausprobieren kann:

10			30
	5		20
15	10	25	50

Wir können nun in einer χ^2 -Tabelle nachsehen, wie unwahrscheinlich es wäre, dass wir die Stichprobe erhalten, wenn die Daten in der Grundgesamtheit zufällig verteilt wären. Wir verwenden hier die Tabelle aus Woods e.a. (1986), *Statistics in Language Studies*, Cambridge Textbooks in Linguistics).

Bei 2 Freiheitsgraden ist ein χ^2 -Wert von 20,8 nur mit einer (viel) geringeren Wahrscheinlichkeit als 0,1% oder 0,001 zu erwarten ist.

Das heißt:

Weniger als eine von 1000 Stichproben würde einen solchen Chi-Quadrat-Wert liefern.

Dass es zwischen den zwei Wortstellungs-eigenschaften eine Beziehung gibt, steht damit mit hoher Wahrscheinlichkeit fest.

Wir sprechen von einem **Signifikanzniveau** von 0.001.

Degrees of freedom	p=							
	97.5	95	50	10	5	2.5	1	0.1
1	0.000982	0.00393	0.45	2.71	3.84	5.02	6.64	10.8
2	0.0506	0.103	1.39	4.61	5.99	7.38	9.21	13.8
3	0.216	0.352	2.37	6.25	7.82	9.35	11.3	16.3
4	0.484	0.711	3.36	7.78	9.49	11.1	13.3	18.5
5	0.831	1.15	4.35	9.24	11.1	12.8	15.1	20.5
6	1.24	1.64	5.35	10.6	12.6	14.5	16.8	22.5
7	1.69	2.17	6.35	12.0	14.1	16.0	18.5	24.3
8	2.18	2.73	7.34	13.4	15.5	17.5	20.1	26.1
9	2.70	3.33	8.34	14.7	16.9	19.0	21.7	27.9
10	3.25	3.94	9.34	16.0	18.3	20.5	23.2	29.6
12	4.40	5.23	11.3	18.5	21.0	23.3	26.2	32.9
15	6.26	7.26	14.3	22.3	25.0	27.5	30.6	37.7
20	9.59	10.9	19.3	28.4	31.4	34.2	37.6	45.3
24	12.4	13.9	23.3	33.2	36.4	39.4	43.0	51.2
30	16.8	18.5	29.3	40.3	43.8	47.0	50.9	59.7
40	24.4	26.5	39.3	51.8	55.8	59.3	63.7	73.4
60	40.5	43.2	59.3	74.4	79.1	83.3	88.4	99.6

Die Größe der Stichprobe und ihre Signifikanz

Die Größe der Stichprobe hat einen Einfluss darauf, wie signifikant sie ist.

➤ ähnlich wie die Frage, ob ein Würfel gezinkt ist, es ratsam ist, den Würfel mehrfach zu werfen!

Beispiel: Stellen wir uns vor, Greenberg hätte eine Stichprobe mit nur 10 Sprachen gezogen, mit einem sonst ähnlichen Ergebnis:

	SOV	SVO	VSO	Σ
P-NP	0	3	2	5
NP-P	4	1	0	5
Σ	4	4	2	10

Wir berechnen daraus einen Chi-Quadrat-Wert von 6.

Wenn wir in der Tabelle nachsehen, finden wir einen p-Wert zwischen 0.10 und knapp vor 0.5.

Es wäre also unter der Nullhypothese zu erwarten, dass mindestens eine von zwanzig Stichproben einen solchen Ch-Quadrat-Wert liefert.

Dies gilt **nicht** mehr als ein signifikantes Ergebnis (p mindestens < 0.5).

Nullhypothese und theoretisch motivierte Hypothese

- In unserem Beispiel arbeiteten wir mit zu erwartenden Werten unter der Annahme, dass es keine Interaktion von Grundwortstellung und Adpositionstyp gibt. (**Nullhypothese**).
- Wir könnten auch Werte eingeben, die nach einer bestimmten Theorie zu erwarten wären, und überprüfen, wie gut die beobachteten Werte den theoretisch vorausgesagten entsprechen.

Chi-Quadrat-Test mit kleinen Werten

Ein Problem des Beispiels ist, dass die Werte in den Zellen recht niedrig sind.

Der Chi-Quadrat-Test wird nur dann empfohlen, wenn alle Erwartungswerte ≥ 5 sind, was in unserem Beispiel bei den VSO-Sprachen nicht der Fall ist.

Wenn wir nur **einen** Freiheitsgrad haben, gibt es eine Sonderregelung (die Yates-Korrektur).

$$\chi^2 = \sum \frac{(|\text{beobachteter} - \text{erwarteter Wert}| - 0,5)^2}{\text{erwarteter Wert}}$$

Berechnung von Chi-Quadrat mit Excel und LibreOffice

geschieht durch die Funktion CHITEST, welche bei Eingabe einer Matrix von beobachteten Werten und von zu erwartenden Werten das Signifikanzniveau liefert.

CHITEST(Beobachtete-Werte; Erwartete-Werte).

Für unser Beispiel müssen wir die beobachteten Daten eintragen und die Erwartungswerte errechnen:

	A	B	C	D	E
1					
2	Beobachtungsdaten				
3		SOV	SVO	VSO	
4	P-NP	0,00	10	6	16
5	NP-P	11,00	3	0	14
6		11,00	13	6	30
7					
8	Erwartungswerte				
9		5,87	6,93	3,20	
10		5,13	6,07	2,80	
11					

Dann können wir in eine Zelle den Befehl CHITEST(B4:D5; B9:D10) eingeben

Für unser Beispiel erhalten wir 3,155E-05, d.h. $3,155 \times 10^{-5}$, also 0,00003155.

Dies ist die Wahrscheinlichkeit, dass die Verhältnisse der Stichprobe „zufällig“ eine Beziehung zwischen den Wortstellungsmustern suggerieren.

Dieser Wert ist sehr klein, die Nullhypothese kann zurückgewiesen werden.

Berechnung von Chi-Square mit R

Mit R ist der Chi-Square-Test sehr einfach durchzuführen.

Wir geben mit `scan` die Daten ein
und konfigurieren sie mit `dim` als 3 × 2-Matrix.

Wir wenden darauf `chisq.test` an:

```
> chisq.test(wortstellung)

Pearson's Chi-squared test

data: wortstellung
X-squared = 20.728, df = 2, p-value = 3.155e-05
```

Hier steht `df` = „degrees of freedom“ – Freiheitsgrade.

Wir erhalten eine Warnung, weil die Zahlenwerte eigentlich zu klein sind,
um den Chi-Quadrat-Test sicher anzuwenden.

```
> wortstellung = scan()
1: 0
2: 11
3: 10
4: 3
5: 6
6: 0
7:
Read 6 items
> wortstellung
[1] 0 11 10 3 6 0
> dim(wortstellung) = c(2, 3)
> wortstellung
      [,1] [,2] [,3]
[1,]    0  10    6
[2,]   11    3    0
```


Bemerkung zu Signifikanzniveaus

In Berichten zu Experimenten werden üblicherweise nicht die genauen Wahrscheinlichkeiten angegeben, nach denen ein Ergebnis als nur zufällig entstanden bewertet werden kann, sondern es wird ein **Signifikanzniveau** α wiedergegeben. Dabei gilt:

- $p \geq 0.05$: **nicht signifikant**; in einer von mind. 20 Stichproben ist das Ergebnis auch dann zu erwarten, wenn die Nullhypothese zutrifft.
- $p < 0.05$ (und $0.01 \leq p$): **signifikant**; das Ergebnis ist in weniger als 20 Stichproben zu erwarten, wenn die Nullhypothese zutrifft. Die Irrtumswahrscheinlichkeit beträgt weniger als 5%.
- $p < 0.01$ (und $0.001 \leq p$): **hoch signifikant**; das Ergebnis ist in weniger als 100 Stichproben zu erwarten, wenn die Nullhypothese zutrifft. Die Irrtumswahrscheinlichkeit beträgt weniger als 1%.
- $p < 0.001$: **höchst signifikant**; das Ergebnis ist in weniger als 1000 Stichproben zu erwarten. Die Irrtumswahrscheinlichkeit beträgt weniger als 0.1%.

Signifikante, hoch signifikante und höchst signifikante Ergebnisse werden mit einem *, zwei ** bzw. drei Sternchen *** angeben; nicht signifikante Beziehungen durch "n.s."

Beispiel: Präsentation eines Testergebnisses

Ein Chi-Quadrat-Test zeigte, dass es einen hoch signifikanten Zusammenhang zwischen der Stellung von S, V und O einerseits und der Vor- oder Nachstellung von Adpositionen andererseits gibt ($\chi^2 = 20.728$, $df = 2$, $p < 0.0001$).

Die Stärke der Interaktion

Ein signifikantes Ergebnis im χ^2 -Test sagt nicht, **wie stark** die Ausprägung des einen Merkmals die Ausprägung eines anderen entspricht. Dies wird durch den statistischen Wert **Cramérs V** bestimmt, der wie folgt berechnet wird, wobei N die Gesamtzahl der Beobachtungen und k die kleinere Zahl der Spalten oder Zeilen ist.

$$\text{Cramérs } V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

In unserem Beispiel ergibt sich der Wert: Quadratwurzel aus $20,8 / (30 \times 1) = 0,83$.

Dies ist ein hoher Effekt. Wir haben im allgemeinen:

Werte für Cramérs V:

- $0,10 < \text{Cramérs } V < 0,30$: geringer Effekt
- $0,30 \leq \text{Cramérs } V < 0,50$: mittlerer Effekt
- $0,50 \leq \text{Cramérs } V$: erheblicher Effekt

Ein weiteres Beispiel:

Zusammenhang Adj-Nomen und Objekt-Verb-Stellung, ein viel größeres Sample wurde betrachtet.

Chi-Quadrat-Test zeigt: Hypothese noch wahrscheinlicher, $p = < 3 * 10^{-13}$

Aber Cramérs V zeigt: Interaktion ist schwächer, 0,232

	OV	VO	
Adj-N	201	100	301
N-Adj	287	404	691
	488	504	992

Der t-Test für zwei Stichproben

Der t-Test eignet sich für **Intervalldaten**.

Wir kommen zurück zu unserem Beispiel der Beobachtung von Behaltensleistung für Vokabeln, die konkrete oder abstrakte Dinge bezeichnen. Wir haben dort festgestellt: Die Mittelwerte für Konkreta und Abstrakta sind verschieden (12,70 vs. 8,97).

Wir fragen uns: Wie groß ist die Wahrscheinlichkeit, dass die Verschiedenheit der Mittelwerte in der **Stichprobe** einer Verschiedenheit der Mittelwerte in der **Grundgesamtheit** entspricht?

Das heißt: Wie wahrscheinlich ist es aufgrund des Versuchs, dass Vokabeln mit konkreten Bedeutungen tatsächlich besser gelernt werden als Vokabeln mit abstrakten Bedeutungen?

Genauer: Man prüft, welche der beiden Hypothesen gerechtfertigt sind:

- H_0 , die Nullhypothese: Die Mittelwerte sind in der Grundgesamtheit gleich.
- H_1 : Die Mittelwerte sind in der Grundgesamtheit nicht gleich.

Das beantwortet der **t-Test für zwei Stichproben**.

Wir machen dabei die Voraussetzung, dass diese Werte in der Grundgesamtheit **normalverteilt** sind.

(Eine Kuriosität: Der Test wurde in der Guinness-Brauerei in Dublin 1908 zur Überwachung der Bierqualität von dem Mathematikstudenten W. Gosset erfunden, der dort in den Ferien jobbte. Die Brauerei erlaubte es nicht, dass er ihn unter eigenem Namen veröffentlichte; seitdem ist er unter Student's t-Test bekannt.)

Z	AA Kon- kreta	AB Ab- strakta
	9	5
	12	8
	19	16
	13	7
	16	10
	14	12
	10	7
	11	3
	12	8
	12	5
	14	9
	17	19
	12	7
	15	18
	13	9
	17	15
	14	8
	11	7
	16	11
	16	14
	14	9
	14	4
	9	6
	10	2
	12	9
	3	4
	10	7
	14	13
	11	8
	11	9
Mittelwert	12,7	8,967

Beispiel für t-Test: Abhängige Stichproben, LibreOffice Calc

In unserem Beispiel haben wir eine **abhängige Stichprobe**: Wir haben für jede Versuchsperson sowohl die Behaltensleistung für konkrete als auch für abstrakte Vokabeln gemessen.

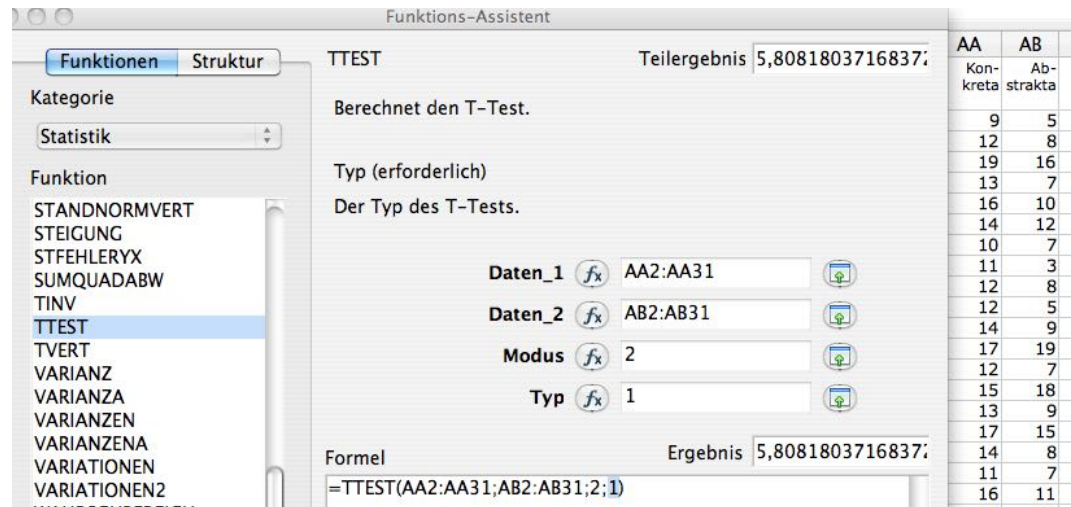
Mit LibreOffice kann der Test mithilfe des Funktionsassistenten durchgeführt werden.

Modus 1 steht für den einseitigen, **Modus 2** für den zweiseitigen Test – wenn man keine Hypothese hat, nach welcher Seite die Abweichung erfolgen sollte.

Typ 1 steht für den Test mit **abhängiger Stichprobe** (gepaart).

Wir erhalten als Ergebnis den Wert $5,8 \cdot 10^{-8}$ als Wahrscheinlichkeit, dass die Nullhypothese stimmt; diese kann daher mit großer Wahrscheinlichkeit (Signifikanzniveau ***) zurückgewiesen werden.

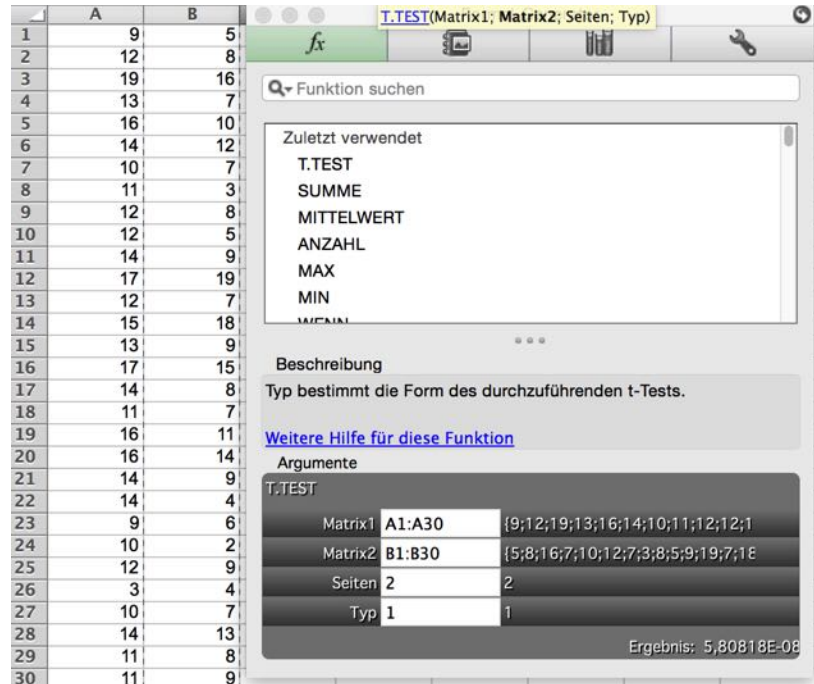
Beachte: Der **einseitige** p-Wert liegt noch niedriger (man könnte ihn verwenden, wenn man von vornherein die Hypothese hat, dass Vokabeln mit konkreten Denotaten schlechter behalten werden).



AA	AB
Konkreta	Abstrakta
9	5
12	8
19	16
13	7
16	10
14	12
10	7
11	3
12	8
12	5
14	9
17	19
12	7
15	18
13	9
17	15
14	8
11	7
16	11

Der t-Test, abhängige Stichprobe, mit Excel

Mit Microsoft Excel kann man den T-Test ebenfalls über den Funktionsassistenten berechnen.



The screenshot shows an Excel spreadsheet with two columns of data, A and B, and the T.TEST function assistant dialog box open. The dialog box displays the function name, a description, and the arguments used for the calculation.

	A	B
1	9	5
2	12	8
3	19	16
4	13	7
5	16	10
6	14	12
7	10	7
8	11	3
9	12	8
10	12	5
11	14	9
12	17	19
13	12	7
14	15	18
15	13	9
16	17	15
17	14	8
18	11	7
19	16	11
20	16	14
21	14	9
22	14	4
23	9	6
24	10	2
25	12	9
26	3	4
27	10	7
28	14	13
29	11	8
30	11	9

T.TEST(Matrix1; Matrix2; Seiten; Typ)

Funktion suchen

Zuletzt verwendet

- T.TEST
- SUMME
- MITTELWERT
- ANZAHL
- MAX
- MIN
- WECHSEL

Beschreibung

Typ bestimmt die Form des durchzuführenden t-Tests.

[Weitere Hilfe für diese Funktion](#)

Argumente

T.TEST

Matrix1	A1:A30	{9;12;19;13;16;14;10;11;12;12;1
Matrix2	B1:B30	{5;8;16;7;10;12;7;3;8;5;9;19;7;18
Seiten	2	2
Typ	1	1

Ergebnis: 5,80818E-08

t-Test mit R

Hierfür rufen wird die folgende Funktion auf. Die ersten beiden Argumente geben die beiden Datenreihen wieder, das dritte Argument bestimmt, dass es sich um den zweiseitigen t-Test handelt.

```
> t.test(vokabeln[,1], vokabeln[,2], paired=TRUE)
```

```
Paired t-test
```

```
data: vokabeln[, 1] and vokabeln[, 2]
t = 7.2316, df = 29, p-value = 5.808e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.677485 4.789182
sample estimates:
mean of the differences
      3.733333
```

Hier wird der Wert von t (= 7,2316) angegeben (für weitere Berechnungen wichtig), die Freiheitsgrade (df = 29, bei 30 Beobachtungen) und der p-Wert (5.808×10^{-8}).

Der t-Test: Unabhängige Stichproben

Nehmen wir an, wir hätten nicht die gleichen Versuchspersonen zwei Mal getestet, sondern **zwei Gruppen von Versuchspersonen**:

Eine für konkrete, eine für abstrakte Wörter.

Wir hätten dann **unabhängige** Stichproben.

In unserem Beispiel ist es sicher besser, abhängige Stichproben zu bilden, da man den Unterschied im Lernverhalten zwischen konkreten und abstrakten Wörtern am besten jeweils innerhalb einer Person misst.

Oft sind abhängige Stichproben aber nicht möglich,

z.B. wenn man verschiedene Lehrmethoden miteinander vergleichen will, oder weil man verschiedene Personen miteinander vergleichen will.

Bei unabhängigen Stichproben kann man nicht mehr einzelne Individuen messen, sondern muss Mittelwerte von unterschiedlichen Gruppen in Verbindung zueinander bringen.

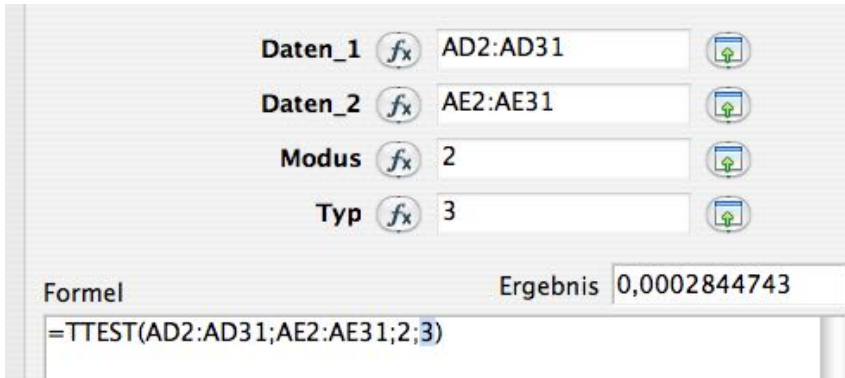
Es gibt Varianten des Tests für Stichproben gleicher oder verschiedener Größe und für die Annahme, dass die Populationen die gleiche oder verschiedene Varianz besitzen.


t-Test unabhängige Stichproben mit LibreOffice Calc


Nehmen wir dasselbe Beispiel wie zuvor, jetzt aber anders interpretiert:
Es wird die Behaltensleistung bei gleichen Vokabeln gemessen,
wir haben jetzt aber zwei Gruppen: **20-jährige** und **40-jährige**:


In unserem Fall sind die **beiden Stichproben gleich groß**,
und wir nehmen an, dass die **Varianzen** in den beiden Grundgesamtheiten
unterschiedlich sein können.


Mit LibreOffice erhalten wir folgendes Resultat:



Daten_1 

Daten_2 

Modus 

Typ 

Formel Ergebnis

Hier steht **Typ 3** für nicht-gepaarte (unabhängige) Stichproben,
ungleiche Varianz (genauer: Keine Annahme gleicher Varianz).
Der p-Wert liegt unter 0,001, das Ergebnis ist höchst signifikant (***).

AD	AE
20-jährige	40-jährige
9	5
12	8
19	16
13	7
16	10
14	12
10	7
11	3
12	8
12	5
14	9
17	19
12	7
15	18
13	9
17	15
14	8
11	7
16	11
16	14
14	9
14	4
9	6
10	2
12	9
3	4
10	7
14	13
11	8
11	9
12,7	8,9666667

Berechnung mit R

Hier werden als Parameter `paired=FALSE` und `var.equal=FALSE` (möglicherweise nicht gleiche Varianz) eingegeben. Es wird dann der Welch-Test durchgeführt, der sicherer (konservativer) ist und ein leicht abweichendes Resultat ergibt:

```
> t.test(vokabeln[,1], vokabeln[,2],paired=FALSE, var.equal=FALSE)
```

```
Welch Two Sample t-test
```

```
data: vokabeln[, 1] and vokabeln[, 2]
t = 3.8853, df = 53.295, p-value = 0.0002845
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.806296 5.660370
sample estimates:
mean of x mean of y
12.700000  8.966667
```

Auch hier wird die Nullhypothese mit einer Wahrscheinlichkeit von 0,0002845 zurückgewiesen.

Einfaktorielle Varianzanalyse

Betrachten wir folgendes hypothetisches Beispiel:

Es wurden 3 Versionen eines Textes erarbeitet, wobei

- A die Ausgangsversion,
- B eine Version mit kürzeren Sätzen und
- C eine Version ist, in der einige Pronomina durch ihre Antezedens-Ausdrücke ersetzt wurden. (z.B. *Hans kaufte ein Auto. Er ist davon / von dem Auto begeistert*).

Es wurde die Lesezeit der Texte in Sekunden (natürlich bei verschiedenen Versuchspersonen, jeweils 16, insgesamt 48) gemessen.

Texttyp; **unabhängige** Variable,
Lesezeit; **abhängige** Variable.

```
> leseleistung
  col1 col2 col3 col4 col5 col6 col7 col8 col9 col10 col11 col12 col13 col14 col15 col16
A   61   59   66   68   55   70   71   62   66   74   69   63   65   62   64   66
B   57   55   56   63   51   54   50   60   54   61   61   59   59   60   54   54
C   61   58   62   68   62   66   51   65   73   63   70   60   61   59   57   62
```

Hypothesen:

- H_0 : Die Mittelwerte der Lesezeiten der Texte unterscheiden sich nicht in der Grundgesamtheit
- H_1 : Es gibt einen Unterschied zwischen den Mittelwerten

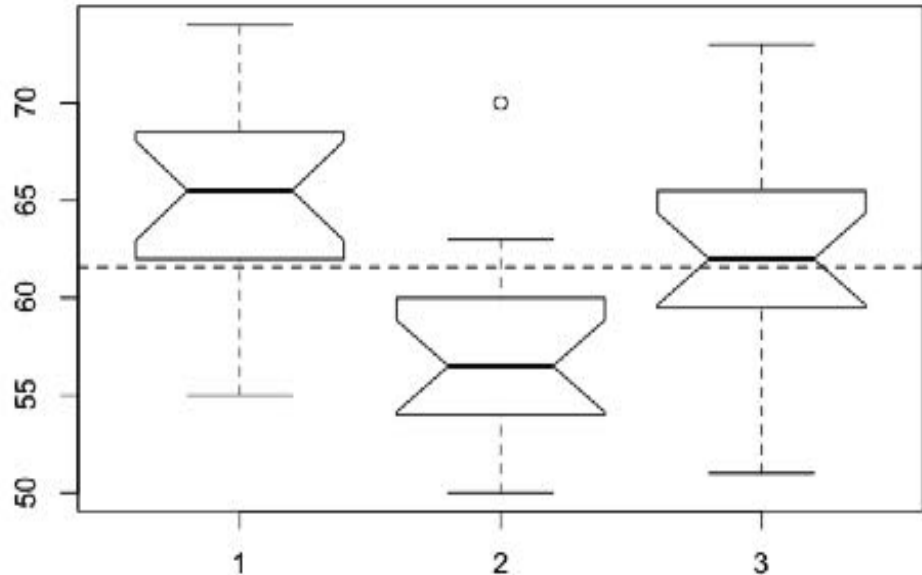
Wir betrachten die Lesegeschwindigkeiten der drei Texte zunächst mithilfe von Boxplots:

- `> boxplot(leseleistung[1,], leseleistung[2,], leseleistung[3,], notch=TRUE)` ; Boxplot
- `> abline(mean(leseleistung), 0, lty=2)` ; fügt Mittelwertlinie zu

Der Unterschied zwischen Texttyp 1 und Texttyp 2 scheint signifikant zu sein; der Unterschied zwischen 2 und 3 ist möglicherweise signifikant, der zwischen 1 und 3 aber eher nicht (Überlappung der Einkerbungen)

Die Boxplots zeigen: Die Daten sind in etwa normalverteilt, d.h. nicht besonders asymmetrisch;

daher kann die Varianzanalyse angenommen werden.



Zur Durchführung einer ANOVA muss die Matrix „leseleistung“ noch in eine geeignete Form gebracht werden:

- Wir kombinieren die Daten zu einem einzigen langen Vektor mit
> leseleistungsliste <- c(leseleistung[“A”,.], leseleistung[“B”,.], leseleistung[“C”,.])
- Wir erzeugen einen Vektor mit 16 Einsen, 16 Zweien, 16 Dreien:
> n <- rep(16,3) (n ist ein Vektor, in dem 16 dreimal wiederholt wird)
> texttyp <- rep(1:3, n) (ein Vektor, in dem die Zahlen 1, 2, 3 je 16mal wiederholt werden)

Bildschirmfoto:

```
> n
[1] 16 16 16
> texttyp
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3
```

Wir schaffen nun einen Data-Frame:

```
> lesedaten <- data.frame(lesesec = leseleistungsliste, texttypen = factor(texttyp))
```

Damit wird die folgende Datenstruktur erzeugt:

```
> lesedaten
  lesesec texttypen
1      61         1
2      59         1
3      66         1
4      68         1
5      55         1
6      70         1
7      71         1
8      62         1
~      ~         -
~~~~~
44     60         3
45     61         3
46     59         3
47     57         3
48     62         3
```

- Berechnung der ANOVA:
 - > `anova(lm(lesesec ~ texttypen, lesedaten))` – hier wird lesesec als abhängige, texttypen als unabhängige Variable gewertet
 - Ergebnis (gleichwertig):
 - Es gibt einen Unterschied zwischen den Gruppen; H_0 wird zurückgewiesen.
 - Df: Freiheitsgrade (3 Gruppen, 46 Beobacht.)
 - Mean Sq: Mittlere Qu.Summe, Varianzmaß für Beobachtungen (Residuals)
 - F-value: Prüfgröße für Signifikanz
 - Pr: Wahrscheinlichkeit, dass dieser F-Wert zufällig ist; hier sehr klein
 - ***: Signifikanzniveau; höchst signifikant
 - Bericht des Ergebnisses: „Einfaktorielle Varianzanalyse $F(2, 45) = 9.8, p < 0.001$ “
 - Genauere Betrachtung mit dem TukeyHSD-Test:
 - höchst signifikant: Gruppe 2-1
 - signifikant: Gruppe 3-3
 - nicht signifikant: Gruppe 3-1; diff: Median-Unterschiede; lwr, upr: lower, upper level des 95% Konfidenz-Intervalls
 - p: Signifikanz
- ```

> summary(aov(lesesec~texttypen, lesedaten))
 Df Sum Sq Mean Sq F value Pr(>F)
texttypen 2 495.54 247.771 9.7965 0.0002938 ***
Residuals 45 1138.12 25.292

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> TukeyHSD(aov(lesesec~texttypen, lesedaten))
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = lesesec ~ texttypen, data = lesedaten)

$texttypen
 diff lwr upr p adj
2-1 -7.7500 -12.0593038 -3.440696 0.0002182
3-1 -2.6875 -6.9968038 1.621804 0.2952178
3-2 5.0625 0.7531962 9.371804 0.0178426

```

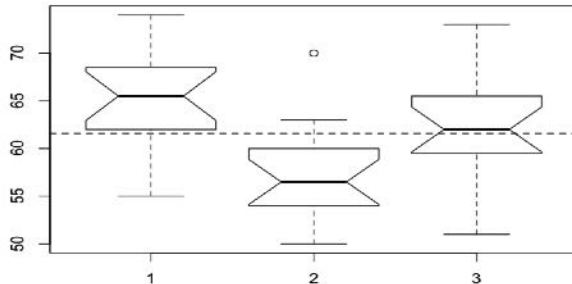
Darstellung des Konfidenzniveaus für die Gruppenvergleiche:

➤ `plot(TukeyHSD(aov(lesesc ~ textypen, lesedaten)))`

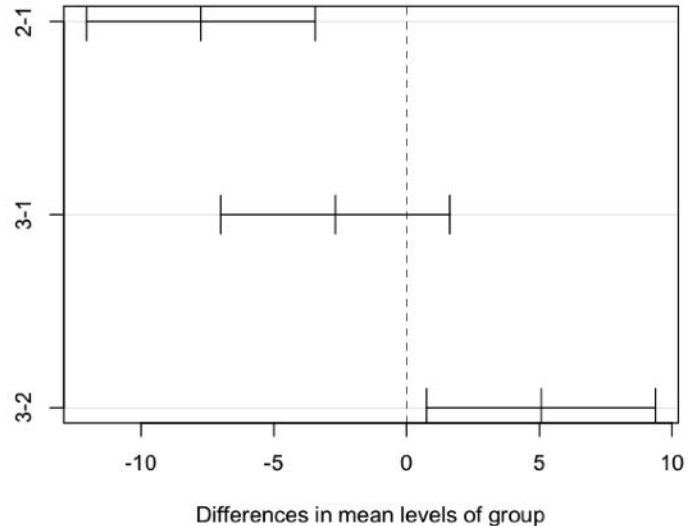
Für 3-1 geht die Null-Linie durch die Anzeige des Konfidenzniveaus; daher: Eine signifikante Differenz kann nicht angenommen werden.

➤ Größere Stichproben können das Intervall verengen und so zu Signifikanz führen.

➤ Zur Erinnerung auch noch einmal die Boxplots:



**95% family-wise confidence level**



Dies ist ein Beispiel für eine einfaktorielle Varianzanalyse; wir haben nur einen Faktor: den Texttyp. Zweifaktorielle Varianzanalyse z.B. wenn die Versuchspersonen zwei verschiedenen Gruppen angehören, z.B. deutschen Muttersprachlern vs. Deutsch-als-Fremdsprache-Lernern.

## *Schlußbemerkungen zur schließenden Statistik*

### Signifikanz und Stichprobengröße

Die Signifikanz einer Stichprobe hängt oft von der Größe der Stichprobe ab. Bei kleineren Stichproben kann sich bereits eine gewisse Tendenz zeigen, die aber noch nicht signifikant ist. Wenn man die Stichprobe vergrößert, können sich diese Tendenz verfestigen, und wir könnten ein signifikantes Ergebnis erhalten.

### Signifikanz und Aussagekraft (*predictive power*)

Mit dem Nachweis, dass sich in einer Stichprobe zwei Gruppen in einer bestimmten Messeigenschaft signifikant unterscheiden, hat man zwar gezeigt, dass sich die beiden Gruppen auch in der Grundgesamtheit mit einer gewissen Wahrscheinlichkeit unterscheiden. Aber wir sagen dabei nichts über die **Stärke** dieser Unterscheidung aus.

Beispiel: Die Körpergröße von erwachsenen Männern und Frauen unterscheidet sich, wie man durch Stichproben nachweisen kann. Allerdings gibt das Wissen, ob jemand weiblich oder männlich ist, nicht viele Anhaltspunkte darauf, wie groß diese Person ist. Das Geschlecht hat wenig Aussagekraft für die Körpergröße.

---

Bitte lesen Sie als Beispiel eines experimentellen Artikels:

Christine Röhrig, Stefan Baumann: Prosodic Marking of Information Structure in German (wird am Mittwoch auf die Moodle-Seite geladen, mit Annotationen).



## Experimenteller Artikel

Wir haben uns bereits einen sprachwissenschaftlichen Artikel näher angesehen, der auf sprachhistorischen Daten aufgebaut war.

Hier sehen wir uns einen Artikel an, der auf experimentellen Daten beruht und der auch eine statistische Auswertung mit einschließt:

Röhr, Christine & Stefan Baumann. 2010. Prosodic marking of information structure in German. *Proceedings 5th International Conference on Speech Prosody*. Chicago.

Es handelt sich also um einen Beitrag zu einer internationalen Konferenz, der in den sog. *Proceedings* dieser Konferenz veröffentlicht wurde.

## Abstract

In a production experiment on read German, we investigate the prosodic marking of discourse referents reflecting different types of information status. Acoustic and phonological analyses reveal an increase in the number of pitch accents as well as higher and later accentual peaks from textually given through textually accessible and inferentially accessible to new referents. Due to the increasing number of produced accents, segmental durations also increase from given to new information. Furthermore, specific accent types lead to different segmental durations. The differences in the prosodic marking of the two types of accessible information suggest a difference in cognitive activation between them, supporting the idea of an activation continuum of discourse referents.

**Index Terms:** prosody, information status, degree of givenness, cognitive activation, pitch accent, effort code

Der Abstract informiert über den Inhalt des folgenden Artikels; wir überschlagen ihn hier, da er sehr voraussetzungsreich ist.

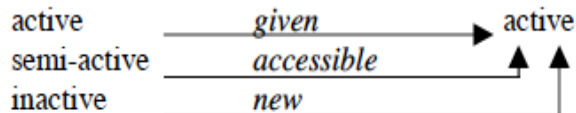
Index terms, auch Key words genannt: für den Artikel relevante Schlagwörter

## Introduction

Dieser Teil führt in die Problemstellung und die Forschungslage ein.

### 1. Introduction

The dimension of 'given' versus 'new' information is a central part in the investigation of information structure. Nevertheless, the various approaches to givenness in the literature differ with respect to the level this notion applies to (see [17] for an overview). In the present paper, we adopt Chafe's [4], [5] cognitive view of givenness defined as the degree of activation of a referent or proposition assumed by the speaker to be in the listener's consciousness at the time of utterance. Following Chafe, we postulate three different types of information status corresponding to three steps on a potentially continuous scale of cognitive activation: if a referent is already active in the listener's consciousness at the time of the utterance, it is *given*; if a referent becomes activated from a previously semi-active state, it is *accessible*; and if a referent becomes activated from a previously inactive state, it is *new* (see Fig.1).



Gegebene / Neue Information,  
wesentliche Begriffe der Informationsstruktur

Diskursreferent:

Ein Objekt, von dem die Rede ist.

Degree of Activation: Grad, zu dem ein Referent im Bewusstsein des Hörers präsent ist.

Drei Grade der Aktivierung:

- Given: Der Referent ist bereits präsent
- Accessible: Der Referent wurde bereits erwähnt, ist aber nicht hoch präsent, wird nun aber präsent gemacht.
- New: Der Referent wird neu erwähnt und damit präsent gemacht.

## Diskussion der Forschungslage

In terms of the prosodic marking of discourse referents, it is commonly assumed for West Germanic languages that new referents are marked by pitch accents whereas given referents get deaccented (see e.g. [8]). The prosody of accessible referents, however, is a matter of some debate. Chafe [5], e.g., postulates that accessible information is marked – like new information – by accented noun phrases, while Lambrecht [15] suggests that accessible referents are either accented or deaccented. However, several studies have shown by now that a simple dichotomy of accentuation versus deaccentuation is inappropriate for an account of accessible information, and for

Pierrehumbert & Hirschberg's study suggests a ternary distinction between high accents for new, low accents for accessible and no accents for given referents (applying to the 'hearer's mutual beliefs'), while Kohler's perception experiments indicate an interrelation between medial/late peaks and some kind of new information on the one hand and between early peaks and 'established' (interpreted here as accessible) information on the other.

Einfaches Bild:

- New: Pitch accent (= Tonhöhenakzent)
- Given: Deakzentuierung, Akzentlosigkeit

Doch diese einfache Zweiteilung genügt nicht.

Pierrehumbert & Hirschberg:

- H (Hochton): New
- L (Tiefton): Accessible
- – (kein Ton): Given

Kohler:

- Late (später Ton auf Silbe): New
- Early (früher Ton auf Silbe): Accessible

In a recent perception experiment on German, Baumann [1] (see also [2]) could show that accessible information cannot be treated as a uniform category and that different types of more or less activated information demand different accent types as linguistic markers. In fact, there is evidence that a range of accent types (including deaccentuation) can be mapped onto the gradient scale of activation degrees, with the pitch height on the accented syllable being the determining factor. Such a mapping suggests a somewhat iconic use of pitch height, which is compatible with Gussenhoven's [13] *Effort Code*: the higher the pitch on a lexically stressed syllable – and, in turn, the higher its prominence – the newer (or more newsworthy) the discourse referent.

In order to test this basic hypothesis in production data, we conduct a reading experiment displaying new, inferentially and textually accessible as well as textually given target referents. In addition to examining the types of pitch accent used, we investigate the alignment of the F0 peaks and valleys with the lexically stressed syllable of the target words, and the duration of these syllables. We hypothesize that speakers make use of

Baumann:

- Mehr Akzente
- Mehr Aktivierungsebenen; bilden eine Skala

Iconic: Ähnlichkeit zwischen Bedeutung und Ausdruck;  
Effort Code: Je mehr Akzent, desto neuer der Referent.

Vorausschau auf das Experiment:

- Lese-Experiment

Aufnahme von:

- Akzentton-Typen
- Früher/später Tongipfel
- Dauer der Silbe

## Method

In diesem Abschnitt geht es um eine genaue, nachvollziehbare Darstellung der Methode, die in dem Experiment angewendet wurde.

## Reading Material

Hier wird das experimentelle Material beschrieben:

The test material consists of ten different target words denoting discourse referents, each of them embedded in four target sentences in three different contexts (see Table 1).

Tabelle 1 gibt ein Beispiel für eines der zehn Target-Wörter, *Banane*.

Es gibt insbesondere vier Aktivierungsfälle:

- New Context 1
- Textually accessible, also vorerwähnt Context 1
- Scenario, aus Weltwissen inferierbar Context 2
- Textually given, unmittelbar vorher erwähnt Context 3

| target word = Ba.na.ne [ba 'na:nə] banana                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>CONTEXT 1: (a) new (b) textually accessible</p> <p>„Was hätten Sie gerne?“ (a) „Ich nehme die <u>Banane</u> mit“, antwortet Thomas dem Obsthändler. Normalerweise ernährt er sich sehr ungesund und isst zwischendurch ständig Süßigkeiten. Außerdem treibt er fast nie Sport und wenn doch, dann am liebsten Minigolf. (b) Er steckt sich die <u>Banane</u> ein. Lecker sieht die Banane aus. Vielleicht wird er demnächst öfter welche kaufen.</p> <p>“What would you like?“ (a) “I’ll take the <u>banana</u> (along)”, says Thomas to the fruit merchant. He usually eats very unhealthily and he is always eating sweets between meals. He hardly ever plays sport, and if he does he prefers mini golf. (b) He pockets the <u>banana</u>. The banana looks delicious. Maybe he’ll buy them more often in future.</p> |
| <p>CONTEXT 2: (c) scenario (inferentially accessible)</p> <p>Thomas darf heute im Zoo seinen Lieblingsaffen füttern. Voller Vorfreude wird er sich gleich auf den Weg zu ihm machen. (c) Er steckt sich die <u>Banane</u> ein. Vorhin war er dafür extra noch auf dem Markt beim Obsthändler.</p> <p>Today Thomas is allowed to feed his favourite monkey in the zoo. With great anticipation he’s about to set off (for the zoo). (c) He pockets the <u>banana</u>. He’s just been to the green grocer’s at the market especially to get one.</p>                                                                                                                                                                                                                                                                           |
| <p>CONTEXT 3: (d) textually given</p> <p>Thomas hat gerade auf dem Markt eine Banane gekauft. (d) Er steckt sich die <u>Banane</u> ein. In Zukunft möchte er sich viel gesünder ernähren.</p> <p>Thomas has just bought a banana at the market. (d) He pockets the <u>banana</u>. In the future he wants to eat much more healthily.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

### **2.2. Speakers and recordings**

Nine native speakers of German (six female, three male), aged between 22 and 31, took part in the experiment. All of them originated from the area around Cologne and Düsseldorf. Before the acoustic recordings, each subject was asked to read through the material thoroughly in order to guarantee full comprehension. After that, the subjects' task was to read out the texts in a contextually appropriate manner. The contexts were presented on separate file cards in pseudo-randomised order. They were repeated three times by each subject, adding up to 120 target sentences per speaker for analysis, i.e. 1080 tokens in total.

Genaue Beschreibung der Durchführung:

- 9 Sprecher; Angabe zu Geschlecht, Alter, Herkunft, Rekrutierung.
- Verfahren der Aufnahme
- Pseudo-Randomisierung: Zufällige Präsentation unter Vermeidung von Reihenfolge-Effekten
- 9 Sprecher,  
10 Target-Wörter,  
4 Target-Wort-Tokens  
(präsentiert in 3 Kontexten)  
je 3 mal gelesen:  
 $9 \times 10 \times 4 \times 3 = 1080$  Tokens insgesamt

## 2.3. Analysis

The acoustic data were annotated manually using the EMU software [3]. At four levels, we marked the beginning and the end of the target word, the foot starting with the lexically stressed syllable, the stressed syllable itself and the vowel included in it.

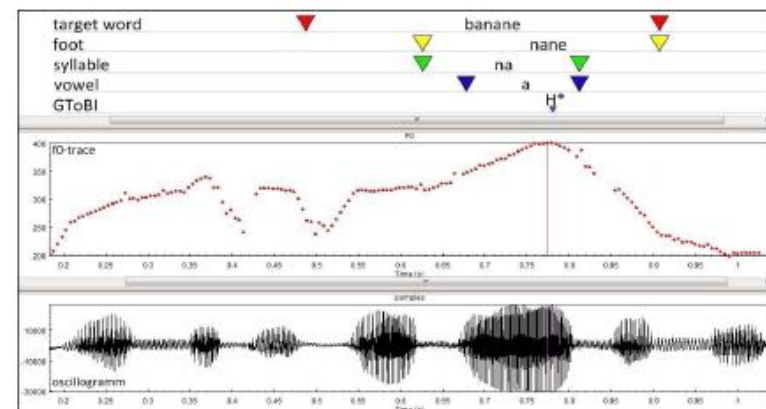


Figure 2: Target sentence *Ich nehme die BaNane mit* ('I'll take the banana (along)'; new, speaker S08) with labels, F0 trace and oscillogram in EMU.

At the tonal level, we annotated the F0 minima and maxima making up pitch accents and categorised them according to GToBI [9], [10]. The structure of the target sentences, with the argument in non-final position, allows the nuclear accent either to fall on the target word or on the sentence-final verbal particle. In the latter case the target word is either deaccented (marked by '0') or receives a prenuclear accent (e.g. in *Er steckt sich die BaNane EIN.*), indicated by an additional 'PN'

## Analyse

Hier wird das Messverfahren beschrieben, am Beispiel von einem von 1080 Tokens:

- Das Target-Wort, der Silbenfuß, die Silbe und der Silbenvokal werden markiert; deren Länge kann somit gemessen werden.
- $F_0$  (= Tonbewegung) Minimum und Maximum in der betonten Silbe werden gemessen;
- Wenn die betonte Silbe des Targetworts den nuklearen Akzent (Hauptakzent des Satzes) trägt, wird sie einem der GToBI-Akzenttypen zugeordnet ( $H^*$ ,  $L+H^*$ ,  $!H^*$ ,  $H+L^*$ ,  $H+!H^*$ ,  $L^*$ , wobei  $H^*$ : Hoch,  $!H^*$ : Downstep,  $L^*$ : Tief)
- Wenn die betonte Silbe des Targetworts nicht den Hauptakzent trägt, sondern einen Nebenakzent: PN
- Wenn die betonbare Silbe nicht akzentuiert (deakzentuiert) wird: 0



## Results and Discussions

Hier werden die Resultate dargestellt und (oft in einem separaten Abschnitt) diskutiert.

### Akzenttypen

Hier wurden zunächst fünf Akzenttypen und die vier Aktivierungszustände als nominale Klassen behandelt und ein Chi-Quadrat-Test durchgeführt;

Ergebnis: Die Wahrscheinlichkeit, dass Akzenttypen und Aktivierungszustände in der Grundgesamtheit zufällig einander zugeordnet sind, ist sehr klein; als Ergebnis ist höchst significant ( $p < 0,001$ ).

Das Schaubild gibt einen klaren Eindruck:

Je neuer die Diskursreferenten, desto ausgeprägter der Akzent, vor allem der H\*-Akzent.

## 3. Results and Discussion

### 3.1. Accent types

As an overall result, the distribution of accent categories (including prenuclear accents and deaccentuation) proves to depend significantly on the referents' information status (chi square:  $p < 0.001$ ), as shown in Fig.3.

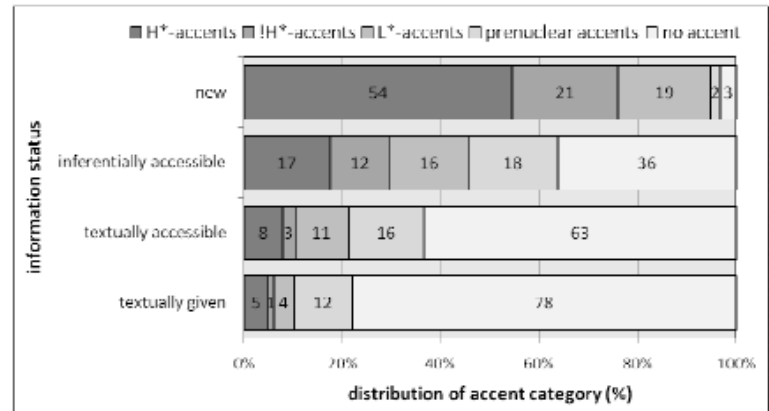


Figure 3: Relative distribution of nuclear accent categories (comprising accent types with the same starred tone), prenuclear accents and deaccentuation on all target words per information status; all speakers pooled.

Wenn man alle acht Akzenttypen mit den vier Aktivierungstypen korreliert, erhält man ein differenzierteres Bild:

- Wir erhalten noch immer ein höchst signifikantes Resultat.
- L+H\* und H\* wird vor allem für neue Information verwendet.
- Aber H\* ist auch der häufigste Akzenttyp für andere Aktivierungstypen, falls überhaupt ein Nuklearakzent verwendet wird.

Position des F<sub>0</sub>-Gipfels in der Silbe:

- Early Peak: H+L\*, H+!H\*, treten zusammen bei „inferentially accessible“ häufiger als H\* auf.

Rolle des Tiefakzents:

- L\* und H+L\* treten zusammen bei „textually accessible“ häufiger als H\* auf.

Table 2. *Relative distribution of accent types, including prenuclear accents (PN) and deaccentuation (0), on all target words per information status; all speakers pooled.*

|                                                  |                        | information status (all speakers & all target referents): tonal marking |          |      |        |        |        |         |         |       |      |
|--------------------------------------------------|------------------------|-------------------------------------------------------------------------|----------|------|--------|--------|--------|---------|---------|-------|------|
| preference for deaccentuation of target referent | low                    | <b>new</b>                                                              | accents  | H* > | L+H* > | !H* >  | H+L* > | H+!H* > | L* >    | 0 >   | PN   |
|                                                  |                        |                                                                         | % of 270 | 40,4 | 14,1   | 12,6   | 11,1   | 8,9     | 7,8     | 3,3   | 1,9  |
|                                                  |                        | <b>inferentially accessible</b>                                         | accents  | 0 >  | PN >   | H* >   | H+L* > | H+!H* > | !H* >   | L* >  | L+H* |
|                                                  |                        |                                                                         | % of 270 | 36,3 | 18,1   | 16,7   | 11,5   | 7,0     | 5,2     | 4,4   | 0,7  |
|                                                  |                        | <b>textually accessible</b>                                             | accents  | 0 >  | PN >   | H* >   | L* >   | H+L* >  | H+!H* > | !H* > | L+H* |
|                                                  |                        |                                                                         | % of 270 | 63,7 | 15,2   | 7,4    | 5,9    | 4,8     | 1,5     | 1,1   | 0,4  |
| high                                             | <b>textually given</b> | accents                                                                 | 0 >      | PN > | H* >   | H+L* > | L* >   | H+!H* > | L+H* =  | !H*   |      |
|                                                  |                        | % of 269                                                                | 78,1     | 11,9 | 4,5    | 2,2    | 1,9    | 0,7     | 0,4     | 0,4   |      |

As in the analysis of the comprised pitch accent categories, the distribution of single pitch accent types shows significant differences between the four types of information status (chi square:  $p < 0.001$ ). The most prominent pitch accent types L+H\* and H\* are preferably used for marking new information. However, Table 2 shows that H\* is the most frequent single type of accent on accessible and given referents as well, if they receive a nuclear pitch accent (the most frequent markers are deaccentuation and prenuclear accents, very often realised as L\*). Thus, H\* qualifies as the default nuclear accent type. Within the information status 'inferentially accessible', however, the two types of early peak accents (H+L\* and H+!H\*) form a larger group than H\* accents (18.5% vs. 16.7%). In comparison, for textually accessible referents the two L\* accent types (L\* and H+L\*) taken together are more frequent than H\* accents (10.7% vs. 7.4%; see also Fig.3).

## Stützung der Hypothese:

### F<sub>0</sub>-Gipfelposition

- Mittlere und späte Gipfel: New
  - Frühe Gipfel: Accessible/Given
- Varianzanalyse zu Position von H\*,  
möglich da wir jetzt Ordinaldaten haben:
- Mehr oder weniger aktiviert,
  - Position von H\* mehr oder weniger vorn.

Wichtige Daten zu Sprecherunterschieden,  
ohne statistische Angabe.

This distribution of accent types supports Kohler's [14] proposal that medial (H\*) and late peaks (L+H\*) rather mark new information while early peaks (H+!H\*, H+L\*) mark established information, which is comparable to our notion of accessibility (and givenness). In terms of alignment differences *within the same accent type*, we only find significantly later H\* accents marking new versus textually accessible referents (ANOVA:  $F(3, 186) = 8.197, p < 0.001$ ).

In sum, the variation in frequency, position and type of accentuation with changes in the referents' information status clearly confirms our hypothesis: the 'newer' the referent, the more likely its marking by a high (and late-peak) nuclear pitch accent. As the degree of activation of the target referent increases, subjects are more likely to use lower (and early-peak) and less prominent accents (if at all). The semantic and structural differences between the test words had no significant effect on their prosodic marking.

However, the results are to some extent speaker-specific. While new information generally gets accented, some subjects hardly differentiate between the prosodic marking of accessible and given information. There seem to be (at least) two different types of speaker: those speakers who usually deaccent accessible and given information generally prefer low pitch accents (for marking new information), while those speakers who often place an accent on accessible and given referents use high pitch accents by default.

## Silbenlänge

Die Silbenlänge ist ebenfalls höchst signifikant mit dem Aktivierungstatus korreliert.

Interessant:

- Silben mit L\* und prä nuklearen Akzente auf Targetwörtern sind signifikant länger.

Erklärung:

- Kompensation für die Betonung bei fehlendem H\*-Akzent

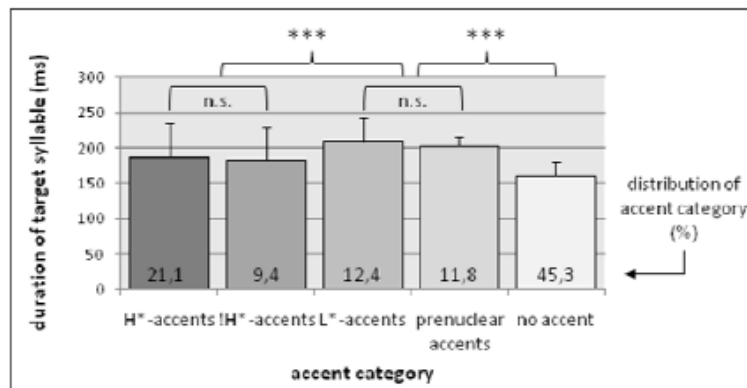


Figure 4: Durations of target referent's stressed syllable per accent category; all speakers pooled.

This means that different accent categories are characterised by different durations (ANOVA:  $F(4, 1080) = 79.766$ ,  $p < 0.001$ ). Surprisingly, however, post hoc tests (Tukey-HSD) show that syllables bearing L\* and prenuclear accents (which are also often low) are significantly longer than syllables carrying H\* and !H\* accents, although the latter two were found to be the most frequent markers of new referents, which are generally longer than accessible and given referents (shortest syllable durations in deaccented target words). In fact, this result confirms the claim that for low accents (nuclear and prenuclear) duration is a decisive prominence-lending feature compensating for a lack of tonal movement (see e.g. [9: 278]).

## 4. Conclusions

Generally, our hypotheses are confirmed. The newer, or less activated, a referent, the more likely it is to be marked by a nuclear pitch accent in read German. Conversely, the higher the degree of a referent's activation, the higher is the preference for deaccentuation. Prenuclear accents are only used if the referent is already accessible or given in the discourse. As for the *types* of accent used, new information is found to be primarily marked by high and relatively late peaks while in accessible and given information the relative proportion of lower and early peak accents increases. Since more pitch accents are produced on less activated referents, a decreasing degree of givenness is reflected by longer segmental durations. In fact, the duration of the target word's stressed syllable partly depends on the type of accent used, with – somewhat surprisingly – syllables carrying low accents being longer than syllables carrying high accents. Thus, segmental durations only indirectly reflect a constituent's degree of activation.

Our results show that the actual prosodic marking of a referent ultimately depends on speaker-specific preferences. Nevertheless, the idea of activation degrees, and in particular their marking by corresponding degrees of prominence, indicated by pitch height and peak alignment, is highly compatible with Gussenhoven's [13] *Effort Code*: higher and later peaks indicate less active referents.

## Zusammenfassung

Ein Rückblick auf die erzielten Ergebnisse,

und eine allgemeine, abschließende Deutung in einem weitverbreiteten Modell („effort code“).

## *Nochmal Abstract*

Es lohnt sich nun, den Abstract zu lesen:

### **Abstract**

**In** a production experiment on read German, we investigate the prosodic marking of discourse referents reflecting different types of information status. Acoustic and phonological analyses reveal an increase in the number of pitch accents as well as higher and later accentual peaks from textually given through textually accessible and inferentially accessible to new referents. Due to the increasing number of produced accents, segmental durations also increase from given to new information. Furthermore, specific accent types lead to different segmental durations. The differences in the prosodic marking of the two types of accessible information suggest a difference in cognitive activation between them, supporting the idea of an activation continuum of discourse referents.

**Index Terms:** prosody, information status, degree of givenness, cognitive activation, pitch accent, effort code

# Theoretische Werkzeuge

## *Theoretische Werkzeuge der diskreten Mathematik*

Ziel: Darstellung von wichtigen theoretischen Werkzeugen für die Beschreibung sprachlicher Phänomene.

Diese Werkzeuge dienen dazu, formale **Modelle** zu entwickeln, die bestimmte Beobachtungen erfassen und die selbst wiederum Voraussagen machen, die man überprüfen kann.

Hier: Werkzeuge der **diskreten** Mathematik;

für die Phonetik sind auch Werkzeuge der **numerischen** Mathematik.

# Mengen

## Was ist eine Menge?

Die Mengenlehre wurde gegen Ende des 19. Jahrhunderts von dem Mathematiker Georg Cantor als theoretische Basis der Mathematik entwickelt.

Ziel: eine elementare, einfache und konsistente Theorie, auf deren Grundlage sich die gesamte Mathematik aufbauen ließe.

Die klassische Definition einer **Menge** (englisch: **set**):

Eine Menge ist eine abstrakte Zusammenfassung bestimmter wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens zu einem Ganzen.



### Betrachten wir die Bestandteile dieser Definition genauer:

- **abstrakt:** Die Objekte brauchen nicht physisch zusammengefasst zu werden, (wie etwa die Marken einer Briefmarkensammlung in einem Album).
- **Zusammenfassung:**  
Es muss klar sein, welche Objekte dazugehören und welche nicht.  
Es handelt sich lediglich um eine Zusammenfassung, nicht um eine Anordnung – die Reihenfolge, in der wir die Elemente angeben, spielt daher keine Rolle.  
(Strukturen, in denen die Reihenfolge relevant ist, heißen **Tupel** oder **Listen**.)
- **wohlunterschieden:**  
Die Objekte müssen identifizierbar sein, das heißt, man muss sie auseinanderhalten können.  
Insbesondere kann in einer Menge ein und dasselbe Objekt nicht mehrfach auftauchen.  
(Strukturen, die auch das mehrfache Vorkommen von Objekten erlauben, heißen **Multisets**.)
- **Anschauung / Denken:** Die Objekte können konkret sein (z.B. die Studenten im Seminarraum) oder abstrakt (wie die sieben Kardinaltugenden oder die natürlichen Zahlen zwischen 3 und 17).  
Es können sogar Mengen (abstrakte Objekte) zu neuen Mengen zusammengefasst werden.

## Einige Beispiele und weitere Begriffe:

- Die Objekte, die zu einer Menge gehören, nennt man **Elemente** der Menge.  
Von den Elementen wird nichts weiter vorausgesetzt.  
Insbesondere kann es sich bei ihnen selbst um Mengen handeln.  
Wir schreiben:  $x \in A$  für “ $x$  ist ein Element der Menge  $A$ ”.
- Mengen können klein sein (wie die Menge der natürlichen Zahlen zwischen 3 und 17) oder groß (wie die Menge der natürlichen Zahlen zwischen 3 und 17 Milliarden).  
Diese Mengen sind **endlich**,  
aber es gibt auch **unendliche** Mengen (z.B. die Menge aller natürlichen Zahlen 1, 2, 3, 4, ...).
- Mengen können nur ein einziges Objekt enthalten (sogenannte **Einermengen**);  
man beachte, dass die Mengen stets abstrakt sind, Elemente oft konkret.
- Mengen können auch überhaupt keine Objekte enthalten (die **leere** Menge,  $\emptyset$ ).

## Identität von Mengen

Wann sind zwei Mengen gleich?

Genau dann, wenn sie dieselben Elemente enthalten:

- Definition:  $A = B$  gdw. für alle  $x$  gilt:  $x \in A$  gdw.  $x \in B$

(Im Englischen wird steht für „genau dann wenn“ *if and only if* – oft abgekürzt zu *iff*).

## Darstellung von Mengen: Aufzählung

Wir können die Elemente einer Menge einfach **aufzählen**. Man verwendet dazu geschweifte Klammern und Kommas, um die Elemente voneinander zu trennen.

- {a, e, i, o, u} die fünf Grundvokale
- {a, e, i, {o, u}} enthält nur 4 Elemente, eines ist selbst eine Menge
- {a} eine Einermenge
- {} die leere Menge, auch  $\emptyset$  geschrieben
- Die Reihenfolge der Aufzählung ist irrelevant: {a, b, c, d, e} = {b, c, d, a, e}.
- Mehrfachvorkommen ist irrelevant: {a, b, c, d, e} = {a, b, c, c, d, e}.

## Darstellung von Mengen: Abstraktion (Prädikatsnotation)

Die Elemente, die zu der Menge gehören, werden beschrieben.

Alle Objekte, auf welche die Beschreibung zutrifft, und nur diese, sind dann in der Menge enthalten.

Schreibweise:

- {Variable | Beschreibung der Variablen}

Typischerweise verwenden wir x, y, z als Buchstaben für Variablen.

- {x | x ist ein Grundvokal}
- {x | x ist eine natürliche Zahl und  $1 \leq x \leq 1000$ }
- {x | x ist eine natürliche Zahl}.

Das erste Beispiel liest man "die Menge aller x so, dass x ein Grundvokal ist".

## Die Teilmengenbeziehung

Eine wichtige Relation zwischen Mengen ist die Beziehung der **Teilmenge**.

Um auszudrücken, dass A eine Teilmenge von B ist, schreiben wir  $A \subseteq B$ ; dies ist wie folgt definiert:

➤ Definition:  $A \subseteq B$  gdw. gilt: Für alle x, wenn  $x \in A$ , dann  $x \in B$ .

Zwei Beispiele für Teilmengenbeziehungen:

➤  $\{a, e, i\} \subseteq \{a, e, i, o, u\}$   
 $\emptyset \subseteq \{a, e, i\}$

Die leere Menge ist Teilmenge jeder anderen Menge A:  $\emptyset \subseteq A$ .

Die Teilmengenbeziehung gehorcht drei wichtigen Gesetzen:

➤ Reflexivität: Für jede Menge A gilt:  $A \subseteq A$   
Transitivität: Wenn  $A \subseteq B$  und  $B \subseteq C$ , dann gilt auch:  $A \subseteq C$   
Antisymmetrie: Wenn  $A \subseteq B$  und  $B \subseteq A$ , dann gilt:  $A = B$

Das Gesetz der Reflexivität besagt: jede Menge ist eine Teilmenge von sich selbst.

Wenn man das ausschließen will, kann man zu der Relation der **echten Teilmenge** greifen, für die wir die Schreibweise  $A \subset B$  verwenden:

➤ Definition:  $A \subset B$  gdw.  $A \subseteq B$ , aber nicht:  $B \subseteq A$ .

## Kardinalität von Mengen

Wir wollen manchmal wissen, wie viele Elemente eine Menge hat.  $\kappa$

Dies wird die **Kardinalität** einer Menge genannt.

Die Kardinalität der Menge A wird durch  $\text{card}(A)$  oder  $\#(A)$  angegeben.

➤  $\text{card}(A) = \#(A) =$  die Zahl der Elemente in der Menge A

Beispiele:

➤  $\#\{a, b, c\} = 3$

$\#\{a, b, b, c\} = 3$

$\#\{\{a, b\}, \{c\}\} = 2$

$\#\{a\} = 1$

$\#(\emptyset) = 0$

Dies sind Beispiele von **endlichen** Mengen.

Daneben gibt es auch **unendliche** Mengen wie z.B. die Menge aller natürlichen Zahlen  $\mathbb{N}$ , oder die Menge aller reellen Zahlen  $\mathbb{R}$ .

Cantor konnte zwischen verschiedenen Arten unendlicher Mengen unterscheiden:

➤ Die natürlichen Zahlen sind **abzählbar unendlich**

(man kann sie sich in einer nicht abbrechenden Reihe vorstellen).

Dies ist die Kardinalität  $\aleph_0$  (Aleph Null)

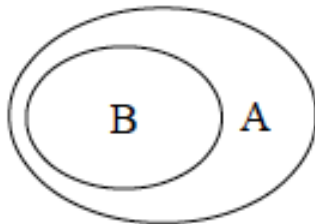
➤ Die reellen Zahlen sind **überabzählbar unendlich**.

## Darstellung von Mengen durch Venn-Diagramme

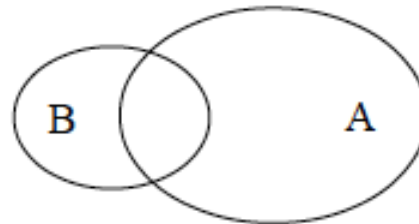
Methode zur Darstellung mengentheoretischer Beziehungen, benannt nach dem Mathematiker John Venn, nach dem Mathematiker Leonard Euler auch Euler-Kreise genannt (didaktisches Hilfsmittel verwendete -- *Lettres à une Princesse d'Allemagne*, 1768).

In einem Venn-Diagramm werden die Elemente durch Punkte in der Ebene dargestellt, und Mengen von Elementen durch geschlossene Flächen.

Die Teilmengenbeziehung kann damit wie folgt dargestellt werden:



$$B \subseteq A, B \subset A$$



$$B \not\subseteq A, B \not\subset A$$

## Mengentheoretische Operationen

Die **Vereinigung** (englisch: **union**) zweier Mengen A und B, geschrieben  $A \cup B$ , ist diejenige Menge, welche alle Elemente, die in A oder B vorkommen, und nur diese enthält.

- Definition:  $A \cup B = \{x \mid x \in A \text{ oder } x \in B\}$
- $\{a, e, i\} \cup \{i, o, u\} = \{a, e, i, o, u\}$   
 $\{a, e, i\} \cup \{o, u\} = \{a, e, i, o, u\}$   
 $\{a, e, i\} \cup \emptyset = \{a, e, i\}$

Der **Durchschnitt** (englisch: **intersection**) zweier Mengen A und B, geschrieben  $A \cap B$ , ist diejenige Menge, welche alle Elemente enthält, die sowohl in A wie auch in B vorkommen:

- Definition:  $A \cap B = \{x \mid x \in A \text{ und } x \in B\}$
- $\{a, e, i\} \cap \{i, o, u\} = \{i\}$   
 $\{a, e, i\} \cap \{o, u\} = \emptyset$   
 $\{a, e, i\} \cap \emptyset = \emptyset$

Die mengentheoretische **Differenz** (englisch: **subtraction**)  $A \setminus B$  ist diejenige Menge, welche genau die Elemente aus A enthält, die nicht in B enthalten sind:

- Definition:  $A \setminus B = \{x \mid x \in A \text{ und } x \notin B\}$
- $\{a, e, i\} \setminus \{i, o, u\} = \{a, e\}$   
 $\{a, e, i\} \setminus \{o, u\} = \{a, e, i\}$   
 $\{a, e, i\} \setminus \emptyset = \{a, e, i\}$

Häufig beschränken wir uns auf eine bestimmte Menge von Objekten, z.B. die Menge der natürlichen Zahlen, und betrachten Teilmengen dieser Menge.

Eine solche Menge nennt man **Universum**, oft mit  $U$  bezeichnet.

Bezüglich eines Universums  $U$  definieren wir das **Komplement** einer Menge  $A$ :

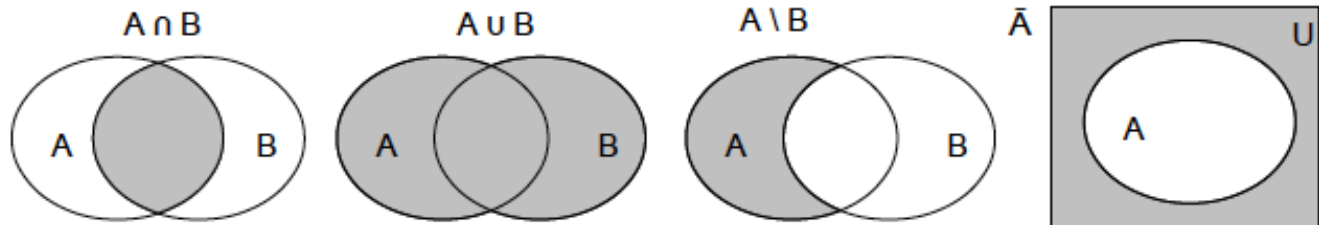
➤ Definition Komplement:  $\bar{A} =_{\text{def}} U \setminus A$ . (auch durch Überstreichung markiert).

Zum Beispiel gilt bezüglich der Grundvokale als Universum

➤  $\overline{\{a, e, i\}} = \{o, u\}$

### Darstellung von Mengenoperationen durch Venn-Diagramme

Wir können die eben eingeführten Operationen wie folgt durch Venn-Diagramme darstellen:





## Mengentheoretische Gesetze

Die mengentheoretischen Begriffe wie Vereinigung, Durchschnitt, Teilmenge, Komplement sind durch strukturelle Beziehungen miteinander verbunden.

Im Folgenden seien A, B, C beliebige Mengen.

Die jeweiligen Gesetze lassen sich mit Hilfe von Venn-Diagrammen darstellen.

- Idempotenz:  $[A \cap A] = A$   
 $[A \cup A] = A$
- Kommutativität:  $[A \cap B] = [B \cap A]$   
 $[A \cup B] = [B \cup A]$
- Assoziativität:  $[A \cap [B \cap C]] = [[A \cap B] \cap C]$   
 $[A \cup [B \cup C]] = [[A \cup B] \cup C]$
- Distributivität:  $[A \cap [B \cup C]] = [[A \cap B] \cup [A \cap C]]$   
 $[A \cup [B \cap C]] = [[A \cup B] \cap [A \cup C]]$
- De Morgan:  $[\overline{A \cap B}] = [\overline{A} \cup \overline{B}]$   
 $[\overline{A \cup B}] = [\overline{A} \cap \overline{B}]$

Mathematische Strukturen mit diesen Gesetzen nennt man eine **Boolesche Algebra**, nach dem irischen Mathematiker George Boole.

Es gibt auch Beziehungen zwischen den mengentheoretischen Operationen und der Teilmengenbeziehung:

➤  $A \subseteq B$  gdw.  $A \cup B = B$

➤  $A \subseteq B$  gdw.  $A \cap B = A$

Die leere Menge  $\emptyset$  und das Universum  $U$  gehorchen folgenden Gesetzmäßigkeiten:

➤  $A \cup \bar{A} = U$

➤  $A \cap \bar{A} = \emptyset$

➤  $\bar{\bar{A}} = A$

➤  $\bar{U} = \emptyset$

➤  $\bar{\emptyset} = U$

# Relationen

## Was ist eine Relation?

Ein weiterer grundlegender Begriff der Mathematik ist die **Relationen**.

Relationen dienen zur Modellierung von Beziehungen zwischen Elementen; zum Beispiel ist die Beziehung "größer als" eine Relation.

Beziehungen werden ausgedrückt durch **geordnete Paare**.

Wir schreiben geordnete Paare in spitzen Klammern,

z.B. ist  $\langle a, b \rangle$  das geordnete Paar, das aus  $a$  als dem ersten und  $b$  als dem zweiten Element besteht.

Im Gegensatz zu Mengen ist bei geordneten Paaren die Reihenfolge wichtig;

wir haben also:  $\langle a, b \rangle \neq \langle b, a \rangle$ .

Ferner gilt auch, dass ein geordnetes Paar wie  $\langle a, a \rangle$  nicht reduziert werden kann auf  $a$ .

Eine **Relation** ist eine Menge von geordneten Paaren.

Zum Beispiel kann man die Relation "geht im Alphabet unmittelbar voraus" als die folgende Menge von geordneten Paaren von Buchstaben angeben:

➤  $\{\langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle, \langle e, f \rangle, \dots \langle x, y \rangle, \langle y, z \rangle\}$

Wie stets bei Mengen, kommt es hierbei auf die Reihenfolge der geordneten Paare nicht an, d.h. wir könnten auch schreiben:

➤  $\{\langle b, c \rangle, \langle a, b \rangle, \langle e, f \rangle, \langle c, d \rangle, \dots \langle x, y \rangle, \langle y, z \rangle\}$

## Kartesisches Produkt und Feld

Wenn wir zwei Mengen A und B haben, können wir die Menge aller Paare bilden, wobei das erste Element aus A und das zweite aus B stammt.

Wir nennen diese Menge das **kartesische Produkt** und schreiben  $A \times B$ .

➤  $A \times B = \{\langle x, y \rangle \mid x \in A \text{ und } y \in B\}$

Beispiel:  $\{a, b, c\} \times \{1, 2, 3, 4\} =$

$$\begin{aligned} &\{\langle a, 1 \rangle, \langle b, 1 \rangle, \langle c, 1 \rangle, \\ &\quad \langle a, 2 \rangle, \langle b, 2 \rangle, \langle c, 2 \rangle \\ &\quad \langle a, 3 \rangle, \langle b, 3 \rangle, \langle c, 3 \rangle \\ &\quad \langle a, 4 \rangle, \langle b, 4 \rangle, \langle c, 4 \rangle\} \end{aligned}$$

Unter dem **Feld** einer Relation versteht man die Menge aller Objekte, die zu irgendeinem Objekt in der Relation stehen.

➤ das Feld von  $R = \{x \mid \text{es gibt ein } y \text{ sodass gilt: } \langle x, y \rangle \in R \text{ oder } \langle y, x \rangle \in R\}$

Beispiel: Das Feld der Relation  $\{\langle a, 1 \rangle, \langle b, 2 \rangle, \langle c, 2 \rangle\}$  ist die Menge  $\{a, b, c, 1, 2\}$ .

## Eigenschaften von Relationen

Viele Relationen weisen bestimmte Eigenschaften auf; die wichtigsten davon:

- R ist **reflexiv** wenn gilt: Wenn  $\langle x, y \rangle \in R$ , dann gilt:  $\langle x, x \rangle \in R$  und  $\langle y, y \rangle \in R$
- R ist **transitiv** wenn gilt: Wenn  $\langle x, y \rangle \in R$  und  $\langle y, z \rangle \in R$ , dann gilt:  $\langle x, z \rangle \in R$
- R ist **symmetrisch** wenn gilt: Wenn  $\langle x, y \rangle \in R$  gdw.  $\langle y, x \rangle \in R$ ,
- R ist **antisymmetrisch** wenn gilt: Wenn  $\langle x, y \rangle \in R$  und  $\langle y, x \rangle \in R$ , dann gilt:  $x=y$

Die Relation “ist ähnlich zu” ist reflexiv: Jedes x ist sich selbst ähnlich.

Sie ist auch symmetrisch: Wenn x y ähnlich ist, dann ist auch y x ähnlich.

Die Relation “ist größer als” ist transitiv:

Wenn x größer als y und y größer als z ist, dann ist x größer als z.

Die Relation “ist mindestens so groß wie” (für Zahlen) ist antisymmetrisch:

Wenn eine Zahl n mindestens so groß ist wie m, und m mindestens so groß ist wie n, dann müssen n und m genau gleich groß sein, und mithin, da es sich um Zahlen handelt, identisch sein.

Mit den eben definierten Eigenschaften können wir wichtige Typen von Relationen definieren.

## Äquivalenzrelationen

Diese drücken aus, dass Objekte in einer bestimmten Hinsicht ähnlich, vergleichbar sind.

Beispiele: “ist ähnlich zu”, “verdient genau so viel wie”, “hat genau so viele Haare auf dem Kopf wie”.

Hier ist die Definition:

- R ist eine **Äquivalenzrelation** wenn gilt:
  - a. R ist reflexiv
  - b. R ist symmetrisch
  - c. R ist transitiv

Das gilt für die angeführten Beispiele. Nehmen wir “ist in derselben Stadt geboren wie”:

- (a) Für jede Person x gilt: x ist in derselben Stadt geboren wie x.
- (b) Wenn x in derselben Stadt geboren ist wie y, dann ist y in derselben Stadt geboren wie x.
- (c) Wenn x in derselben Stadt geboren ist wie y, und y in derselben Stadt geboren wie z, dann ist x in derselben Stadt geboren wie z.

Äquivalenzrelationen zerlegen eine Menge in nicht-überlappende Teilmengen, die hinsichtlich der Relation äquivalent sind.

Wir sprechen hier von einer **Partition**.

Beispiel: Wenn Hans, Maria und Peter in Berlin geboren sind, Otto und Sophia in Lübeck und Karl in Buxtehude, dann erhalten wir folgende Partition:

- $\{\{Hans, Maria, Peter\}, \{Otto, Sophia\}, \{Karl\}\}$

## Präordnung

Beispiel: “ist größer als” (für Personen):

- R ist eine (strenge) **Präordnung** gdw:
  - a. R ist transitiv
  - b. R ist asymmetrisch, d.h.: Wenn  $\langle x, y \rangle \in R$ , dann gilt:  $\langle y, x \rangle \notin R$

Wenn z.B. Maria größer als Hans ist und Hans größer als Eva, dann ist Maria größer als Eva.  
Und wenn Maria größer als Hans ist, dann ist Hans nicht größer als Maria.

Es gibt auch den Begriff der **nicht-strengen Präordnung**; dieser schließt alle Paare  $\langle x, x \rangle$  mit ein, nicht-strikte Präordnungen sind also reflexiv.

Ein Beispiel: “ist mindestens so groß wie”.

## Halbordnung

- R ist eine **Halbordnung** (engl. partial order) gdw:
  - a. R ist reflexiv
  - b. R ist transitiv
  - b. R ist antisymmetrisch

Beispiel: Die Teilmengenbeziehung.

- a. Sie ist reflexiv: Für alle Mengen A gilt:  $A \subseteq A$ .
- b. Sie ist transitiv: Wenn  $A \subseteq B$  und  $B \subseteq C$ , dann  $A \subseteq C$ .
- c. Und sie ist antisymmetrisch: Wenn  $A \subseteq B$  und  $B \subseteq A$ , dann gilt:  $A = B$ .

Bei **strengen** Halbordnungen werden reflexive Paare der Art  $\langle x, x \rangle$  entfernt.

Ein Beispiel ist die Relation der echten Teilmenge,  $\subset$ .

## Totale Ordnung

Abschließend führen wir noch den Begriff der **totalen Ordnung** (auch **lineare** Ordnung oder einfach **Ordnung**) ein.

Dies ist eine Halbordnung, die auch die Eigenschaft der Totalität erfüllt:

- R ist eine **totale Ordnung** gdw.
  - a. R ist eine Halbordnung
  - b. R ist total, d.h. für alle  $x, y$  im Feld von R gilt:  $\langle x, y \rangle \in R$  oder  $\langle y, x \rangle \in R$ .

Dies trifft zu z.B. für die Relation  $\leq$  für natürliche Zahlen:

- sie erfüllt die Halbordnung (sie ist reflexiv, transitiv und antisymmetrisch),
- und sie ist total: Für je zwei natürliche Zahlen  $n, m$  gilt entweder  $n \leq m$  oder  $m \leq n$  (oder beides, wenn  $n=m$ ).

Man kann eine totale Ordnung in Form einer Kette darstellen.

Dies trifft nicht zu auf die Teilmengenbeziehung:

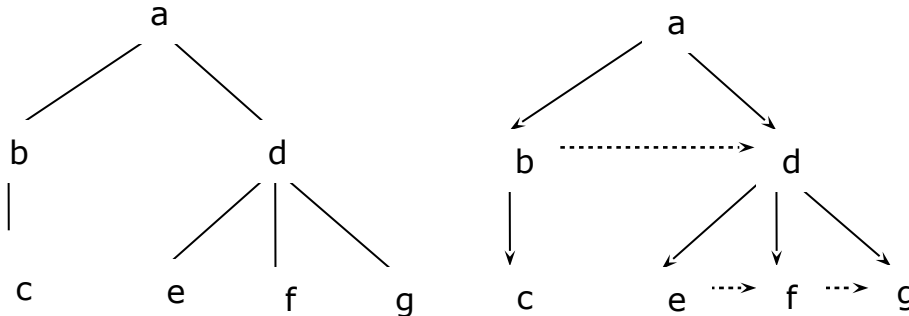
- Es gilt nicht für zwei beliebige Mengen  $A, B$ :  $A \subseteq B$  oder  $B \subseteq A$



## Anwendung: Syntaktische Bäume

Ein wichtiges Instrument zur Beschreibung von syntaktischen Strukturen sind **Bäume**. Dies ist ein Untertyp von **Graphen**.

Beispiel eines Baumes in normaler Darstellung und mit expliziten Dominanz- und Präzedenzrelationen.



Graphen bestehen aus **Knoten** und **Kanten**.

Die Kanten, die jeweils zwei Knoten verbinden, können als (zweistellige) **Relation** dargestellt werden, die **Dominanzrelation** (oben – unten).

Die Knoten stehen in einer weiteren Anordnung, von links nach rechts (diese Anordnung ist also nicht zufällig).

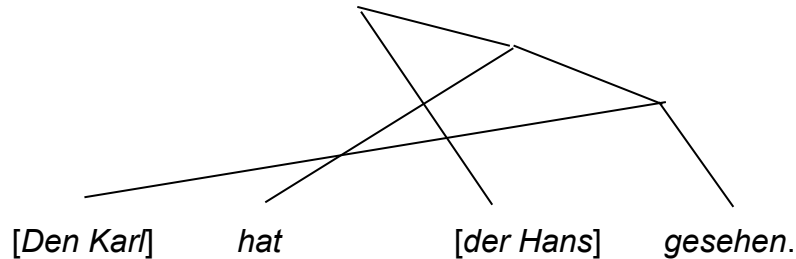
Dies ist die **Präzedenzrelation**.

Ein Baum hat die folgenden Eigenschaften:

- Es gibt einen obersten Knoten;  
alle anderen Knoten sind direkt oder indirekt mit diesem Knoten verbunden.
- Die Kanten zwischen den Knoten drücken die sogenannte **Dominanzrelation**  $D$  aus.  
Diese Relation ist eine **Halbordnung**.  
Die Dominanzrelation wird insbesondere als **transitiv** verstanden.  
Zum Beispiel dominiert der Knoten  $a$  den Knoten  $d$ , und der Knoten  $d$  dominiert den Knoten  $f$ ;  
damit dominiert  $a$  auch  $f$ .  
Wir schreiben z.B.  $\langle a, d \rangle \in D$ .
- Zwischen Knoten gibt es ferner die Relation der Reihenfolge oder **Präzedenz**  
(von links nach rechts; diese Relation wird nicht eigens durch Linien dargestellt).  
Diese Relation ist ebenfalls eine **Halbordnung**.  
Sie ist daher transitiv. Zum Beispiel geht der Knoten  $e$  dem Knoten  $f$  voran, und der Knoten  $f$  dem  
Knoten  $g$ ; also geht auch  $e$   $g$  voran.  
Wir schreiben z.B.  $\langle e, f \rangle \in P$ .
- Die Relationen der Dominanz und der Präzedenz schließen sich gegenseitig aus:  
Wenn zwei Knoten in Dominanz-Relation zueinander stehen, dann stehen sie nicht in Präzedenz-  
Relation zueinander, und umgekehrt.  
Für alle  $\langle x, y \rangle$  gilt: Wenn  $\langle x, y \rangle \in D$ , dann  $\langle x, y \rangle \notin P$ ,  
Wenn  $\langle x, y \rangle \in P$ , dann  $\langle x, y \rangle \notin D$
- Zwischen  $D$  und  $P$  gibt es einen Zusammenhang, der sich kreuzende Kanten verhindert:  
Wenn immer gilt:  $\langle x, y \rangle \in D$  und  $\langle x, z \rangle \in D$  und  $\langle y, z \rangle \in P$ ,  
dann gilt: Wenn  $\langle y, y' \rangle \in D$  und  $\langle z, z' \rangle \in D$ , dann  $\langle y', z' \rangle \in P$

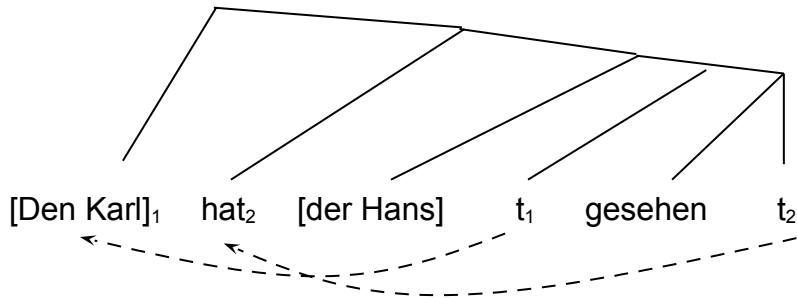
Eine Konsequenz dieses Baumbegriffes ist z.B., dass man bei wortstellungsfreien Sprachen wie dem Deutschen keine Analysen der folgenden Art vornehmen darf:

➤ *Den Karl hat der Hans gesehen.*



Stattdessen wird in der Regel angenommen, dass linguistische Strukturen mit **Spuren** erzeugt werden, die mit den vorkommenden Ausdrücken in Verbindung stehen.

Diese Beziehung kann man – metaphorisch gesprochen – so verstehen, dass die entsprechenden Ausdrücke von ihrer Ausgangsposition in ihre Endposition **bewegt** worden sind.



# Funktionen

## Funktionen als rechtseindeutige Relationen

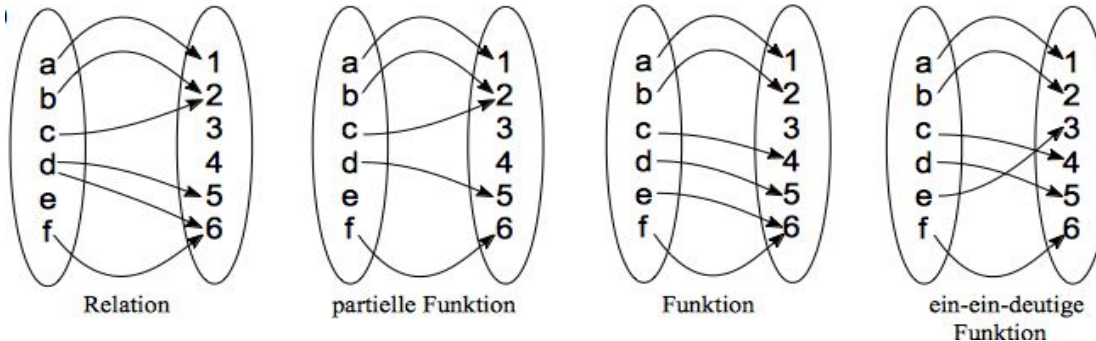
**Funktionen** sind Relationen, die **rechtseindeutig** sind.

Ein Element  $x$  darf immer nur mit genau einem Element  $y$  in der Verbindung  $\langle x, y \rangle$  stehen.

➤ Eine Relation  $R$  ist rechtseindeutig, wenn für alle  $x, y, z$  gilt:

Wenn  $\langle x, y \rangle \in R$  und  $\langle x, z \rangle \in R$ , dann gilt:  $y = z$

Pfeilnotation: “ $x \rightarrow y$ ” steht dafür, dass sich  $\langle x, y \rangle$  in der Relation befinden.



Eine Funktion, die nicht jedes Element der linken Menge abbildet, nennt man **partiell**;  
eine Funktion, die auch links-eindeutig ist, nennt man **ein-eindeutig**.

Beispiel für eine Funktion:

“ $y$  ist Vater von  $x$ ”. Jeder Mensch hat nur einen (biologischen) Vater.

Hingegen ist “ $y$  ist Tochter von  $x$ ” keine Funktion;  $x$  kann mehrere Töchter haben.

## Schreibweise

Wir geben Funktionen in der Regel mit den Buchstaben  $f$ ,  $F$  oder  $f$  an.

Die Eigenschaft der Rechtseindeutigkeit erlaubt eine neue Schreibweise, die aus dem Matheunterricht in der Schule bekannt ist.

Anstelle von  $\langle x, y \rangle \in f$  oder einer äquivalenten Notation können wir Folgendes schreiben:

➤  $y = f(x)$ .

Sprechweisen:

➤  $x$  ist das **Argument** und  $y$  der **Wert**.

➤  $f$  wird auf  $x$  **angewendet**.

➤  $f$  **bildet**  $x$  auf  $y$  **ab**.

Weitere Begriffe:

➤ **Definitionsbereich (domain)** einer Funktion  $f$ ,  $\text{DOM}(f)$ :

$\{x \mid \text{es gibt } y \text{ so, dass } \langle x, y \rangle \in f\}$

➤ **Wertebereich (range)** einer Funktion  $f$ ,  $\text{RNG}(f)$ :

$\{y \mid \text{es gibt } x \text{ so, dass } \langle x, y \rangle \in f\}$

Im Definitionsbereich von  $f$  sind also die möglichen Argumente von  $f$  enthalten, im Wertebereich von  $f$  die möglichen Werte von  $f$ .

Beispiel:

Der Definitionsbereich der Vater-Funktion ist die Menge der Personen (jeder hat einen Vater), der Wertebereich der Vater-Funktion ist die Menge der Väter.

## Beschreibung von Funktionen

Für die Beschreibung von Funktionen haben sich verschiedene Schreibweisen eingebürgert.

➤ Aufzählung

$f =$  Isaak  $\mapsto$  Abraham

Jakob  $\mapsto$  Isaak

Esau  $\mapsto$  Isaak

...

➤ Zuweisungsvorschrift:

$f:$  Väter  $\mapsto$  Personen,

$x \mapsto$  der Vater von  $x$ .

Die erste Zeile gibt die Menge an, die durch die Funktion, hier  $v$ , abgebildet werden.

Die zweite Zeile beschreibt für jedes  $x$  aus dieser Menge, auf welches Objekt  $x$  durch die Funktion  $v$  abgebildet wird.

➤ Lambda-Notation

$f = \lambda x[\text{Vater von } x]$

Wir nennen solche Ausdrücke **Lambda-Terme**.

Sie haben die folgende Struktur:

➤  $\lambda$  Variable [Beschreibung des Wertes der Variablen]

Die Lambda-Notation hat sich in der linguistischen Semantik weitgehend durchgesetzt. Sie macht es möglich, den Wert einer Funktion angewendet auf ein Argument anzugeben, indem man einfach die Lambda-Variable durch das Argument ersetzt (sog. Lambda-Konversion)

$$\begin{array}{ll} \lambda x[\text{Vater von } x](\text{Isaak}) & \lambda x[x^2 + x + 1](3) \\ = \text{Vater von Isaak} & = 3^2 + 3 + 1 \\ = \text{Abraham} & = 9 + 3 + 1 \\ & = 13 \end{array}$$

Man kann auch Funktionen definieren, die Funktionen als Werte liefern oder Funktionen als Argumente erwarten:

$$\begin{array}{ll} \lambda x[\lambda y[2 \cdot x + y]](4)(5) & \lambda f [f(2) + f(3)](\lambda x[x^2 + 1]) \\ = \lambda y[2 \cdot 4 + y](5) & = [\lambda x[x^2 + 1]](2) + \lambda x[x^2 + 1](3) \\ = [2 \cdot 4 + 5] & = [2^2 + 1 + 3^2 + 1] \\ = 13 & = 15 \end{array}$$

Bei  $\lambda x[\lambda y[\dots]]$  können wir auch die Klammern weglassen und einfach schreiben:  $\lambda x \lambda y[\dots]$ . Dann ist es aber wichtig, die Argumentreihenfolge zu beachten:

In dem Ausdruck

$$\lambda x \lambda y[2x + y](4)(5)$$

geht das Argument 4 in die Position von  $x$ , und das Argument 5 in die Position von  $y$ .

Wie wir gesehen haben, stellen Lambda-Terme eine Beziehung her zwischen einem Operator  $\lambda x$  und Vorkommen von Variablen [...x...].

Man sagt, dass der Operator die Vorkommen der Variablen **bindet**.

Welche Variable dabei gewählt wird, ist unerheblich; der Ausdruck  $\lambda x[x^2 + x = 1]$  und der Ausdruck  $\lambda y[y^2 + y + 1]$  stehen für dieselbe Funktion.

Es kann nun dazu kommen, dass Vorkommen derselben Variablen von unterschiedlichen Operatoren gebunden werden:

➤  $\lambda x[\lambda f [f(x) + 1](\lambda x[2x + 1])]$

In solchen Fällen ist es besser, eine Umbenennung der Variablen vorzunehmen, z.B. von  $x$  zu  $x'$ :

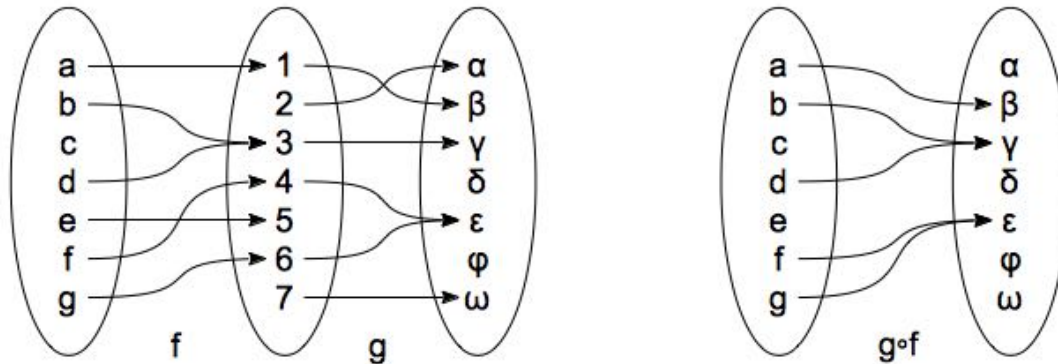
➤  $\lambda x[\lambda f [f(x) + 1](\lambda x'[2x' + 1])]$



## Verknüpfung von Funktionen

Wir können Funktionen nicht nur auf Argumente anwenden; wir können Funktionen (und allgemeiner Relationen) auch miteinander verknüpfen.

Beispiel:



Wir schreiben  $g \circ f$  für die Verknüpfung der Funktion  $f$  mit der Funktion  $g$ . Diese Operation ist wie folgt definiert:

➤ Für alle  $x$  im Definitionsbereich von  $f$ :  $g(f(x)) = g \circ f(x)$

## Charakteristische Funktionen

Eine Menge zu kennen bedeutet, die Elemente jener Menge zu kennen.

Wir müssen von jedem Objekt wissen, ob es in der Menge enthalten ist oder nicht.

Diese Information kann man als Funktion angeben,

die jedem Objekt in dem Universum genau einen von zwei **Wahrheitswerten** zuordnet: den Wert 1, falls das Objekt in der Menge enthalten ist, und den Wert 0, falls nicht.

Solche Funktionen heißen **charakteristische Funktionen** einer Menge.

Wir schreiben  $\chi_A$  ("chi-A") für die charakteristische Funktion der Menge A.

- $\chi_A$  = diejenige Funktion  $f$ , sodass gilt:  
für jedes  $x$ , wenn  $x \in A$ , dann  $f(x) = 1$ , sonst  $f(x) = 0$ .

Beispiel:

- Sei das Universum U die Menge der Vokale {a, e, i, o, u}.  
Dann gilt:  $\chi_{\{e, i\}} = \{\langle a, 0 \rangle, \langle e, 1 \rangle, \langle i, 1 \rangle, \langle o, 0 \rangle, \langle u, 0 \rangle\}$

Man kann also eine Menge A immer auch als eine charakteristische Funktion  $\chi_A$  angeben.

Damit können wir Mengen durch die Lambda-Notation ausdrücken:

- $\{x \mid x \text{ ist rot}\}$  kann auch ausgedrückt werden durch:  $\lambda x[x \text{ ist rot}]$

Der Lambda-Term  $\lambda x[x \text{ ist rot}]$  ist als die charakteristische Funktion der Menge  $\{x \mid x \text{ ist rot}\}$  zu verstehen.

Er bezeichne die Funktion, die jedem Objekt  $x$  den Wahrheitswert 1 zuweist, falls  $x$  rot ist, und den Wert 0, falls dies nicht der Fall ist.

## Logik

Ein weiterer wichtiger Werkzeugkasten, vor allem für die Semantik, sind die formalen logischen Sprachen.

### Aussagenlogik

Wir betrachten hier lediglich die Sprache der **Aussagenlogik**.

In dieser Sprache sind die einfachsten Ausdrücke Sätze, die einen **Wahrheitswert** haben, die also wahr oder falsch sein können.

Solche Sätze können durch sog. **Konnektoren** verknüpft werden.

## Die Konnektoren der Aussagenlogik

- Wenn  $\Phi$  ein Aussagesatz ist, dann ist  $\neg\Phi$  ein Aussagesatz, die **Negation** von  $\Phi$ , gelesen “nicht  $\Phi$ ”
- Wenn  $\Phi$  und  $\Psi$  Aussagesätze sind, dann ist  $[\Phi \wedge \Psi]$  ein Aussagesatz, die **Konjunktion** von  $\Phi$  und  $\Psi$ , gelesen “ $\Phi$  und  $\Psi$ ”
- Wenn  $\Phi$  und  $\Psi$  Aussagesätze sind, dann ist  $[\Phi \vee \Psi]$  ein Aussagesatz, die **Disjunktion** von  $\Phi$  und  $\Psi$ , gelesen “ $\Phi$  oder  $\Psi$ ”.
- Wenn  $\Phi$  und  $\Psi$  Aussagesätze sind, dann ist  $[\Phi \rightarrow \Psi]$  ein Aussagesatz, die (**materiale**) **Implikation** oder das **Konditional**, gelesen “Wenn  $\Phi$  dann  $\Psi$ ”
- Wenn  $\Phi$  und  $\Psi$  Aussagesätze sind, dann ist  $[\Phi \leftrightarrow \Psi]$  ein Aussagesatz, die (**materiale**) **Äquivalenz** oder das **Bikonditional**, gelesen “ $\Phi$  genau dann, wenn  $\Psi$ ”

Damit können wir die **wohlgeformte Formeln der Aussagenlogik** aufbauen. Ein Beispiel:

- $p_1$   $p_1, p_2, p_3$  seien einfache (atomare) Sätze
- $\neg p_1$
- $[\neg p_1 \vee p_2]$  Es handelt sich um eine **rekursive Definition**
- $\neg[\neg p_1 \vee p_2]$  des Begriffs „wohlgeformte Formel“:
- $[p_3 \rightarrow \neg[\neg p_1 \vee p_2]]$  Komplexe Formeln werden sukzessive auf
- $[p_1 \wedge [[p_3 \rightarrow \neg[\neg p_1 \vee p_2]]]]$  einfachere zurückgeführt.
- ...
- Bemerke: Die Syntax natürlicher Sprachen
- ist ebenfalls rekursiv.

Die Klammerungen sind nötig, um Ambiguität zu vermeiden:

- $\neg p_1 \vee p_2$  a.  $\neg [p_1 \vee p_2]$
- b.  $[\neg p_1 \vee p_2]$

## Die Interpretation der Aussagenlogik

Aussagesätze werden durch **Wahrheitswerte** interpretiert:

ihre Bedeutungen sind eine der beiden Wahrheitswerte 0 (falsch) oder 1 (wahr).

Über die Bedeutung der elementaren Aussagen  $p_1, p_2$  usw. kann die Aussagenlogik dabei nichts weiter sagen.

Sie zeigt aber, wie die Bedeutung von komplexen Sätzen auf die Bedeutung der Teilsätze zurückgeführt werden kann.

Diese Regeln der Zurückführung können durch sogenannte **Wahrheitwert-Tafeln** erfasst werden:

| $\Phi$ | $\neg\Phi$ |
|--------|------------|
| 0      | 1          |
| 1      | 0          |

| $\Phi$ | $\Psi$ | $[\Phi \wedge \Psi]$ |
|--------|--------|----------------------|
| 0      | 0      | 0                    |
| 0      | 1      | 0                    |
| 1      | 0      | 0                    |
| 1      | 1      | 1                    |

| $\Phi$ | $\Psi$ | $[\Phi \vee \Psi]$ |
|--------|--------|--------------------|
| 0      | 0      | 0                  |
| 0      | 1      | 1                  |
| 1      | 0      | 1                  |
| 1      | 1      | 1                  |

| $\Phi$ | $\Psi$ | $[\Phi \rightarrow \Psi]$ |
|--------|--------|---------------------------|
| 0      | 0      | 1                         |
| 0      | 1      | 1                         |
| 1      | 0      | 0                         |
| 1      | 1      | 1                         |

| $\Phi$ | $\Psi$ | $[\Phi \leftrightarrow \Psi]$ |
|--------|--------|-------------------------------|
| 0      | 0      | 1                             |
| 0      | 1      | 0                             |
| 1      | 0      | 0                             |
| 1      | 1      | 1                             |

- Die Disjunktion zweier Sätze ist genau dann wahr, wenn mindestens ein Teilsatz wahr ist (sog. **inklusive** Disjunktion; wir können auch eine **exklusive** Disjunktion formulieren).
- Das Konditional  $[\Phi \rightarrow \Psi]$  besagt: Wenn der Vordersatz  $\Phi$  wahr ist, dann ist auch der Folgesatz  $\Psi$  wahr. (Vorder- und Folgesatz heißen auch **Antezedens** und **Konsequens**).  
Wenn der Vordersatz falsch ist, ist die Implikation immer wahr.  
(Das ist beim natürlichsprachlichen *wenn-dann* anders).

Mithilfe dieser Regeln kann man nun den Wahrheitswert eines komplexen Satzes bestimmen, wenn die Wahrheitswerte der einfachen Sätze bekannt sind.

Beispiel; wir nehmen an, dass  $p_1$  wahr,  $p_2$  falsch und  $p_3$  wahr ist.

| $p_1$ | $p_2$ | $p_3$ | $\neg p_1$ | $[\neg p_1 \vee p_2]$ | $\neg[\neg p_1 \vee p_2]$ | $[p_3 \rightarrow \neg[\neg p_1 \vee p_2]]$ | $[p_1 \wedge [p_3 \rightarrow \neg[\neg p_1 \vee p_2]]]$ |
|-------|-------|-------|------------|-----------------------|---------------------------|---------------------------------------------|----------------------------------------------------------|
| 1     | 0     | 1     | 0          | 0                     | 1                         | 1                                           | 1                                                        |

Wir haben gesehen, dass die Sprache der Aussagenlogik jeder zulässigen Zeichenfolge genau eine syntaktische Struktur zuweist. Sie erlaubt also keine Ambiguitäten wie die natürliche Sprache. Sie haben immer genau einen Wahrheitswert. Dies zeigt sich in den folgenden Beispielen:

$$\begin{array}{cc} \neg [p_1 \wedge p_2] & [\neg p_1 \wedge p_2] \\ \begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} & \begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \\ 1 & 0 \end{array}$$

Der linke Satz erhält den Wahrheitswert wahr, der rechte falsch; die beiden Sätzen bestehen zwar aus denselben Zeichen, sie sind jedoch unterschiedlich geklammert.

## Tautologien und Kontradiktionen

Die Logik ist vor allem an Tautologien und Kontradiktionen interessiert, d.h. an Sätzen, die aus logischen Gründen wahr oder falsch sind. Ein Beispiel für eine Tautologie ist der folgende Satz.

$$[p_1 \vee [[p_3 \rightarrow \neg [p_1 \wedge p_2]]]]$$

Bei 3 elementaren Sätzen gibt es insgesamt  $2^3$ -viele Möglichkeiten, also 8 Möglichkeiten, den Sätzen Wahrheitswerte zuzuweisen. Unter jeder dieser Möglichkeiten ist der Gesamtsatz wahr:

| $p_1$ | $p_2$ | $p_3$ | $[p_1 \wedge p_2]$ | $\neg[p_1 \wedge p_2]$ | $[p_3 \rightarrow \neg[p_1 \wedge p_2]]$ | $[p_1 \vee [p_3 \rightarrow \neg[p_1 \wedge p_2]]]$ |
|-------|-------|-------|--------------------|------------------------|------------------------------------------|-----------------------------------------------------|
| 0     | 0     | 0     | 0                  | 1                      | 1                                        | 1                                                   |
| 0     | 0     | 1     | 0                  | 1                      | 1                                        | 1                                                   |
| 0     | 1     | 0     | 0                  | 1                      | 1                                        | 1                                                   |
| 0     | 1     | 1     | 0                  | 1                      | 1                                        | 1                                                   |
| 1     | 0     | 0     | 0                  | 1                      | 1                                        | 1                                                   |
| 1     | 0     | 1     | 0                  | 1                      | 1                                        | 1                                                   |
| 1     | 1     | 0     | 1                  | 0                      | 1                                        | 1                                                   |
| 1     | 1     | 1     | 1                  | 0                      | 0                                        | 1                                                   |

Eine **Kontradiktion** ist ein Satz, der unter jeder möglichen Zuweisung von Wahrheitswerten für die elementaren Sätze falsch ist. Es gilt natürlich, dass die Negation einer Tautologie eine Kontradiktion ist, und umgekehrt die Negation einer Kontradiktion eine Tautologie.

## Tautologien und logische Folgerung

Wie können wir nachweisen, dass eine logische Folgerung  $\Phi \Rightarrow \Psi$  gerechtfertigt ist?

Nach der Definition der logischen Folgerung muss gelten: In jedem Fall, in dem  $\Phi$  wahr ist, ist auch  $\Psi$  wahr. Dies können wir nun aber im einzelnen überprüfen.

Ein Beispiel: Wir wollen zeigen, dass aus der Prämisse  $[\neg p_1 \wedge [p_2 \Rightarrow p_1]]$  die Konklusion  $\neg p_2$  folgt. Hierzu betrachten wir alle Möglichkeiten, für die die Prämisse wahr ist.

| $p_1$ | $p_2$ | $[p_2 \rightarrow p_1]$ | $\neg p_1$ | $[\neg p_1 \wedge [p_2 \rightarrow p_1]]$ | $\Rightarrow$ | $\neg p_2$ |
|-------|-------|-------------------------|------------|-------------------------------------------|---------------|------------|
| 0     | 0     | 1                       | 1          | 1                                         |               | 1          |
| 0     | 1     | 0                       | 1          | 0                                         |               | 0          |
| 1     | 0     | 1                       | 0          | 0                                         |               | 1          |
| 1     | 1     | 1                       | 0          | 0                                         |               | 0          |

Es gibt hier nur einen einzigen Fall, zu dem die Prämisse wahr ist; für diesen Fall ist auch das Konklusion wahr, und damit ist die logische Folgerung gültig.

Wie dieses Beispiel andeutet, gibt es einen engen Zusammenhang zwischen der logischen Folgerung (und der logischen Äquivalenz) und dem Begriff der Tautologie. Es gilt nämlich:

- Die logische Folgerung  $\Phi \Rightarrow \Psi$  besteht gdw.  $[\Phi \Rightarrow \Psi]$  eine Tautologie ist.
- Die logische Äquivalenz  $\Phi \Leftrightarrow \Psi$  besteht gdw.  $[\Phi \Leftrightarrow \Psi]$  eine Tautologie ist.



Die Ähnlichkeit zwischen den Symbolen für logische Folgerung und Äquivalenz zwischen Sätzen,  $\rightarrow$  und  $\leftrightarrow$ , und den Symbolen  $\Rightarrow$  und  $\Leftrightarrow$  zum Aufbau von Sätzen sind also motiviert.

Wir können zeigen, dass der Satz  $[[\neg p_1 \wedge [p_2 \rightarrow p_1]] \rightarrow \neg p_2]$  eine Tautologie ist:

| $p_1$ | $p_2$ | $[p_2 \rightarrow p_1]$ | $\neg p_1$ | $[\neg p_1 \wedge [p_2 \rightarrow p_1]]$ | $\neg p_2$ | $[[\neg p_1 \wedge [p_2 \rightarrow p_1]] \rightarrow \neg p_2]$ |
|-------|-------|-------------------------|------------|-------------------------------------------|------------|------------------------------------------------------------------|
| 0     | 0     | 1                       | 1          | 1                                         | 1          | 1                                                                |
| 0     | 1     | 0                       | 1          | 0                                         | 0          | 1                                                                |
| 1     | 0     | 1                       | 0          | 0                                         | 1          | 1                                                                |
| 1     | 1     | 1                       | 0          | 0                                         | 0          | 1                                                                |

Im folgenden Beispiel wird eine Äquivalenz nachgewiesen:

Zeige:  $\neg[\neg p_1 \wedge p_2] \Leftrightarrow [\neg p_2 \vee p_1]$

Beweis: Wir zeigen:  $[\neg[\neg p_1 \wedge p_2] \leftrightarrow [\neg p_2 \vee p_1]]$  ist eine Tautologie.

| $p_1$ | $p_2$ | $\neg p_1$ | $[\neg p_1 \wedge p_2]$ | $\neg[\neg p_1 \wedge p_2]$ | $\neg p_2$ | $[\neg p_2 \vee p_1]$ | $[\neg[\neg p_1 \wedge p_2] \leftrightarrow [\neg p_2 \vee p_1]]$ |
|-------|-------|------------|-------------------------|-----------------------------|------------|-----------------------|-------------------------------------------------------------------|
| 0     | 0     | 1          | 0                       | 1                           | 1          | 1                     | 1                                                                 |
| 0     | 1     | 1          | 1                       | 0                           | 0          | 0                     | 1                                                                 |
| 1     | 0     | 0          | 0                       | 1                           | 1          | 1                     | 1                                                                 |
| 1     | 1     | 0          | 0                       | 1                           | 0          | 1                     | 1                                                                 |

Interessant ist nun folgendes:

Wenn ein Satz  $\Phi$  eine Kontradiktion ist, d.h. immer falsch ist, dann ist der Satz  $[\Phi \rightarrow \Psi]$  immer wahr, egal welche Wahrheitswerte  $\Psi$  annimmt.

Das folgt aus der Wahrheitstafel für das Konditional,  $\rightarrow$

Dann aber der Schluß  $\Phi \Rightarrow \Psi$  logisch gültig.

Wir sagen: Aus einem Widerspruch folgt jeder beliebige Satz;

auf Latein: Ex falso quodlibet.

Dieses Prinzip ist fundamental für die Beweisführung in der Logik.

## Aussagenlogische Gesetze

Die Interpretation der Satzkonnectoren in der Aussagenlogik bedingen, dass eine Reihe von Gesetzmäßigkeiten zwischen Formeln der Aussagenlogik bestehen. Hier sind diese Gesetzmäßigkeiten unter ihrer üblichen Bezeichnung zusammengestellt:

- a. Idempotenz:  $[\Phi \wedge \Phi] \Leftrightarrow \Phi$   
 $[\Phi \vee \Phi] \Leftrightarrow \Phi$
- b. Kommutativität:  $[\Phi \wedge \Psi] \Leftrightarrow [\Psi \wedge \Phi]$   
 $[\Phi \vee \Psi] \Leftrightarrow [\Psi \vee \Phi]$
- c. Assoziativität:  $[\Phi \wedge [\Psi \wedge \Omega]] \Leftrightarrow [[\Phi \wedge \Psi] \wedge \Omega]$   
 $[\Phi \vee [\Psi \vee \Omega]] \Leftrightarrow [[\Phi \vee \Psi] \vee \Omega]$
- d. Distributivität:  $[\Phi \wedge [\Psi \vee \Omega]] \Leftrightarrow [[\Phi \wedge \Psi] \vee [\Phi \wedge \Omega]]$   
 $[\Phi \vee [\Psi \wedge \Omega]] \Leftrightarrow [[\Phi \vee \Psi] \wedge [\Phi \vee \Omega]]$
- e. De Morgan:  $\neg[\Phi \wedge \Psi] \Leftrightarrow [\neg\Phi \vee \neg\Psi]$   
 $\neg[\Phi \vee \Psi] \Leftrightarrow [\neg\Phi \wedge \neg\Psi]$
- f. Konditionalgesetze:  $[\Phi \Rightarrow \Psi] \Leftrightarrow [\neg\Phi \vee \Psi]$   
 $[\Phi \Rightarrow \Psi] \Leftrightarrow [\neg\Psi \Rightarrow \neg\Phi]$
- g. Bikonditionalgesetz:  $[\Phi \leftrightarrow \Psi] \Leftrightarrow [[\Phi \Rightarrow \Psi] \wedge [\Psi \Rightarrow \Phi]]$

Die letzten Gesetze zeigen, dass man mit weniger Satzkonnectoren auskommt.

Wir können zum Beispiel  $\leftrightarrow$  mithilfe von  $\wedge$  und  $\rightarrow$  definieren, und wir können  $\Rightarrow$  mithilfe von  $\neg$  und  $\vee$  definieren.

Wenn wir nun  $\top$  und  $\perp$  als einen tautologischen bzw. kontradiktorischen Satz nehmen, dann können wir zusätzlich die folgenden Gesetzmäßigkeiten formulieren:

Komplementgesetze:

|                                                |                                     |
|------------------------------------------------|-------------------------------------|
| $[\Phi \vee \neg\Phi] \Leftrightarrow \top$    |                                     |
| $[\Phi \wedge \neg\Phi] \Leftrightarrow \perp$ | (Satz vom ausgeschlossenen Dritten) |
| $\neg\neg\Phi \Leftrightarrow \Phi$            | (Doppelte Negation)                 |
| $\neg\perp \Leftrightarrow \top$               |                                     |

Mithilfe von solchen aussagenlogischen Gesetzen lässt es sich schneller und eleganter beweisen, ob zwei Ausdrücke logisch äquivalent sind, oder ob einer aus einem anderen folgt. Aufgabe ) kann nun wie folgt gelöst werden:

|                                                |                        |
|------------------------------------------------|------------------------|
| $\neg[\neg p_1 \wedge p_2]$                    |                        |
| $\Leftrightarrow [\neg\neg p_1 \vee \neg p_2]$ | (Gesetz von de Morgan) |
| $\Leftrightarrow [p_1 \vee \neg p_2]$          | (Doppelte Negation)    |
| $\Leftrightarrow [\neg p_2 \vee p_1]$          | (Kommutativität)       |

Desiderata:

Statistik: z-scores

Elementare Wahrscheinlichkeitsrechnung?

# Aufgaben

- Ist in den folgenden Fällen (i) oder (ii) die Negation des Satzes?
  - Hier regnet es immer. (i) Hier regnet es nie.  
(ii) Hier regnet es nicht immer.
  - Jemand hat mir geholfen. (i) Jemand hat mir nicht geholfen.  
(ii) Niemand hat mir geholfen.
  - Es ist noch hell. (i) Es ist noch nicht hell.  
(ii) Es ist nicht mehr hell.
  - Viele haben geklatscht. (i) Viele haben nicht geklatscht.  
(ii) Nicht viele haben geklatscht.
- Definieren Sie die Beziehungen “ $\Phi$  und  $\Psi$  sind äquivalent”, “ $\Phi$  und  $\Psi$  sind konträr”, und “ $\Phi$  und  $\Psi$  sind kontradiktorisch” mithilfe der logischen Folgerung,  $\Rightarrow$
- Geben Sie die logischen Verhältnisse zwischen den folgenden Sätzen an (Implikation, Äquivalenz, Kontrarität, Kontradiktion, Kontingenz).
  - Das Glas ist leer.*
  - Das Glas ist halb voll.*
  - Das Glas ist halb leer.*
  - Das Glas ist voll.*
  - Das Glas ist nicht leer.*
  - Das Glas ist nicht voll.*
- Welche der folgenden Zeichenketten sind wohlgeformte Formeln (Sätze) der Aussagenlogik?
  - $[p_1 \Rightarrow p_2]$
  - $[p_1 \vee p_2 \wedge p_3]$
  - $p_1 \Rightarrow [p_2 \vee p_3]$
  - $[p_1 \vee p_2 \Rightarrow p_3]$
  - $[p_1 \wedge p_2] \Rightarrow p_1$
  - $[p_1 \vee p_3] \Leftrightarrow p_4]$
- Berechnen Sie den Wahrheitswert des folgenden Satzes, unter der Annahme der folgenden Wahrheitswerte für die Teilsätze:  $p_1: 0$ ,  $p_2: 1$ ,  $p_3: 0$ ,  $p_4: 1$   
 $[ \leftarrow [[p_1 \vee p_2] \wedge \leftarrow p_4] \Rightarrow [p_1 \vee \leftarrow p_3]]$
- Welche der folgenden aussagenlogischen Formeln sind Tautologien, Kontradiktionen oder kontingente Sätze?
  - $[[p_1 \wedge p_2] \Rightarrow \leftarrow p_2]$
  - $[[p_1 \vee p_2] \Rightarrow \leftarrow p_2]$
  - $[p_1 \wedge \leftarrow [p_1 \vee p_2]]$
  - $[[p_1 \vee p_2] \wedge [p_2 \Rightarrow p_1]]$
  - $[\leftarrow [p_1 \vee p_2] \Rightarrow \leftarrow p_2]$
  - $[[p_1 \Rightarrow p_2] \Rightarrow [p_2 \Rightarrow p_1]]$
- Desambiguieren Sie die folgenden Sätze mithilfe der aussagenlogischen Notation (wobei  $p_1$ : ‘Es regnet’,  $p_2$ : ‘Es blitzt.’,  $p_3$ : ‘Es donnert’).
  - Es regnet und es blitzt oder es donnert.*
  - Es regnet und blitzt nicht.*

