

Technische Übungen

Prof. Dr. Manfred Krifka

5220156 BA Germanistische Linguistik

Modul 1: Technische Hilfsmittel Di 12-14, DOR24, 1.102

Zweck der Übung ist es, den Studierenden die Werkzeuge und Techniken zu vermitteln, die für die Rezeption linguistischer Fachartikel, die Durchführung von eigenen Forschungen und die Präsentation fremder und eigener Arbeiten in Referaten und Papieren nötig sind. Dies schließt die folgenden Themen ein: Grundlagen der wissenschaftlichen Methode, Organisations- und Publikationsformen der Sprachwissenschaft, Recherchetechniken in Bibliotheken und im Internet, Argumentationsformen in sprachwissenschaftlichen Arbeiten, die Befragung der eigenen Intuition und der von Informanten, grundlegende korpuslinguistische, phonetische und psycholinguistische Methoden, essentielle mathematische Begriffe (Mengen, Funktionen, Phrasenstrukturgrammatiken), elementare statistische Auswertungsverfahren, Planung und Durchführung einer referierenden oder eigenständigen wissenschaftlichen Arbeit, sowie die Verwendung von Textverarbeitungs- und Präsentationssoftware für Referate und Hausarbeiten.

Für die Veranstaltung wurde ein Tutorium eingerichtet:
Mittwoch 8 – 10, Raum 1.401, Sören Schalowski.

Koordinaten

- Sprechstunde Mi 13-15, bitte bei Frau Klein anmelden, und nach Vereinbarung
- E-Mail: krifka@rz.hu-berlin.de, bitte "Technische Hilfsmittel" in der Subjekt-Zeile
- Webseite: <http://amor.rz.hu-berlin.de/~h2816i3x/lehrstuhl.html>
- Moodle-Seite des Kurses: <http://lms.cms.hu-berlin.de/moodle/>, Schlüssel "Hilfsmittel" direkt: <http://lms.hu-berlin.de/moodle/course/view.php?id=7492>, auf der Seite werden Skripten und Texte verfügbar gemacht.
- Telefon: 030-20939670
- Büro: Hegelplatz 2, 3.303
- Tutor: Sören Schalowski, schalows@student.hu-berlin.de
- Hilfskraft: Sophia Döring, sophiadoering@gmx.de
- Sekretärin: Anina Klein
Büro: 3.306, Tel. 030-20939639, e-mail: anina.klein@cms.hu-berlin.de

Überblick über die Veranstaltung:

14. 10.	Einführung in die Veranstaltung. Wissenschaftliche Methode; Publikationstypen: Zeitschriften, Monographien, Sammelwerke, Handbücher, Lehrbücher, populärwissenschaftliche Sachbücher. Veranstaltungen: Workshops, Tagungen, Kongresse
21. 10.	Literaturrecherche: Bibliotheken in Berlin; Online-Kataloge, Online-Literatur technische Werkzeuge: Web-Browser, Adobe Reader, Google Scholar Bibliographieprogramme, Zotero.
28. 10.	Textverarbeitung (MS Word, OpenOffice, LaTeX); forgeschr. Arbeiten mit MS Word: Formatvorlagen, Numerierungen Konventionen des wissenschaftlichen Schreibens: Titel, Abstract, gegliederter Text, Bibliographie, Index, Inhaltsverzeichnis Linguistische Konventionen: Numerierungen, Objektsprache, Zitierweisen
4. 11.	Textverarbeitung: Verweise, Sonderzeichen, Tabellen, Grafiken; Textmanipulation: Suchern/Ersetzen, Änderungen anzeigen, Korrekturhilfen Makros, Autokorrektur, Bibliotheksverwaltung, Drucken: regulär und pdf Tabellenkalkulation: Grundlagen von Excel
11. 11.	Darstellung linguistischer Objekte: IPA und IPA-Sonderzeichen, Glossierungskonventionen, Baumdarstellungen, tabellarische Darstellungen,
18. 11.	Referat: Handout, Powerpoint-Präsentation. Seminararbeit: Planen, Durchführen, Schreiben.
25. 11.	Stoff-Wiederholung oder Bibliotheksbesuch?
2. 12.	Korpuslinguistik: Verfügbare Korpora, Abfragemöglichkeiten
9. 12.	Korpuslinguistik: Möglichkeiten bestimmter Korpora (DWDS, TIGER)
16. 12.	Annotationswerkzeuge: Exmeralda
	Akademische Ferien
6. 1.	Werkzeuge für phonetische Untersuchungen: PRAAT
13. 1.	Experimente: Planung und Durchführung von linguistischen Befragungen und psycholinguistischen Experimenten
20. 1.	Beschreibende Statistik: Datentypen, beschreibende Parameter von Daten, Darstellung von Daten (mit R)
27. 1.	Schließende Statistik. Stichproben, Wahrscheinlichkeit, Signifikanz, grundlegende Tests.
3. 2.	Werkzeuge der Modellierung: Mengen, Relationen, Funktionen.
10. 2.	Werkzeuge der Modellierung: Phrasenstrukturgrammatiken und Automaten

1. Wissenschaftliche Methode und Wissenschaftskommunikation

1.1 Die wissenschaftliche Methode

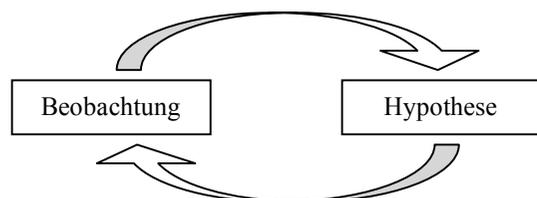
Was ist Wissenschaft? Wir können hier nicht eine Einführung in die Wissenschaftsphilosophie geben (siehe hierzu z.B. W. Balzer, *Die Wissenschaft und ihre Methoden*, 2002). Wir wollen aber mit einer Darstellung der wissenschaftlichen Methode beginnen.

Wir beginnen mal nicht direkt mit den alten Griechen, sondern mit Abū 'Alī al-Ḥasan ibn al-Ḥasan ibn al-Haytham, auch bekannt als Alhazen (965-1039), der nicht nur in einer ganzen Reihe von eigenständigen wissenschaftlichen Beiträgen etwa zur Optik bekannt ist, sondern der auch als Begründer einer reflektierten wissenschaftlichen Methode gilt.



Die Grundidee der wissenschaftlichen Methode kann man als einen Zyklus von Beobachtung, Hypothese und Beobachtung beschreiben:

- Wir beobachten ein Phänomen A.
- Wir versuchen das Phänomen A mit einer Hypothese H zu erklären.
- Wir folgern: Wenn Hypothese H gilt, dann sollten wir Phänomen B beobachten; wir machen eine Voraussage.
- Wir versuchen, Phänomen B zu beobachten (in der Natur oder in einem Experiment).
- Wenn wir tatsächlich B beobachten, können wir dies als eine Stärkung der Hypothese H auffassen. Wenn wir stattdessen B' beobachten, müssen wir Hypothese H modifizieren.



Dabei ist es von grundsätzlicher Bedeutung, dass die Beobachtungen und auch Folgerungen der Art "Wenn H gilt, dann sollte man B beobachten" intersubjektiv überprüfbar sind. Wissenschaft ist ein essentiell demokratisches Unternehmen. Das heißt, Experimente müssen wiederholbar und nachvollziehbar sein.

Dies schließt ein, dass die wissenschaftliche Diskussion öffentlich ist, d.h. dass wissenschaftliche Ergebnisse (Beobachtungen und Hypothesen) öffentlich zugänglich gemacht werden:

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

1.2 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

1.3 Sprachwissenschaftliche Gesellschaften

Es gibt heute zahlreiche sprachwissenschaftliche Gesellschaften; hier einige wichtige:

- Deutsche Gesellschaft für Sprachwissenschaft (DGfS), <http://www.dgfs.de/>
Jährliche Jahrestagungen, 2010 an der HU; 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/>

1.4 Sprachwissenschaftliche Zeitschriften

Schon in dem relativ kleinen Feld der Sprachwissenschaften finden wir eine große Anzahl wissenschaftlicher Zeitschriften:

Wichtige (meist) deutschsprachige Zeitschriften

- Zeitschrift für Sprachwissenschaft (ZS), das Organ der DGfS: <http://www.degruyter.de/journals/zs/detail.cfm>
- Linguistische Berichte (LB): http://www.buske.de/index.php?cPath=4_13&content=recherche
- Zeitschrift für germanistische Linguistik (ZGL): <http://www.degruyter.de/journals/zgl/detail.cfm>
- Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB): <http://www.mediaevum.de/zeitschriften/pbb4.php>

Wichtige internationale englischsprachige Zeitschriften

- Language (Lg), das Organ der LSA: <http://www.lsadc.org/info/pubs-language.cfm>
- Linguistics: <http://www.degruyter.de/journals/linguistics/detail.cfm>
- 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.de/journals/theolin/detail.cfm>
Daneben gibt es Dutzende von Zeitschriften mit engerem Fokus.

1.5 Nachschlagewerke: Lexika und Handbücher

Lexika

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

- Hadumod Bußmann, *Lexikon der Sprachwissenschaft*. 3. Aufl.. Kröner, Stuttgart 2002.

- Helmut Glück, *Metzler Lexikon Sprache*, 3. Aufl. 2005.
Zugang zur CD über: <http://www2.hu-berlin.de/cd-rom-service/index.php4?frame=start.php4?progname=digibib34>
- R.E. Asher & J.E. Simpson (1993), *The Encyclopedia of Language and Linguistics*, 10 Vols., Pergamon Press.

Online-Lexika:

- Norbert Fries, Online Lexikon Linguistik, <http://lexikon.anaman.de/lexikon.htm>
- Glottopedia, im Wiki-Format: <http://www.glottopedia.org/>

Handbücher

Bei Handbüchern handelt es sich um Werke mit umfangreicheren Artikeln zu bestimmten Spezialgebieten der Sprachwissenschaft. Sie erlauben es im besten Fall, 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 u.a.: Dialektologie, Sprachgeschichte, Soziolinguistik, Computerlinguistik, 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.

1.6 Wissenschaftliche Monographien und Sammelwerke

Buchveröffentlichungen zu Spezialthemen, Monographien werden zu bestimmten Themen geschrieben; in der Regel sind sie länger als Zeitschriftenartikel. Sammelwerke versammeln verschiedene Artikel, oft zu einem Thema oder einem Themenkomplex, und oft auch von verschiedenen Autoren.

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

1.7 Wissenschaftliche Qualifikationsarbeiten

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

- Dissertationen zum Erlangen des Doktorgrades
- Magister- oder Diplomarbeiten (“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, eine kleine Anzahl an verschiedenen Bibliotheken zu deponieren. Dissertationen der USA sind ferner seit 1938 per Mikrofilm verfügbar (siehe UMI, <http://www.proquest.com/brand/umi.shtml>).

Viele Dissertationen und Magisterarbeiten sind heute über das Internet von Webseiten abrufbar; es gibt insbesondere auch zentral geführte Webseiten, die eine Kontinuität der Präsentation anstreben.

1.8 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*, die in jährlichen Jahresbänden und übers Internet verfügbar ist: <http://www.blonline.nl/searchfields1.htm>

1.9 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.

1.10 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.

1.11 Konferenzen, Sommerschulen, Diskussionsforen

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

- Jahrestagung der DGfS (Ende Februar)
- Annual Meeting, Linguistic Society of America (Anfang Januar)
- International Congress of Linguists (CIL), alle 5 Jahre
- Syntax of the World’s Languages III, FU Berlin, September 2008
- International Conference of Quotation II, ZAS Berlin, 16-18. 10. 2008

Über aktuelle Veranstaltungen in Berlin und Potsdam informiert die folgende Webseite:

- http://www.zas.gwz-berlin.de/index.html?events_berlin

Neben Konferenzen sind – vor allem für Studenten – verschiedene Sommerschulen zu erwähnen, bei denen auch Studienleistungen erbracht werden können. Beispiele:

- Die Sommerschule der LSA (alle 2 Jahre),
- die Sommerschule der DGfS (alle 3 Jahre),
- die jährliche Sommerschule der GLOW (“Generative Linguistics in the Old World”, <http://glow.uit.no/>)
- die European Summer School of Language, Logic and Information (ESSLLI)

Das wichtigste Internet-Diskussionsforum ist die *Linguist List*. Hier gibt es Diskussionen zu speziellen Themen, Umfragen, Buchbesprechungen, Hinweise auf Konferenzen, Stellenangebote, Software und vieles mehr: <http://www.linguistlist.org/issues/>

2. Literatur-Recherche

2.1 Relevante Bibliotheken in Berlin

Auch wenn sich hier darum bemüht wurde, die wichtigsten Punkte zu den jeweiligen Bibliotheken und ihrer Nutzung aufzulisten, so ist es doch von großer Wichtigkeit, selber einen Blick auf die einzelnen Homepages zu werfen und sich so einen eigenen Überblick über die Bibliotheken selbst und deren Leistungsangebot zu verschaffen. Die Auflistung ist daher als ein erster Appetizer zu sehen, der den Einstieg erleichtern soll! Darüber hinaus erhebt die Liste keinen Anspruch auf Vollständigkeit. Vielmehr werden Bibliotheken aufgeführt, die sich bei der Recherche sprachwissenschaftlicher Literatur als hilfreich erwiesen haben.

2.1.1 Zentralbibliothek der Humboldt-Universität (ZUB)

Anschrift: Hessische Str. 1-2 (provisorischer Standort) → U6 Zinnowitzer Str.

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 gibt es auf der Homepage und an den Bibliotheksschaltern.

Einführungskurs online: <http://www.ub.hu-berlin.de/onlinekurs/>

Anmeldung Schulungen: <http://www.ub.hu-berlin.de/service/schulungen/>

Liste aller ZwB: <http://www.ub.hu-berlin.de/bibliothek/zweigbibliotheken/>

2.1.2 Zweigbibliothek Germanistik der HU (ZwB Germanistik)

Anschrift: Dorotheenstr. 24 (DOR 24) → U6 Friedrichstr.

Homepage: <http://www.ub.hu-berlin.de/bibliothek/zweigbibliotheken/german/german.html>

Online-Katalog: Über Online-Katalog der ZUB

Anmeldung/Ausleihe:

Wie in der ZUB funktioniert dies hier über den Studierendenausweis. Jedoch handelt es sich in der ZwB Germanistik meist um Präsenzbestände und eine Ausleihe ist i.d.R. unter der Woche ab 17 Uhr bis zum folgenden Tag um 12 Uhr oder über das Wochenende (Fr ab 17 bis Mo 12 Uhr) möglich. Teilweise ist jedoch auch ein Freihandbestand vorhanden (Ausleihe: 4 Wochen). Um welche Typ es sich bei einer Publikation handelt, ist dem Eintrag im Online-Katalog zu entnehmen.

Einführung Recherche:

http://www.ub.hu-berlin.de/bibliothek/zweigbibliotheken/german/german_hilfe.pdf

Weitere Informationen:

Die ZwB Germanistik beinhaltet die Teilbibliotheken (TB) Skandinavistik und Historisch-Vergleichende Sprachwissenschaft.

2.1.3 Zweigbibliothek Fremdsprachliche Philologen der HU (ZwB Fremdsprachl. Philo.)

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

→ U6 Friedrichstr.

Homepage: <http://www.ub.hu-berlin.de/bibliothek/zweigbibliotheken/philologie/frephil.html>

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

2.1.4 Bibliothek des Zentrums für Allgemeine Sprachwissenschaft (ZAS)

Schützenstrasse 18, 2. Stock.

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

Online-Katalog: <http://dbserv.gwz-berlin.de/cgi-bin/acwww25/maske.pl?db=zas>

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.

2.1.5 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>

2.1.6 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.

2.2 Online-Angebote: ebenfalls eine kleine Auswahl

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.

2.2.1 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?bib_id=ubhub&lett=l&colors=&ocolors=

Online-Tutorial: http://www.ub.hu-berlin.de/onlinekurs/ik_modul_c/05_01_dbis_01.html

2.2.2 Elektronische Zeitschriftenbibliothek der HU (EZB)

Über die EZB ist es möglich, einen Überblick darüber zu erlangen, welche Zeitschriften online von der HU aus zugänglich sind.

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 folgende Webseite:

https://ssl.cms.hu-berlin.de/dana-na/auth/url_default/welcome.cgi

2.2.3 Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV)

Hierbei handelt es sich um das 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/>

Online-Katalog:

<http://vs13.kobv.de/V?portal=KOBV&institute=KOBV&func=meta-1&mode=advanced>

2.2.4 Zeitschriftendatenbank (ZDB)

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

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

Online-Katalog: <http://dispatch.opac.d-nb.de/DB=1.1/>

2.2.5 Karlsruher Virtueller Katalog (KVK)

Hierbei handelt es sich um eine Art 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. Beim KVK handelt es sich also nicht um eine eigenen Datenbank. Er verknüpft eine Vielzahl von Zielsystemen, aus denen ausgewählt werden kann.

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

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

2.2.6 Google Scholar

Mithilfe des Web-Interfaces Google Scholar können Sie Internet-Suchen auf wissenschaftliche Webseiten einschränken. Insbesondere erfüllt Google Scholar auch die Funktionen eines Zitationsindizes: Man kann herausfinden, welche wissenschaftlichen Beiträge von welchen anderen zitiert werden. Dies ist gerade bei fortgeschrittener wissenschaftlicher Arbeit wichtig, da man auf diese Weise nachverfolgen kann, wie bestimmte empirische Befunde und Hypothesen weiter diskutiert wurden.

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

2.2.7 Google Print

Google Print ist das größte einer Reihe von Projekten, Druckwerke über das Internet verfügbar zu machen. Neuere Werke sind aus Copyright-Gründen in der Regel nur in Auszügen oder in kleinen „Snippets“ verfügbar. Dennoch ist dies ein sinnvolles Instrument, um nach dem Vorkommen von Wörtern in Büchern zu suchen oder auch, um einen ersten Eindruck von einem Werk zu gewinnen.

Homepage: <http://www.google.de/books?hl=de>, <http://print.google.com/>

2.2.8 Freie Suche

Zahlreiche wissenschaftliche Arbeiten sind auf den Webseiten von Instituten, auf eigenen Server-Seiten zu bestimmten Themen oder auf persönlichen Webseiten verfügbar. Man kann

diese mithilfe von Suchmaschinen finden (Namen von Autoren, Schlüsselwörter von Titeln, Schlüsselphrasen (mit Anführungszeichen eingeben)).

2.3 Technische Werkzeuge: Web-Browser und PDF-Programme

Es gibt verschiedene Web-Browser (wie Internet Explorer, Firefox, Safari). Empfohlen: Firefox, ein "offener" Browser.

Wichtig: Legen Sie sich ein gut organisiertes Verzeichnis von "Bookmarks" an, das spart Arbeit!

2.4 Bibliographie-Programme

Es gibt 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/>
- Zotero (freie Version, siehe eigene Darstellung) <http://www.zotero.org/>
- Bookends (für Macintosh, kostenpflichtig), <http://www.sonnysoftware.com/>

Hier stelle ich Bookends vor, das ich selbst verwende.

2.5 Zotero

1. WAS IST ZOTERO?

Zotero ist eine kostenlose Erweiterung für den Firefox-Browser, mit der man unterschiedlicher Online- und Offline-Quellen sammeln, verwalten und zitieren kann. Es ist hilfreich für wissenschaftliches Arbeiten, da es die Bearbeitung von bibliografischen Angaben und Literaturlisten unterstützt.

Zotero wurde vom Center for History and New Media der George Mason University entwickelt.

2. WAS KANN ZOTERO?

- Zotero findet bibliographische Informationen auf Webseiten
- Speichern der Informationen mit einem Mausklick in einer lokalen Datenbank
- Kopien von Onlinequellen speichern
- Datenbankeinträge mit Tags, Notizen und Verlinkungen versehen
- Einträge in verschiedenen Formaten als Literaturverzeichnis exportieren (z.B. in MS Word)
- verschiedene Metadatenformate (z.B. BibTeX, EndNote und MODS) importieren und exportieren

3. SYSTEMVORAUSSETZUNGEN

Firefox 2.0. oder 3.0, Netscape Navigator 9.0 oder Flock 0.9.1 für Windows, Mac oder Linux

Download von: <http://www.zotero.org/>

4. ZOTERO GRUNDLAGEN

- Zotero-Symbol rechts in der unteren Leiste des Browsers (öffnet Zotero)



• Symbole in der Adresszeile des Browsers:

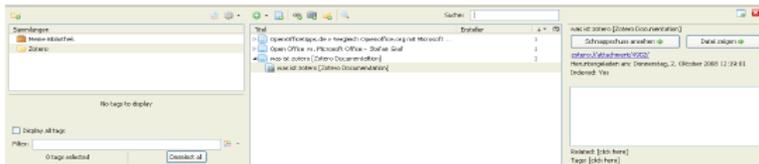
- Buch (z.B. im Bibliothekskatalog)



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



• Die drei Spalten von Zotero



LINKE SPALTE:

„Meine Bibliothek“, alle angelegten Sammlungen.

SYMBOL:

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

MITTLERE SPALTE:

Alle Einträge der links markierten Sammlung.

SYMBOL:

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

RECHTE SPALTE:

Informationen über den in der mittleren Spalte ausgewählten Eintrag.

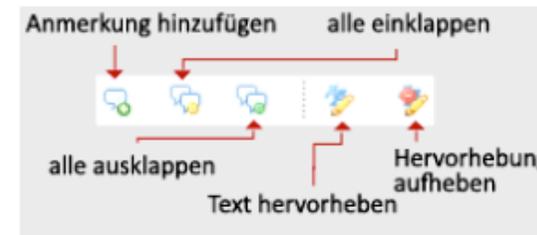
SYMBOL:

- Vollbildmodus
- Zotero schließen

• Einträge verknüpfen (Rechte Spalte)

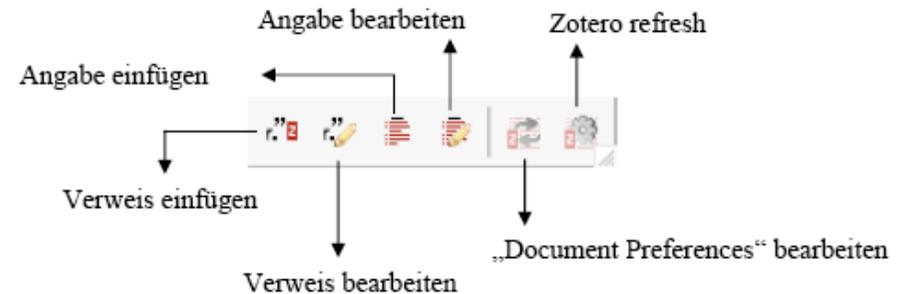
- *Infos*: Zitationsinformationen (können geändert werden)
- *Notizen*
- *Anhänge*: z.B. Dateien, PDFs, Bilder, Links und Schnappschüsse von Webseiten
- *Tags*: Schlagwörter (werden auch automatisch aus Bibliothekskatalogen übernommen)
- *Zugehörig*: andere, zugehörige Einträge

• Gespeicherte Website bearbeiten



• Einträge exportieren

- als HTML-Dateien, Rich-Text-Dateien
- in Bibliographieprogramme (z.B. EndNote, Refworks)
- direkt in MS Word oder Open Office (Plugins notwendig):



- Format für Literaturverzeichnis auswählen: *Document Preferences*
- weitere Zitationsstile können hinzugefügt werden

3. Textverarbeitung

3.1 Programme und Optionen

Für eine Textverarbeitung, die wissenschaftlichen Ansprüchen genügt, stehen Ihnen eine Reihe von Programmen zur Verfügung. Hier einige Optionen:

3.1.1 LaTeX

LaTeX wird oft in mathematisch-naturwissenschaftlichen Publikationen und in der Informatik verwendet, da sich damit insbesondere Formeln gut darstellen lassen. Es handelt sich um eine Makro-Sprache für das Textsatzprogramm TeX. Es trennt das Schreiben von Texten von der tatsächlichen Darstellung, die durch globale Befehle gesteuert werden kann.

Die Textdarstellung auf dem Bildschirm und der Ausdruck sieht daher ziemlich verschieden aus – LaTeX verfolgt nicht das Prinzip WYSIWIG, “What you see is what you get”, d.h. was man auf dem Bildschirm sieht, das erscheint auch im Ausdruck. Ein Beispiel, das einen Quelltext und einen Ausdruck gegenüberstellt:

(1)

```
## 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

LaTeX kann man mit einfachen Texteditoren produzieren, es gibt hierfür jedoch Entwicklungsumgebungen, die mehr Kontrollmöglichkeiten erlauben. Der Wikipedia-Eintrag hält Hinweise auf freie Programme für alle Betriebssysteme bereit:

<http://de.wikipedia.org/wiki/LaTeX>

3.1.2 Open Office

Dies ist ein freies Software-Paket, das ebenfalls für verschiedene Betriebssysteme (Linux, OSX und Windows) zur Verfügung steht. Es enthält neben dem Textverarbeitungsprogramm auch ein Grafikprogramm, ein Modul zur Formatierung von Formeln, eine Tabellenkalkulation, eine Präsentations-Software und eine Datenbank.

Da es nichts kostet, mittlerweile recht stabil läuft, weitgehend mit Microsoft Word kompatibel ist und sehr viele Funktionen von MS Word besitzt, ist es insgesamt sehr zu empfehlen. Im Oktober 2008 ist die letzte Version 3 veröffentlicht worden.

Hinweise und Download: Siehe Wikipedia-Artikel, <http://de.wikipedia.org/wiki/OpenOffice>

3.1.3 Microsoft Word

Dieses Programm ist zum Standard in der Textverarbeitung geworden. Es ist in den Versionen seit 2004 ein ausgezeichnetes Programm mit sehr vielen Funktionen und auch gut für das Schreiben wissenschaftlicher Texte geeignet.

Da die meisten Kursteilnehmer mit MS Word arbeiten, konzentriert sich die Darstellung hier auch auf dieses Programm. Es auch sinnvoll, gut mit der Standard-Textverarbeitung vertraut zu sein. Auf der anderen Seite empfehlen wir aber auch nachhaltig die Verwendung des kostenlosen Open Office oder des ebenfalls kostenlosen LaTeX.

Weitere Hinweise zur Verwendung von MS Word in der Linguistik siehe <http://www.linguistics.ucsb.edu/faculty/cumming/WordForLinguists/Index.htm>

3.2 Arbeiten mit MS Word: Grundsätzliches

Sie können Ihre Arbeit ganz wesentlich erleichtern, wenn Sie sich mit den Möglichkeiten von MS Word (oder der Textverarbeitung, die sie verwenden) gut vertraut machen. Es lohnt sich durchaus, ein paar Stunden oder auch Tage sich damit zu beschäftigen und eine gute, funktionierende und vertraute Arbeitsumgebung einzurichten.

3.2.1 Grundeinstellungen

Die Grundeinstellungen können in WORD → EINSTELLUNGEN...verändert werden. Sie sollten unbedingt

- Unter “Allgemein” die “Liste zuletzt geöffneten Dateien” auf den Maximalwert 9 einstellen (der ist leider immer noch ziemlich niedrig).
- Unter “Allgemein” die Maßeinheiten am besten auf Zentimeter festlegen.
- Unter “Rechtschreibung” sollte man “Rechtschreibung während der Eingabe überprüfen” deselektieren, das stört während des Schreibens nur. Ebenfalls “Grammatik während der Eingabe überprüfen”. Man kann die Rechtschreibung später separat überprüfen.

Gehen Sie auch die anderen Einstellungsoptionen durch und versuchen Sie, diese zu verstehen.

Weitere wichtige Grundeinstellungen können unter EXTRAS → ANPASSEN vorgenommen werden. Dort haben Sie die Möglichkeit, die sichtbaren Symbolleisten am Rand und die Tastatur anzupassen.

3.2.2 Dokumentvorlagen

Wenn Sie Word aus dem Paket heraus frisch starten, finden Sie unter DATEI → PROJEKT-KATALOG eine Reihe von Dokumentformaten. Die von Microsoft gelieferten sind nicht zu gebrauchen; am besten gehen Sie von der Betriebssystem-Oberfläche in den entsprechenden Ordner und löschen sie diese.

Als erstes legen Sie ein neues Dokumentformat an:

- Gehen Sie auf DATEI → NEUES LEERES DOKUMENT. Ein neues Dokument öffnet sich.
- Gehen Sie auf DATEI → SEITE EINRICHTEN.
- Stellen Sie sicher, dass das Format hochgestellt ist, nicht quergestellt (falls Sie das wünschen).
- Stellen Sie in dem Untermenü SEITENEINSTELLUNGEN als Papierformat DIN A4 ein.
- Stellen Sie in dem Untermenü MICROSOFT WORD die SEITENRÄNDER ein. Zu empfehlen ist zunächst ein Abstand oben, unten, links und rechts von 2 cm, ein Bundsteg von 0 cm und ein Abstand von Kopfzeile und Fußzeile von 2 cm.
- Sie können nun gleich Kopfzeile und Fußzeile festlegen. Die minimale Version ist, dass Sie die Kopfzeile freilassen und in der Fußzeile automatisch die Seitenzahl drucken. Hierzu gehen Sie auf ANSICHT/KOPF- UND FUßZEILE, gehen dann auf die Fußzeile am Ende des Textes, zentrieren den Text (Sie drücken hierzu die entsprechende Schaltfläche) und fügen das Symbol für die Seitenzahl ein. Dieses findet sich auf einer speziellen Schaltfläche unter dem Symbol “#”.
- Speichern Sie das Dokumentformat: DATEI → SPEICHERN UNTER..., bei FORMAT:wählen Sie DOKUMENTVORLAGE. Sie werden wahrscheinlich aufgefordert, die Datei in dem Ordner EIGENE VORLAGEN zu speichern. Es ist besser, sie in dem darüberliegenden Ordner VORLAGEN zu speichern, da kann man direkter auf sie zugreifen. Geben Sie dieser Vorlage einen entsprechenden Namen, z.B. NORMALDOKUMENT.

Der besondere Tipp: Die Dokumentvorlagen werden alphabetisch sortiert dargeboten. Wenn Sie wollen, dass sie später als erste Option auftritt, empfiehlt es sich, den Namen mit einem Leerzeichen beginnen zu lassen.

Sie können jetzt das Dokument schließen – es lautet “Normaldokument.dot”, das “dot” steht für Formatvorlagen. Gehen Sie dann auf “Datei/Projektkatalog” und wählen Sie “Normalvorlage” aus. Sie erhalten eine leere Datei, die nach Ihren Vorgaben vordefiniert ist.

3.2.3 Formatvorlagen (“Style Sheets”)

Es ist ganz wesentlich für das strukturierte und effiziente Arbeiten, dass Sie Texte nicht einfach eintippen und bei Überschriften, Fußnoten, Beispielsätze usw. von Fall zu Fall die Größe oder die Art der Schrift, den Einzug, den Paragraphenabstand usw. eingeben. Das kostet Zeit, führt zu Inkonsistenzen, und zieht viel Mehrarbeit nach sich, wenn man das Format verändern will. Sie müssen mit **Formatvorlagen** arbeiten.

Die Formatvorlage “Standard” oder “Normal”

Sie haben eine Datei im Dokumentformat “Normalvorlage” geöffnet. Sie beginnen damit, das Grundformat zu definieren.

- Wählen Sie FORMAT → FORMATVORLAGE. Wahrscheinlich können Sie aus dem Menü eine Formatvorlage “Standard” oder “Normal” auswählen; wenn nicht, dann definieren Sie diese Formatvorlage neu.

Ich habe für “Standard” die folgenden Werte definiert: Schriftart Times New Roman, 12 pt, Deutsch, Links, Zeilenabstand: genau 12 pt, Abstand Vor: 6 pt, Nach: 6 pt, Absatzkontrolle. Diese Werte kann man definieren über die Optionen des Untermenüs ÄNDERN → FORMAT. Dort kann man die Schriftart, die Schriftgröße, die Sprache, die Ausrichtung (linksbündig, justiert) usw. festlegen. Die linksbündige Ausrichtung (Flattersatz) ist meist besser lesbar als die justierte. Als Zeilenabstand stelle ich einen genauen Wert ein anstelle des Werts

“einfach”, weil bei diesem Wert der Zeilenabstand bei der Verwendung von Sonderzeichen aus dem Symbol-Zeichensatz größer wird. Als Abstand vor und nach jedem Absatz habe ich 6pt gewählt, das führt insgesamt bei einer 12 pt-Zeile zu einer Zeile Zwischenraum. Zusätzlich oder alternativ dazu kann man auch eingezogene Zeilen einstellen. Die Einstellung “Absatzkontrolle” heißt, dass isolierte Einzelzeilen am Seitenkopf oder –fuß vermieden werden. Zusätzlich sollten Sie die Tabstopps markieren. Ich habe sie so eingestellt, dass alle 0,5 cm ein Tabstopp erfolgt, also bei 0,5 cm, bei 1 cm, bei 1,5 cm usw. Wenn Sie aus dem Untermenü herausgehen, nicht vergessen, bei der Option “Zur Dokumentvorlage hinzufügen” ein Häkchen zu setzen, bei der Option “Automatisch aktualisieren” aber kein Häkchen zu setzen.

Weitere Formatvorlagen

Von diesem Grundformat ausgehend können Sie nun weitere Formatvorlagen definieren. Bei der Definition von neuen Formatvorlagen ist es wichtig, dass Sie jeweils angeben, auf welcher Formatvorlage sie beruhen soll; dies ist typischerweise “Normalvorlage”. Ich halte die folgenden Formatvorlagen für nützlich:

- “Titel”, für den Titel eines Papiers. Zentriert, größere Schrift (z.B. 18 pt), fett, größerer Abstand (z.B. 24 pt). Nachfolgende Formatvorlage: Autor.
- “Autor”, für den Autor eines Papiers. Zentriert, größere Schrift (z.B. 14 pt), fett, größerer Abstand (z.B. 18 pt). Nachfolgende Formatvorlage: Standard.
- “Liste”, für Aufzählungen wie die vorliegende. Unter “Format”/“Numerierung” ein “Aufzählungszeichen” aussuchen, das am Anfang der Liste stehen soll; einen geeigneten Abstand vorher und nachher einstellen (hier beträgt dieser 2 pt). Nachfolgende Formatvorlage: Liste.
- “Quotation”, für Zitate; diese sind definiert durch einen linken und einen rechten Einzug von jeweils 1 cm.
- “Referenz”, für bibliographische Referenzen; hierfür ist der Absatz mit einer “hängenden Zeile” definiert, wobei die Folgezeilen um 0,5 cm nach rechts gerückt werden.
- “Beispiel”, für ausgeworfene (und später zu numerierende) Beispiele. Diese habe ich ebenfalls als Absätze mit hängenden Zeilen mit einem Einzug von 1 cm definiert, wobei der erste Tabstopp bei 0,5 cm gelöscht wird.
- “Überschrift unterstrichen”, für unterstrichene, nicht numerierte Überschriften. Zeichenformat unterstrichen, Zeilen nicht trennen (d.h. die Überschrift befindet sich immer auf derselben Seite), Absätze nicht trennen, d.h. der folgende Absatz beginnt immer auf derselben Seite wie die Überschrift.

Sie können natürlich noch weitere Formate definieren.

3.2.4 Formatvorlagen für Überschriften

Es gibt verschiedene vordefinierten Formatvorlagen-Namen. Besonders wichtig sind hierfür die Formatvorlagen “Überschrift 1” bis “Überschrift 9”, die insgesamt bis zu 9 Überschriften-Ebenen erlauben. Realistischerweise braucht man selbst für Habilitationsschriften nicht mehr als vier. Diese ersten vier sollte man definieren. Ziel ist es, dass die Überschriften automatisch nach dem Dezimalsystem durchnumeriert werden, wie dies auch in dem vorliegenden Text geschieht.

Gehen Sie auf FORMAT → FORMATVORLAGE..., öffnen Sie die die Formatvorlage “Überschrift 1”, gehen Sie auf “Ändern”, und stellen Sie sicher, dass die Formatvorlage auf “Standard” bzw. “Normal” beruht und ferner die nachfolgende Formatvorlage ebenfalls “Standard” bzw. “Normal” ist. Dann ändern Sie Schriftgröße (Vorschlag: 14 pt) und Schriftart (fett). Bei “Absatz” stellen Sie sicher, dass “Zeilen nicht trennen” und “Absatz-

kontrolle" ausgewählt sind, und dass vor dem Absatz 18 pt und danach 6 pt Abstand zum nächsten Absatz liegen sollen. Ferner soll bei der Option "Hängend" der Wert 0,5 cm gewählt werden.

Nun wird die automatische Numerierung eingerichtet. Es gibt hier verschiedene Möglichkeiten; hier wird die Dezimalnumerierung besprochen. Gehen Sie bei Auswahl von "Ändern" auf "Numerierung...", dann auf "Gliederung" und wählen Sie die Schaltfläche aus, die die entsprechende Dezimalnumerierung zeigt (also erste Zeile 1., zweite Zeile 1.1, dritte Zeile 1.1.1). Beim Verlassen des Menüs nicht vergessen, "Zur Dokumentvorlage hinzufügen" auszuwählen.

Die Formatvorlage "Überschrift 2" sollte auf "Überschrift 1" beruhen, allerdings sollte die Schrift nun nicht fett sein, der Abstand zum vorhergehenden Absatz sollte nur 12 pt betragen, und bei der Option "hängend" den Wert 1 cm wählen.

Formatvorlage "Überschrift 3" basiert auf "Überschrift 2", die Schrift ist nun aber wieder fett, dafür nur 12 pt groß. Formatvorlage "Überschrift 4" basiert auf "Überschrift 3", nun ist aber die Schrift kursiv.

3.3 Automatische Nummerierung und Querverweise

3.3.1 Automatische Numerierungen

Wir haben die automatische Numerierung bei Überschriften kennengelernt. In wissenschaftlichen Werken gibt es häufig weitere Dinge, die nummeriert werden sollen, vor allem, weil man im Text darauf Bezug nehmen will – Beispiel: Abbildungen, Tabellen, Formeln, oder – in der Linguistik – Sprachbeispiele.

Das Nummerieren bei Hand hat den Nachteil, dass nachträgliche Veränderungen – wie z.B. Erweiterungen, Verkürzungen, Umstellungen – die Numerierung durcheinanderbringen. Das kann man vermeiden, indem man die automatische Numerierung und automatische Verweise auf nummerierte Elemente verwendet.

Word bietet hierfür drei Standardformate an: Abbildung, Tabelle und Gleichung. Sie können unter EINFÜGUNG → BESCHRIFTUNG aufgerufen werden. Die Auswahl fügt Zeilen der folgenden Art in den Text ein:

Abbildung 1

Gleichung 1

Tabelle 1

Die Zahl wird hierbei automatisch erzeugt. Wenn Sie z.B. noch einmal EINFÜGUNG → BESCHRIFTUNG → ABBILDUNG auswählen, wird die nächsthöhere Zahl eingefügt:

Abbildung 2

3.3.2 Querverweise

Sie können auf diese automatisch erzeugten Nummern mit Querverweisen referieren. Gehen Sie hierzu auf EINFÜGUNG → QUERVERWEIS, wählen Sie im linken Menü "Abbildung" aus und im rechten "Nur Bezeichnung und Nummer". Gehen Sie in dem Feld "Für welche Beschriftung" auf "Abbildung 1" und drücken Sie die Enter-Taste. Es wird in Ihren Text der Verweis "Abbildung 1" eingefügt.

Diese Einfügung wird automatisch erzeugt, d.h. wenn sich die Nummer des Ziels des Verweises ändert, dann ändert sich der Verweis auch. Dies können Sie sehen, indem Sie in

dem Text **vor** der Stelle, in dem Sie die Beschriftung "Abbildung 1" eingefügt haben, eine weiteres Mal eine Beschriftung dieses Typs einfügen. Markieren Sie dann den gesamten Text durch Drücken der Tastenkombination Befehl+a, und drücken Sie dann die Tastenkombination BEFEHL+WAHL +UMSCHALT+U. Aus der ursprünglichen "Abbildung 1" wird nun "Abbildung 2", und der Querverweis verändert sich entsprechend auf "Abbildung 2".

3.3.3 Ein neues Beschriftungsformat definieren

Sie können neue Beschriftungsformate definieren, z.B. werden Sie in einem englischsprachigen Text statt **Tabelle** sicher **Table** stehen haben wollen. Hier wird die Neudefinition eines Beschriftungsformats gezeigt, indem wir ein neues Beschriftungsformat **Beispiel** für linguistische Beispiele definieren. Dieses soll einfach nummerierte Beispiele der folgenden Art erzeugen:

(1) Dies ist ein automatisch nummeriertes Beispiel.

Für neue Beschriftungsformate gibt es unter "Einfügung/Beschriftung" die Option "Neue Bezeichnung". Wir rufen diese auf und definieren als Namen der neuen Bezeichnung eben "Beispiel". Wir können nun auch diese Kategorie automatisch einfügen und erhalten:

Beispiel 1

Wir wollen jedoch keinesfalls immer das Wort "Beispiel" vor der Nummer stehen haben. Wir können dem abhelfen, indem wir das automatische Beispiel wie folgt einfügen. Wir gehen auf "Einfügen/Feld...", und wählen unter der Kategorie "Numerierung" die Option "Seq" aus. In die Zeile des Menüs, die nun mit "Seq" beschriftet ist, schreiben wir "Beispiel" und drücken die Eingabe- oder Enter-Taste. Wir sehen jetzt nur die entsprechende Nummer des Beschriftungsformats:

2

Wir können uns auf diese neue Beschriftungen durch Querverweise beziehen, wovon Sie sich gleich überzeugen können (durch EINFÜGUNG → QUERVERWEIS...).

3.4 Ein Makro für die neue Beschriftungskategorie

Wir wollen nun häufig folgende Aktionen nacheinander ausführen:

- Wechsel zu der Formatvorlage "Beispiel"
- Einfügen einer öffnenden Klammer "("
- Einfügen einer automatisch erzeugten Beispielnnummer
- Einfügen einer schließenden Klammer ")"

Einfügen eines Tabstopps Für solche Fälle sind sog. **Makros** gedacht. Wir werden nun ein solches Makro definieren und unter einer bestimmten Tastenkombination – ich schlage hier Kontroll+B vor – ablegen.

Gehen Sie hierzu auf EXTRAS → MAKRO → AUFZEICHNEN... Geben Sie dem Makro den Namen "Beispiel" und klicken Sie auf der Schaltfläche "Tastaturbelegung". Drücken Sie dann die Tasten "Kontroll+B" und dann auf "OK". Jetzt erscheint links oben eine kleine Schaltfläche. Geben Sie die vier oben angeführten Aktionen in der angegebenen Reihenfolge an und drücken Sie am Ende auf das Quadrat der Schaltfläche. Das Makro ist jetzt gespeichert. Wenn immer Sie nun Kontroll+B drücken, werden die oben angeführte Aktionen ausgeführt, und sie können sofort danach Ihr linguistisches Beispiel einfügen.

Es lohnt sich, jetzt einmal aus Word ganz herauszugehen und das Programm dann neu zu starten, denn nur dann werden Makros dauerhaft übernommen.

4. Konventionen wissenschaftlichen Schreibens

Es gibt eine Reihe von Konventionen wissenschaftlichen Schreibens, die sich bewährt haben und die man einhalten muss. Einige wichtige Regeln für linguistische Texte:

- Klarer, aussagekräftiger Titel
- Häufig: Abstract (Zusammenfassung) des Textes am Anfang.
- Der Text ist mit Kapitelüberschriften zu gliedern, aber nicht zu feinteilig. Die Kapitel sind zu numerieren.
- Wenn ein Abschnitt gegliederte (numerierte) Unterabschnitte hat, dann sollte nach der Überschrift sofort (oder nach einem kurzen einleitenden Text) die Überschrift des ersten Unterabschnitts stehen. Numerierte Abschnitte mit nur einem Unterabschnitt sind zu vermeiden.
- Die Referenz auf Objektsprache im Text erfolgt mit *Kursivschrift* (ohne Anführungszeichen). Kursivschrift wird für keinen anderen Zweck verwendet. Bedeutungen werden in einfachen Anführungszeichen wiedergegeben, phonetische Umschrift in eckigen Klammern, Phoneme in Schrägstrichen, Grapheme in spitzen Klammern. Beispiel: Das englische Wort *luck* [lʌk] ‘Glück’ zeigt den Verlust eines initialen /g/
- Längere objektsprachliche Beispiele werden in eigenen Abschnitten numeriert dargestellt. Wir werden die Behandlung fremdsprachlicher Beispiele (Interlinearübersetzung, idiomatische Übersetzung) später ausführlich behandeln.
- Auf Beispiele und andere numerierte Elemente kann man im Text durch die Nummern bezugnehmen: “Wie Beispiel (27.a) zeigt, ...”
- Zitate werden in doppelten Anführungszeichen wiedergegeben oder in einem eigenen Abschnitt, mit Einrückung, dargestellt.
- Endnoten sind zu vermeiden, Fußnoten sind sparsam einzusetzen.
- Der Bezug auf Sekundärliteratur sollte im Text selbst geschehen, nicht in Fußnoten. Format: “Nach Partee (1987) besitzen NPn verschiedene semantische Typen, die durch Typenverschiebung ineinander überführt werden können.” Dabei sind Seitenangaben möglich: “Nach Partee (1987: 324ff.)...”
- Bei Doppel- oder Mehrfachautorschaft kann man das &-Zeichen einsetzen: “Wie Gundel, Hedberg & Zacharski (1993) zeigten, ...”. Bei mehr als drei (manchmal mehr als zwei) Autornamen schreibt man “e.a.” für “et alii”, z.B. “Wie Gundel e.a. (1993) zeigten, ...”
- Bei längeren Werken (ab der Bachelorarbeit) ist ein Inhaltsverzeichnis nötig. Es kann automatisch erzeugt werden (mehr dazu später).
- Alle verwendete Literatur muss am Ende in einer Bibliographie zusammengefasst werden. Hierfür gibt es viele verschiedene Formatvorlagen, die zwischen verschiedenen Publikationsformen unterscheiden.

Auf der Moodle-Seite finden Sie das “Unified Style Sheet”, das von Herausgebern linguistischer Zeitschriften erarbeitet wurde und an das man sich halten kann.

Hier einige Beispiele für Zitationsweisen aus diesem Stylesheet.

Example references

- Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.
- Casali, Roderic F. 1998. Predicting ATR activity. *Chicago Linguistic Society (CLS)* 34(1). 55-68.
- Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.
- Coetsem, Frans van. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.
- Franks, Steven. 2005. Bulgarian clitics are positioned in the syntax. http://www.cogs.indiana.edu/people/homepages/franks/Bg_clitics_remark_dense.pdf(17 May, 2006.)
- Iverson, Gregory K. 1983. Korean /s/. *Journal of Phonetics* 11. 191-200.
- Iverson, Gregory K. 1989. On the category supralaryngeal. *Phonology* 6. 285-303.
- 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.
- Murray, Robert W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59(3). 514-528.
- Oxford English Dictionary*, 2nd edn. 1989. Oxford: Oxford University Press.
- Pedersen, Johan. 2005. The Spanish impersonal *se*-construction: Constructional variation and change. *Constructions* 1, <http://www.constructions-online.de>. (3 April, 2007.)
- Rissanen, Matti. 1999. Syntax. In Roger Lass (ed.), *Cambridge History of the English Language*, vol. 3, 187-331. Cambridge & New York: Cambridge University Press.
- Stewart, Thomas W., Jr. 2000. *Mutation as morphology: Bases, stems, and shapes in Scottish Gaelic*. Columbus, OH: The Ohio State University dissertation.
- Webelhuth, Gert (ed.). 1995. *Government and binding theory and the minimalist program: Principles and parameters in syntactic theory*. Oxford: Blackwell.
- Yu, Alan C. L. 2003. *The morphology and phonology of infixation*. Berkeley, CA: University of California dissertation.

5. Weiteres zu Textverarbeitung

5.1 Sonderzeichen: Grundsätzliches

In der Sprachwissenschaft verwendet man verschiedene Typen von Sonderzeichen, z.B. phonetische, logisch-mathematische Zeichen, griechische Buchstaben, lateinische Zeichen mit diversen Akzenten und Erweiterungen und um die Schriften anderer Sprachen.

Gute Informationsquelle: <http://tlt.its.psu.edu/suggestions/international/index.html>

5.1.1 Phonetische Zeichen

Im wesentlichen die Zeichen des Internationalen Phonetischen Alphabets. Siehe...

- Anhang 1 für die offizielle Tabelle,
- http://de.wikipedia.org/wiki/Internationales_Phonetisches_Alphabet mit detaillierten Informationen und der Möglichkeit, Zeichen herauszukopieren
- <http://www.arts.gla.ac.uk/IPA/ipachart.html> für die Beschreibung des Phonetischen Alphabets durch die International Phonetic Association.

5.1.2 Logisch-Mathematische Zeichen

wichtigsten gebräuchliche Zeichen:

(2)

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

5.1.3 Fremdsprachliche Schriften

Heute können zahlreiche fremdsprachliche Schriften geschrieben werden, auch 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

5.2 Kodierung von Sonderzeichen

Die Eingabe von Sonderzeichen (und deren Ausdruck) ist nicht immer einfach, die Dinge haben sich aber in den letzten Jahren entscheidend verbessert.

Der wesentliche Schritt gelang mit der Einführung des Kodierungsformats Unicode. Dies ist ein international vereinbarter Standard, der umfassend genug ist, dass darin alle jemals

verwendeten Zeichensysteme Platz finden (auch z.B. das Schriftsystem des Chinesischen). Für weiterführende Informationen siehe <http://de.wikipedia.org/wiki/Unicode>.

Der Unicode-Standard UTF8 ist am weitesten verbreitet; er wird von modernen Textverarbeitungsprogrammen, Webbrowsern, E-mail-Programmen usw. unterstützt. Er kodiert Zeichen variabel (zwischen 1 Byte und 4 Byte pro Zeichen) und geht daher sparsam mit Ressourcen um.

Der Unicode-Zeichensatz enthält insbesondere alle Phonetik-Zeichen und diakritische Symbole. Dies gilt allerdings nur für wenige Schriftfamilien, wie **Lucida Grande**, **Arial Unicode** und insbesondere **Doulos SIL** (ein spezialisierter Zeichensatz der Times-Familie). Doulos SIL kann von der folgenden Webseite heruntergeladen und installiert werden: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=DoulosSILfont

5.3 Sonderzeichen in MS Word und deren Eingabe

5.3.1 Allgemeines

MS Word (ab Version 2004) erlaubt in größerem Umfang die Verwendung von Unicode-Sonderzeichen. Frühere Versionen hatten Spezialzeichensätze wie *Symbol* (auch in modernen Versionen verfügbar).

5.3.2 Eingabe über Menüs innerhalb von MS Word

Man geht auf EINFÜGEN → SONDERZEICHEN (oder in Windows: SYMBOL). Hier kann man eine Reihe von Schriftarten auswählen. Für unsere Zwecke sind interessant:

- Arial Unicode: Ein umfangreicher Unicode-Zeichensatz
- Symbol: Ein Zeichensatz mit nicht-akzentuierten griechischen Zeichen und mathematischen Symbolen:

!	∇	#	∃	%	&	∞	(*	+	,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;		
<	=	>	?	≡	A	B	X	Δ	E	Φ	Γ	H	I	Θ	K	Λ	M	N	O	Π	Θ	P	Σ	T	Y	ς	Ω
Ξ	Ψ	Z	[:.]	⊥	_	˘	α	β	χ	δ	ε	φ	γ	η	ι	κ	λ	μ	ν	ο	π	θ	ρ	σ	
τ	υ	ω	ξ	ψ	ζ	{		}	~	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	
⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	
←	↑	→	↓	°	±	″	≥	x	∞	∂	•	+	≠	≡	=	...		—	⌞	⌟	⌠	⌡	⌢	⌣	⌤	⌥	
⌦	⌧	⌨	〈	〉	⌫	⌬	⌭	⌮	⌯	⌰	⌱	⌲	⌳	⌴	⌵	⌶	⌷	⌸	⌹	⌺	⌻	⌼	⌽	⌾	⌿	⌿	
™	Σ	∏	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	∫	

- Wingdings: Ein (oder mehrere) Zeichensätze mit vielen Spezialsymbolen, z.B. etc.

5.3.3 Eingabe über Tastenbelegungen

Man kann die Tasten über die Tabelle auswählen. Besser ist es – zumindest bei häufig belegten Tasten – dafür **Tastaturbelegungen** zu wählen. Dies ist eine Option, die bei EINFÜGEN → SONDERZEICHEN angezeigt wird.

Man kann als Tastenbelegung eine Taste angeben (typischerweise eine Kombination mit Umschalttaste, z.B. Ctrl+A), oder zwei (eine Kombination mit Umschalttaste gefolgt von einer einfachen Taste oder einer Kombination mit Umschalttaste). Am besten die letztere Option wählen und dabei mnemotechnisch günstige Kombinationen wählen:

- ctrlG für Griechisch, z.B. ctrlG+a für α, ctrlG+b für β usw.
- ctrlM für Mathematisch, z.B. ctrlM+v für ∨ (Disjunktion), ctrlM+k für ∧ (Konjunktion)

5.3.4 Eingabe über die Autokorrektur-Funktion

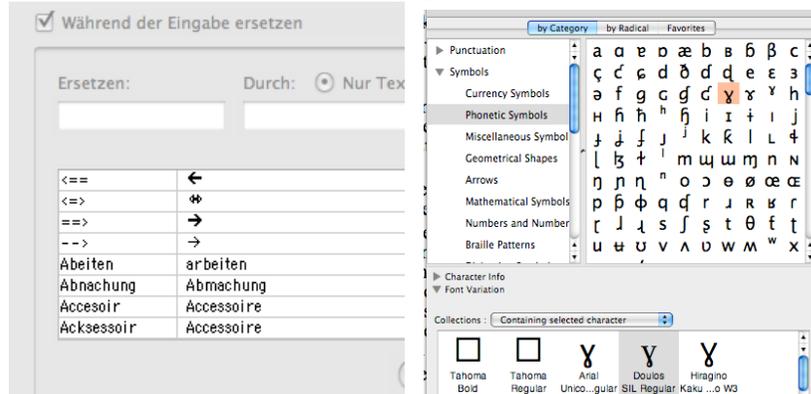
MS Word hat, wie andere Textverarbeitungsprogramme, eine **Autokorrektur**-Funktion. Wenn diese Option eingeschaltet ist (EXTRAS → AUTOKORREKTUR → WÄHREND DER EINGABE ERSETZEN) werden Zeichenketten durch andere Zeichenketten ersetzt. Dies kann für verschiedene Zwecke verwendet werden (siehe Abbildung unten).

- Für die Selbstkorrektur von häufig falsch geschriebenen Wörtern, die ursprüngliche Intention. Gehen Sie auf EXTRAS → AUTOKORREKTUR..., dann sehen Sie eine lange Liste von möglichen Schreibfehlern und deren Ersetzungen. Diese Liste können Sie selbst erweitern.
- Für andere Fälle des Ersatzes einer Zeichenkette durch eine andere. Z.B. können Sie die Kette *mfg* durch *Mit freundlichen Grüßen* ausbuchstabieren lassen.
- Hier relevant: Für die Ersetzung einer Beschreibung durch ein Zeichen. Am besten definieren Sie ein Escape-Symbol, z.B. den Schrägstrich, auf den dann eine Beschreibung des Zeichen folgt. Wenn Sie diese Zeichenfolge später schreiben, wird die Zeichenfolge durch das Symbol ersetzt. Sie können z.B. */alpha* durch den griechischen Buchstaben α ersetzen lassen, */teilmenge* durch ⊆ usw. Im Vergleich zu der Tastenbelegung bedeutet das meist etwas mehr Tippaufwand, aber dafür kann man sich die Sache besser merken. Nebenbemerkung: Mithilfe des Autokorrektur-Menüs kann man auch auf die Option AUTOTEXT kommen. Sie erlaubt es, auch größere Textabschnitte einzufügen, z.B. auch Tabellen von bestimmtem Format.

Wichtig: Wenn man die Sonderzeichenangabe so einrichtet, dann wird dies auch von der Präsentationssoftware PowerPoint übernommen – eine große Erleichterung beim Schreiben von Sonderzeichen bei Präsentationen.

5.3.5 Eingabe über Menüs außerhalb von MS Word

Für das OSX-Betriebssystem von Macintosh gibt es eine Funktion *Character Palette*, mit dem man Zeichen eingeben kann. Man kann Zeichen nach ihrer Unicode-Kategorie angeben, man kann eine eigene Liste von *Favoriten* zusammenstellen. Chinesische Zeichen können auch nach ihrem Radikal sortiert eingegeben werden.



Autokorrektur-Funktion

Einfügen durch Character-Palette

5.3.6 Eingabe von Akzenten

Akzente (diakritische Zeichen) können auf verschiedene Weise eingegeben werden:

- Gängige Akzente können durch Tastenkombinationen eingegeben werden (mit alt/option in OSX, andere Kombinationen mit Windows, siehe Symbol-Menü)

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 ä	opt+v	Ogonek ą	opt+m
Ring å	opt+k	Doppel-Akut ŷ	opt+j
Punkt unten ạ	opt+x	Punkt oben è	opt+w

So können auch weitere sprachspezifische Sonderzeichne 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.

- Bei den obengenannten Akzenten werden definierte Unicode-Zeichen angesprochen. Diese können auch direkt über Character-Palette: Schrift "Latin" eingegeben werden.
- Darüber hinaus können Akzente mit beliebigen Buchstaben kombiniert werden. Tippen Sie den Buchstaben + shift+opt+eine für das diakritische Zeichen typische Taste.

5.3.7 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)
- Unterstreich (underline) für Hervorhebungen – sparsam verwenden.
- **KAPITÄLCHEN (SMALL CAPS)** für Fachausdrücke, Bedeutungen, manchmal Autornamen.
- Weiteres: Doppelt unterstrichen, punktiert unterstrichen, durchstrichen, schattiert

5.3.8 Steuerungszeichen

Mithilfe von EINFÜGEN → SYMBOL (SONDERZEICHEN) können auch eine Reihe von Steuerzeichen eingegeben werden (z.B. Geviertstriche —, Halbgeviertstriche –, geschützte Trennstriche (bei denen nicht getrennt wird), bedingte Trennstriche (bei denen getrennt wird, wenn sie am Wortende stehen) usw. Es gibt hierfür auch Tastenkombinationen

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

5.4 Textmanipulation

5.4.1 Suchen und Ersetzen

Selbstverständlich kann man in allgen gängigen Textverarbeitungsprogrammen suchen und Such-Strings auch durch andere ersetzen. In MS Word geschieht das in BERARBEITEN → SUCHEN / ERSETZEN. Hierbei gibt es eine Reihe von wichtigen Optionen:

- Suchrichtung: Abwärts, Aufwärts, Gesamt, Alle geöffneten Dokumente

- Groß/Kleinschreibung beachten, nach ganzem Wort suchen
- Mustervergleich ist die Bezeichnung für die Suche mit regulären Ausdrücken. Hier können z.B. ? für einzelne Buchstaben, * für Ketten von Buchstaben verwendet werden, und anderes mehr.

Sie können nach Formatierungen suchen und nach Dingen wie Absatzende (^p), Tabulatorpositionen (^t) usw. und diese entsprechend ersetzen.

Suchen mit Mustervergleich -- Beispiele

- Suche "H?nd" findet *Hund, Hand*; ? steht für ein Zeichen
- Suche "au*erdm" findet *außerdem, ausserdem*; * steht für Zeichenketten. Eingrenzen auf ein einzelnes Wort durch @, "(au)[a-z]@(erdm)".
- Eingrenzung der suche durch eckige Klammern, z.B. "[0-9]" Eingrenzung auf Ziffern, "[a,e,i,o,u]" Eingrenzung auf Vokale. Ausschluss von Zeichen: "[!0-9]" findet alle Zeichen, die keine Zifferen sind.
- Suche nach Wortanfang oder Wortende durch ">" oder "<". Beispiel: "<(Muster)" liefert *Mustermann, Musterhaus*, "(muster)>" findet *Farbmuster, Produktmuster*.
- Suche nach aus mehreren Elementen zusammengesetzten Suchtermen gruppiert man durch runde Klammern. Beispiel: Suche nach fünfstelligen Postleitzahlen durch "([0-9][0-9][0-9][0-9][0-9])"
- In geschweiften Klammern "{}" legt man fest, wie oft eine Suchabfrage mindestens oder genau vorkommen soll. Beispiel: Suche nach Postleitzahlen durch "([0-9]{5})"
- Mit "@" erweitert man die Suche auf eine unbestimmte Anzahl von Platzhaltern: "([0-9]@)" findet Zahlen aller Länge.
- Um nach einem als Platzhalter definierten Zeichen zu suchen, gibt man davor einen umgekehrten Schrägstrich (den Backslash "\") ein. Suche nach Fragezeichen: "?\".
- Man kann Such-Variablen in das ERSETZEN-Feld übertragen. Das geschieht mit "\" gefolgt von der Nummer des Ausdrucks im Feld "Suchen".
Beispiel: Voranstellung eines D- zu Postleitzahlen: Suche nach fünfstelligen Postleitzahlen durch "([0-9]{5})", Ersetzen durch "D-\1".
Beispiel: Voranstellung des Währungszeichens, 250 EUR zu EUR 250: Suche nach "([0-9]@) (EUR)", ersetzen durch "\2 \1".

Siehe für komplexe Suchen und Ersetzen: <http://www.akademie.de/office-programme/office-microsoft-word-lernen/kurse/profiwissen-word-praxis-addins/word-schnelle-bearbeitung/powersuche.html>

5.4.2 Ausschneiden und Einsetzen

Im Menü BEARBEITEN können Sie bekanntlich Textstellen ausschneiden und kopieren und an anderer Stelle wieder einsetzen. Diese Art von Textmanipulation können Sie auch direkt durch Markieren und Verschieben durchführen. Wenn Sie mehrere Textabschnitte gleichzeitig im Griff behalten wollen, gibt es ein "Scrapbook" unter EXTRAS → SCRAPBOOK.

5.4.3 Alphabetisch sortieren

Sie können eine Liste von Absätzen (und z.B. Wörtern oder Tabellenzeilen) alpaetisch sortieren: TABELLE → SORTIEREN.

Aufgabe: Nehmen Sie einen beliebigen Text und erzeugen Sie mit dem Suchen-Befehl und der Sortierfunktion eine alphabetisch sortierte Liste der Wörter dieses Textes.

5.4.4 Änderungen nachverfolgen

Vor allem für die Zusammenarbeit mehrerer Autoren an einem Text ist die Funktion EXTRAS → ÄNDERUNGEN NACHVERFOLGEN sinnvoll. Damit kann man einen Text redigieren, wobei alle Veränderungen angezeigt werden und der ursprüngliche Autor den Veränderungen zustimmen oder sie ablehnen kann.

5.4.5 Funktion "Rückgängig machen"

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 rekonstituieren. Zugriff auf diese Funktion am besten über die Symbolleiste.

5.4.6 Automatische Rechtschreibkorrektur

Unter EXTRAS → RECHTSCHREIBUNG UND GRAMMATIK kann man ein Rechtschreibkorrektur-Programm einsetzen. Dabei muss natürlich für den Text die richtige Sprache eingestellt werden (die auch je nach Formatvorlage variieren kann). Man darf sich auf die Rechtschreibkorrektur nicht verlassen, sie stellt aber eine wichtige Unterstützung vor der Abgabe von Arbeiten dar.

Für die Komposition von Texten kann auch die Thesaurus-Funktion manchmal sinnvoll eingesetzt werden.

5.5 Tabellen

Systematische Daten können oft klarer in Tabellenform dargestellt werden. MS Word bietet dafür sehr gute Möglichkeiten mit der Tabellenfunktion. Um diese nützen zu können, müssen sie unter ANSICHT → SYMBOLLEISTE die Symbolleiste "Tabelle und Rahmen" aktiviert haben.

Sie können zunächst eine einfache Tabelle einfügen und diese dann nachbearbeiten.
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>

6. Grafiken, Bilder, Baumdarstellungen, Glossierungskonventionen

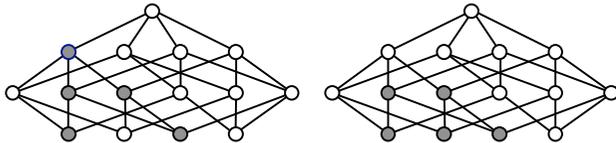
6.1 Grafiken

Für manche Zwecke sind grafische Darstellungen ein wichtiges Kommunikationsmittel. MS Word enthält eine anständige Komponente von Grafik-Werkzeugen, die durch ANSICHT → SYMBOLLEISTEN → ZEICHNEN aktiviert werden kann. Um die Grafik zu sehen, muss man die Ansichtsoption ANSICHT → SEITENLAYOUT einstellen.

Im einzelnen:

- Standardformen, Blockpfeile, Flussdiagramme, Legenden
- Striche und Pfeile
- Textfelder; Ausrichtung des Textes
- Strichstärken, Farben, Hintergründe
- Gruppieren (auch von Teilen einer Grafik – Markieren), am Raster ausrichten, Vordergrund/Hintergrund, Drehen/Kippen
- Verankerung einer Grafik relativ zu einem Paragraphen

Beispiel:



Wenn Sie das Grafikprogramm innerhalb von MS Excel aufrufen, haben Sie als zusätzliche Option “Verbindungen” für Linien zwischen Blöcken. Sie haben dort auch zusätzliche Möglichkeiten, z.B. bei der Ausrichtung von Text in Textfeldern.

6.2 Bilder

Sie können in MS Word auch Bilder (Fotos) integrieren, und zwar 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. Die Dateien sind wesentlich kleiner, manchmal müssen Qualitätsabstriche in Kauf genommen werden..

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.

Es gibt rudimentäre Bearbeitungsmittel in MS Word: Helligkeit, Kontrast, Verwandlung von Farbbildern in Schwarzweiß-Bilder (Graustufe). Sie können das Bild vergrößern (was natürlich die Bildqualität beeinträchtigt) und verkleinern (was ebenfalls die Bildqualität verschlechtern kann).

Das Bild kann auf verschiedene Weise im Text verankert sein. Die einfachste Option ist, das Bild “mit Text in Zeile” zu setzen. Wichtig: Dann müssen Sie die Zeilenhöhe auf “Einfach” setzen; wenn Sie die Zeilenhöhe auf “Genau” gesetzt haben, dann sehen sie nur den unteren Rand des Bildes.

Sie können Bilder und Grafiken mischen, z.B. für Beschriftungen von Teilaspekten von Bildern, für Pfeile usw.

6.3 Baumdarstellungen

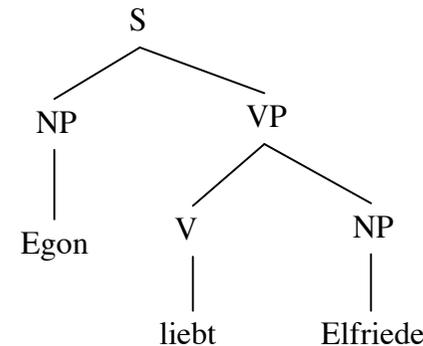
Syntaktische Strukturen werden typischerweise durch Bäume dargestellt. Hierfür gibt es im wesentlichen drei Optionen:

Die Verwendung eines spezifischen Zeichensatzes

Es handelt sich um den Zeichensatz “Arboreal”, <http://www.cascadilla.com/arboreal.html>, zu beziehen für 20\$. Damit kann man Bäume durch Auswahl von Zeichen aus einem Menü erzeugen.

Die Verwendung der Grafik-Option

Mithilfe der Grafikoption “Verbinden” (verfügbar in Excel) kann man auf flexible Weise linguistische Bäume konstruieren.



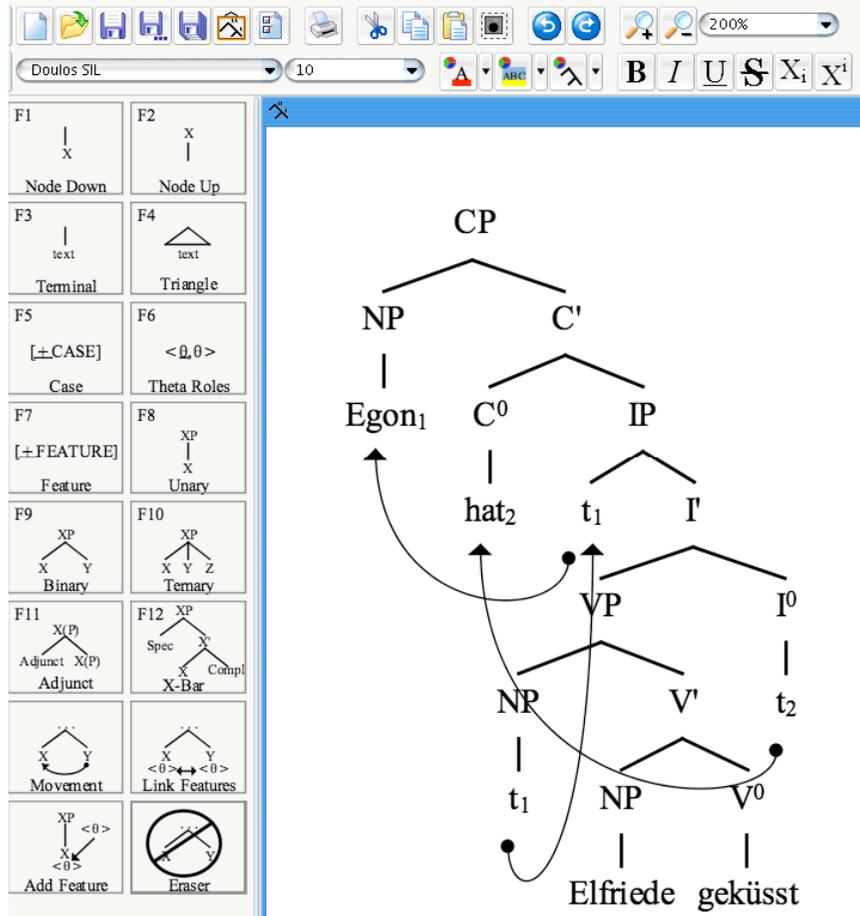
Achtung: Wenn man diese in MS Word kopiert, geht leider gerade die “Verbindungs”-Option verloren. Man sollte am besten ein Arbeitsblatt in Excel haben, in dem man alle Bäume speichert.

Die Verwendung eines speziellen Programmes für Bäume: TreeForm

Es gibt Zusatzprogramme, die Bäume erzeugen können. Beispiel: Der Formeleditor Mathtype (allerdings zum stolzen Preis von \$ 57).

Speziell für Syntaxbäume wurde der Editor TreeForm entwickelt, siehe <http://sourceforge.net/projects/treeform/?abmode=1>. Es handelt sich um OpenSource-Software, die kostenlos erhältlich ist. Es ist ein Java-Programm, das unter Windows, Macintosh OS X und Linux läuft.

Beispiel: TreeForm, Erzeugen eines Syntaxbaums



Man hat in TreeForm einigen Einfluss auf die Gestaltung (Breite) eines Baumes, indem man vor und hinter den Knoten Leerzeichen eingibt. Bei den Linien für syntaktische Bewegung gibt es allerdings keine direkten Einflussmöglichkeiten. Hier kann man z.B. auch mit der Grafikfunktion von MS Word arbeiten.

6.4 Darstellung von linguistischen Glossen

6.4.1 Interlinearübersetzungen: Die Leipzig Glossing Rules

Bei der Darstellung von fremdsprachlichem Material, vor allem aus eher weniger bekannten Sprachen, hat sich das Darstellungsformat die **Interlinearübersetzung** herausgebildet. Sie finden eine gute Darstellung eines weitverbreiteten Formats – der *Leipzig Glossing Rules* – unter <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 idiomatische Übersetzung, in einfachen Anführungszeichen.

Falls es nicht durch den Kontext klar wird, sollte auch der Name der Sprache und die Quelle mit angegeben werden.

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 hamisalug guguna amuq'-da-c
 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.

(5) Turkish

çık-mak
 come.out-INF
 'to come out'

(7) German

unser-n Väter-n
 our-DAT.PL father.PL-DAT.PL
 'to our fathers'

(16) Italian

and-iamo
 go-PRES.1PL
 'we go'

- Darstellungsmöglichkeiten nicht-overter Elemente und von Reduplikationen (Tilde)

(18) Latin

puer
 boy[NOM.SG]
 'boy'

or:

puer-Ø
 boy-NOM.SG
 'boy'

(25) Hebrew

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

- Darstellungsmöglichkeiten für Zirkumfixe und Infixe

(21) German

ge-seh-en
 FTCP-see-FTCP
 'seen'

or:

ge-seh-en
 FTCP-see-CIRC
 'seen'

(23) Latin

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

- Darstellung inhärenter Merkmale wie Genus

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

ož-di-g xōxe m-uq'e-r
 boy-OBL-AD tree(G4) G4-bend-PRET
 'Because of the boy the tree bent.'

(G4 = 4th gender)

7. Schriftliche Arbeiten, Referate, Präsentationen, Plakate

In diesem Abschnitt geht es um die Arbeitstechniken für die Herstellung von verschiedenen Typen eigener schriftlicher Arbeiten. Wir werden später Verfahren kennenlernen, die uns bei der eigentlichen wissenschaftlichen Untersuchung unterstützen können; hier konzentrieren wir uns auf die Tätigkeiten, die unmittelbar mit dem Schreiben zusammenhängen.

7.1 Zur Einleitung

Wozu?

Neben schriftlichen und mündlichen Prüfungen spielen eigenständige Arbeiten eine sehr wichtige Rolle in der Bewertung der Studienleistung. Aber sie sind zugleich viel mehr als das:

- Eine eigenständige Arbeit zwingt dazu, ein Stoffgebiet zum Teil selbständig zu erarbeiten. Man durchdenkt die Daten und ihre möglichen Interpretationen genauer und mit größerer Konsequenz. Man bezieht Stellung zu den Annahmen, die vertreten werden, und beginnt selbst Hypothesen zu entwickeln. Man entwickelt eigene Interessen an bestimmten Datensätzen und wie diese verstanden werden sollen. Man überlegt sich, welche Methoden man einsetzen kann, um Hypothesen zu stützen oder zu widerlegen, und lernt auch, diese Methoden anzuwenden. Auch wenn das Thema der Arbeit vielleicht nicht eines ist, das uns den Rest des Lebens begeistern wird, kann man dabei doch lernen, wie man systematische Forschung zu einem bestimmten Thema angeht.
- Eine eigenständige Arbeit zwingt auch dazu, sich zu überlegen, wie man die Resultate der eigenen Forschungsarbeit am besten anderen Personen mit ähnlichem Hintergrund und ähnlichen Interessen darstellt. Dies ist eine Fähigkeit, die später vielleicht wichtiger sein wird als die vertiefte Kenntnis des Stoffes, den sie sich während der Vorbereitung und des Schreibens angeeignet haben.
- Die Arbeit kann später zur Grundlage von weiterführenden Untersuchungen werden. Zwar ist es sinnvoll, sich nicht zu früh auf bestimmte Themen festzulegen – es ist gut, wenn die Masterarbeit ein deutlich anderes Thema hat als die Bachelorarbeit, und das Thema der Dissertation sollte sich deutlich von dem der Masterarbeit unterscheiden. Aber es ist durchaus möglich, dass man später mit neuen Daten und neuer Perspektive auf alte Themen zurückkommt und die z.B. zu Publikationen ausarbeitet.
- Wenn Sie sich etwa mit einem BA-Abschluss oder einem Master-Abschluss bewerben (für berufliche Stellen oder für weiterführende Studien) kann das Thema der Abschlussarbeit ein wesentliches Argument dafür sein, dass Sie angenommen werden.

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

Worüber?

Worüber sollen Sie in einem Referat, einer Hausarbeit oder einer Abschlussarbeit schreiben? Das hängt zum gewissen Maße natürlich von dem jeweiligen Seminar ab. Im Allgemeinen gilt:

- Das Referat im Seminar stellt typischerweise eine wissenschaftliche Arbeit vor. Um das gut zu machen, müssen Sie aber natürlich den Forschungshintergrund kennen. Sie sollten nicht jeden einzelnen Punkt der Arbeit darstellen, Sie sollten vielmehr raffen, auswählen, zuspitzen, auf den Punkt bringen.
- Die Seminararbeit beruht in der Regel auf dem Referat, oft mit zusätzlichem Material, und stellt dies in einem anderen, schriftlichen, Medium dar. Sie soll zeigen, dass der

Autor, die Autorin die wissenschaftliche Arbeit anderer angemessen und kritisch darstellen kann. Oft enthält sie auch zusätzliches Material im Vergleich zum Referat. Im besten Fall enthält die Seminararbeit auch eigenes empirisches Material oder theoretische Vorschläge.

- Die Abschlussarbeit (z.B. Bachelorarbeit, Masterarbeit) geht typischerweise über die Darstellung 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.
- Bei der Themenwahl einer wissenschaftlichen Arbeit gilt: 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 Slawischen Sprachen*, sondern: *Der Sprechakt des Sich Entschuldignens 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*.

Wie?

Wenn das Thema feststeht, stellt sich die Frage: Wie führe ich die Arbeit durch?

- Fangen Sie frühzeitig mit der Vorbereitung und der Ausführung an. Für die Bachelorarbeit haben Sie zwei Monate Zeit, für die Masterarbeit 5 Monate, für eine Dissertation sollten Sie ca. 2 – 3 Jahre ansetzen. Vor allem bei Bachelor- oder Masterarbeiten sollte man sich schon vorher über die Tragfähigkeit des Themas einigermaßen klar sein.
- Trennen Sie die Phasen des Stoffsammelns, des Organisierens 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. Es lohnt sich z.B., einmal einen wissenschaftsjournalistischen Artikel zu einem Thema zu verfassen. Ich habe auch schon die Erstellung von Forschungsanträgen und von Wikipedia-Artikeln als Aufgabe vergeben.

7.2 Materialsammeln

Eine Abschlussarbeit wird stets Teile enthalten, in dem Sie bestehende Literatur aufarbeiten und auf den speziellen Zweck hin darstellen. Hierzu müssen Sie Literatur systematisch lesen und **das für Ihr Thema Nötige** mit eigenen Worten und auch unter Verwendung von Zitaten herausschreiben und für sich neu organisieren. Typischerweise wird nur ein kleiner Teil dessen, was sie exzerpieren, auch wirklich Verwendung finden.

7.2.1 Karteikästen vs. Computer

Das klassische Instrument hierzu ist der **Karteikasten**. Karteikarten bestehen aus zwei Teilen:

- Einen Thema-Teil, in dem Sie z.B. den Autor und das Werk bezeichnen (wenn Sie publikationsweise experieren) 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) Originalzitaten, freien Inhaltsangaben und Ihren eigenen Bemerkungen klar unterscheiden (z.B. Zitate in Anführungsstrichen, mit Seitenzahlen; eigene Bemerkungen in eckigen Klammern).

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.

7.2.2 Verwendung eines Textverarbeitungsprogramms

Sie können mit Ihrem Textverarbeitungs-Programm in einer oder mehreren Dateien Informationen festhalten.

Es ist sinnvoll, dabei zumindest in einem Punkt dem Karteikarten-System zu folgen: Jede neue Informationseinheit beginnt mit einer neuen Seite (EINFÜGEN → WECHSEL → SEITENWECHSEL), und hat eine Überschrift (oder auch Unterüberschrift). Vorteil: Sie können die Haupttexte ausblenden und sich nur die Überschriften anzeigen lassen.

Beispiel: Volle Anzeige

1→Language-specific observations on information structure

1.1→Russian

1.1.1→Word order, topic and focus: Tracy King, 1995

Topics precede the finite verb. King distinguishes repeated topics, shift topics, and extended topics, following Yokoyama. Good example test on p. 69. There are cases with more than one topic, which suggests adjunction structures:

Det'mi ona ne zanimaetsja

children she NEG occupies "She doesn't care for her children"

Focus is realized in distinct ways. In non-emotive speech items in focus are sentence-final:

Q: *Kto šil?* who sewed?

A: *Šila mne èto plat'e [Inna]* 'INNA sewed me that dress'

Beispiel: Gliederungs-Anzeige (ANSICHT → GLIEDERUNG)

1→Language-specific observations on information structure

1.1→Russian

1.1.1→Word order, topic and focus: Tracy King, 1995

Topics precede the finite verb. King distinguishes repeated topics, shift topics, and extended topics, following Yokoyama. Good example test on p. 69. There are cases with more than one topic, which suggests adjunction structures:

Det'mi ona ne zanimaetsja

children she NEG occupies "She doesn't care for her children"

Focus is realized in distinct ways. In non-emotive speech items in focus are sentence-final:

Beispiel: Kollabierte Anzeige: Ebene 1 und 1.1

1→Language-specific observations on information structure

1.1→Russian

1.2→Germanic

1.2→Hungarian

1.3→Japanese

1.4→Turkish

Beispiel: Kollabierte Anzeige Ebene 1 und 1.1 und 1.1.1

1→Language-specific observations on information structure

1.1→Russian

1.1.1→Word order, topic and focus: Tracy King, 1995

1.1.2→Focus in yes/no questions

1.2→Germanic

1.2.1→Dutch: No scrambling of arguments

1.2.2→Word order and scrambling: Weerman 1989, Neelman 1994

1.2.3→Object shift vs. scrambling: Vikner 1994

Sie können in MS Word 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.

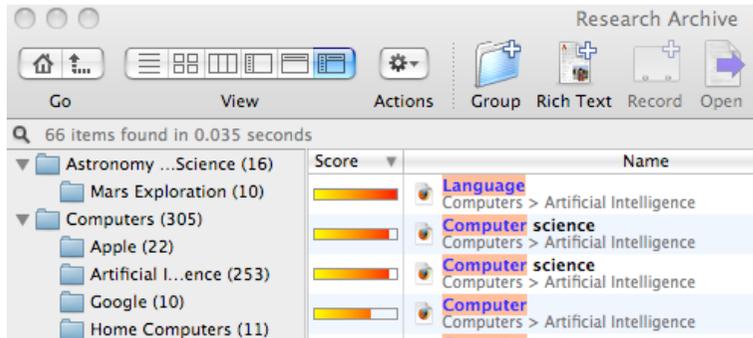
Nachteil: Eingeschränkte Such- und Sortiermöglichkeiten. Sie können nur in geöffneten Dokumenten suchen (aber in allen geöffneten Dokumenten gleichzeitig).

7.2.3 Notetaking-Programme

Eine andere Möglichkeit: Notetaking-Programme. Als Bestandteil von MS Office (für Windows) gibt es z.B. OneNote, und in der Macintosh-Version die ANSICHT → NOTIZBLOCK. Es gibt darüberhinaus zahlreiche weitere Programme (vgl.

http://www.brunnbauer.ch/wissensmanagement/index.php?title=Digitaler_Zettelkasten).

Beispiel eines kommerziellen Produkts: 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 Folder und Unter-Folder angeordnet werden, Beispiel: Devonthink, Suchbegriff "Computer" und "Sprache"



7.3 Formales zu Seminar- und Abschlussarbeiten

Hierzu gibt es ein Formblatt. Einige Punkte:

- Titelblatt mit Seminarangaben, Titel, Name, Adresse, e-mail, Matrikelnummer
- Ab Bachelor-Arbeit: Inhaltsangabe
- Zeilenabstand 1½, justiert (rechter Randausgleich), Silbentrennung

Siehe http://www2.hu-berlin.de/philfak2/lehre/bachelor/ba0405/ba_arbeit.htm und Moodle-Seite.

7.4 Referate

7.4.1 Zweck und Aufbau

Im Gegensatz zu Hausarbeiten ist es ziemlich klar, an wen Sie ein Referat adressieren: An 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 – bei Kurzreferaten sind es vielleicht nur zehn Minuten. Auch Vorträge bei Konferenzen sind oft knapp (manchmal nur 20 Minuten). Wie kann ich diese Zeit optimal nutzen? Bleibt dann auch noch genügend viel Zeit (mindestens 5 Minuten) für die Diskussion übrig?
- Referate sollten nicht abgelesen werden! Dann kann man nämlich gleich den Text lesen, was den Vorteil hat, dass man als Leser den Stoff mit der eigenen Geschwindigkeit verarbeiten kann. 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 alle ihre Vor- und Nachteile; vielleicht brauchen Sie sogar gar keine solchen Hilfen. Handouts (Tischvorlagen) haben

meist klare Vorteile: Die Zuhörer können sie mit Kommentaren für die Diskussion beschreiben und nach Hause mitnehmen, sie zwingen den Zuhörer nicht sklavisch an die jeweilige Stelle der Darstellung, und der Kursleitende kann sie evtl. für Benotung mit heranziehen. Sie sind daher das Medium der Wahl.

7.4.2 Tischvorlagen

Handouts sollen das mündliche Referat unterstützen, nicht ersetzen. Sie sollten also nicht den Text des ganzen Referats enthalten (das ja ohnehin nicht abgelesen werden sollte). Es gibt zwei extreme Formen des Handouts, eine Minimalversion und eine Maximalversion:

- Handouts, die nur das Material zeigen, über das man spricht (also Beispielsätze, Zitate, Bilder, evtl. Tabellen und Grafiken)
- Handouts, die daneben auch die einzelnen Fragestellungen, Behauptungen, Argumentationen darstellen.

Handouts, die dem zweiten Typ zuneigen, sind bessere Abbilder der Referate, und daher insgesamt vorzuziehen. 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”)

Das Handout soll darüber hinaus auch die verwendeten Quellen angeben.

Es ist manchmal schwierig für Zuhörer, vom Handout nicht abgelenkt zu werden und immer zu wissen, worüber gerade die Rede ist. Abhilfe: Der Referierende verweist oft auf Seiten- und Abschnittsnummern und bei Beispielen stets auf die Nummer der Beispiele.

7.4.3 Präsentations-Software

Eine alternative Möglichkeit besteht in der Verwendung von Overhead-Folien und von Präsentations-Software wie MS Powerpoint oder 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 “nach Hause tragen” und 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 (die Stelle der gegenwärtigen Folie) fokussieren.
- 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).

7.5 Powerpoint

Einige Hinweise: Hintergrund möglichst weiß, serifenlose Schrift, Zeilenfall sinnvoll einsetzen, Zusammengehörendes möglichst auf eine einzige Folie. Die Möglichkeit des Mediums nutzen, z.B. dynamische Erzeugung von Schaubildern. Es ist sinnvoll, auch einen Ausdruck mitgeben (4 Folien auf eine DIN A 4-Seite).

7.6 Plakate

8. Linguistische Korpora

8.1 Gegenstand der Korpuslinguistik

Bei der Korpuslinguistik handelt es sich um eine Methode, mit der linguistische Fragestellungen anhand von Sprachdaten empirisch untersucht werden können. Die Korpuslinguistik beschäftigt sich dabei mit dem Aufbau, der Auszeichnung und der Auswertung von Korpora.

Zentrale Fragestellungen

- Welche linguistischen Fragestellungen können anhand von (größeren) Textmengen behandelt werden?
- Nach welchen Kriterien können diese Texte zusammengestellt werden?
- Wie können sie nutzbar gemacht und annotiert (mit linguistischer Information angereichert) werden?
- Wie können sie durchsucht und für weitere Schritte bearbeitet werden?

Ein Korpus wird als eine Sammlung von Texten verstanden, die nach bestimmten (sprachwissenschaftlichen) Kriterien ausgewählt werden. Bei diesen Texten kann es sich sowohl um Schriftsprache (bspw. Zeitungsartikel, Romane), als auch um mündliche Sprache (bspw. Mitschnitte von Gesprächen zwischen Freunden, politische Reden) handeln. Korpora können anhand der gesammelten Daten unterschieden werden in:

- Textkorpora
Sammlung von Texten (geschriebener Text, transkribierte gesprochene Sprache)
- Baubanken
Texte mit Anzeige der syntaktischen Struktur der Sätze
- Sprachkorpora
Sprachdaten (Sprachsignal u. U. mit Transkription, phonetische Annotation)
- Multimodale Korpora
Sprachdaten mit Transkription und weiteren Information wie Mimik, Gestik...

Unterschiedliche Arten von Korpora lassen sich jedoch hinsichtlich einer Vielzahl von Parametern festmachen (für eine differenzierte Darstellung siehe Scherer 2006: 17). Wichtig ist, dass es sich bei diesen Korpora um Ausschnitte einer Sprache handelt und dass sich die Aussagen, die hinsichtlich einer eingangs formulierten Fragestellung gemacht werden können, ausschließlich auf das genutzte Korpus und die darin enthaltenen Texte beziehen. Es ist daher wichtig, sich vorab Gedanken darüber zu machen, welche Textsorten in welchem Umfang in das Korpus mit aufgenommen werden sollen. Dies hängt u.a. vom Untersuchungsgegenstand ab. Dabei ist zu klären, welche Grundgesamtheit das Korpus abbilden soll. Geht es bei der Untersuchung bspw. um Jugendsprache, so sollte das Korpus andere Texte beinhalten als bei einer Untersuchung zum Satzbau im Althochdeutschen.

8.2 Aufbau eines Korpus

Korpora liegen typischerweise in elektronischer Form vor und können so mit Hilfe des Computers relativ schnell bearbeitet, durchsucht und analysiert werden. Die Qualität der Untersuchung hängt dabei u.a. von der Vorverarbeitung der sprachlichen Daten ab.

8.2.1 Schritte der Vorverarbeitung:

- Tokenisierung
- POS-Tagging (Wortartenzuweisung)

➤ Lemmatisierung

Die Vorverarbeitung dient dazu, das Korpus bzw. die darin enthaltenen Texte für eine Vielzahl linguistischer Fragestellungen nutzbar zu machen. An die Vorverarbeitung schließt sich die Anreicherung des Korpus mit linguistischer Information (Annotation). Die Frage ist dabei, welche linguistischen Informationen zu den Sprachdaten hinzugefügt und mit ihnen aligniert werden sollen. Auch dies richtet sich nach dem Untersuchungsgegenstand und den Forschungsfragen. Auf allen Ebenen der Vorverarbeitung und auch bei der Annotation werden linguistische Entscheidungen getroffen, die Auswirkungen auf die Nutzbarkeit eines Korpus haben und deshalb explizit formuliert werden müssen. Sowohl die Annotation als auch die Vorverarbeitung stellen eine Interpretation der sprachlichen Daten dar, wodurch erst ein Großteil der linguistischen Informationen entsteht, die aus einem Korpus extrahiert werden können.

Tokenisierung

Bei der Tokenisierung erfolgt die Festlegung von kleinsten Einheiten innerhalb des Korpus (Token). Eine gängige Definition für ein Token ist eine von Leerzeichen oder Interpunktion begrenzte Folge von Buchstaben oder Ziffern. Im Zuge der Tokenisierung werden alle Satzzeichen von den Wörtern abgetrennt. In (2) handelt es sich also um 12 Token.

(6) Text: „Das können wir aber besser“, sagt der Trainer.

Token: „ Das können wir aber besser ” , sagt der Trainer .

Nachdem nun diese Zerlegung erfolgt ist, können diese kleinsten Einheiten danach wieder in größere Einheiten zusammengefügt werden. So wird bspw. die Stadt *New York* im Zuge der Tokenisierung in die beiden Token *New* und *York* zerlegt. Da jedoch eine feste Verbindung zwischen diesen beiden Teilen besteht, würde man sie evtl. als ein und nicht als zwei Token werten wollen. Um diese Zusammenfügung durchzuführen, wird meistens mit regulären Ausdrücken oder wie im Fall von *New York* mit Listen gearbeitet, die die Ausdrücke enthalten, die nach der Tokenisierung wieder verbunden werden sollen. Solche Listen und regulären Ausdrücke kommen auch bei der Trennung von Tokens zum Einsatz. Beispiele dafür sind Wörter wie *siehste* und *gibts*. Bei dieser Arbeit mit Listen und regulären Ausdrücken handelt es sich also erneut um Schritte, bei denen linguistische Entscheidungen in den Korpusaufbau mit einfließen, die sich auf alle weiteren Verarbeitungsebenen auswirken. Ein linguistisch aufbereitetes Korpus ist daher nicht theorieneutral.

Ein weiterer Bestandteil der Tokenisierung ist die Satzendeerkennung. Dabei geht es um die Mehrdeutigkeitsauflösung des Punktes. Dieser kann sowohl ein Satzende markieren, als auch in Abkürzungen oder Zahlen auftreten. Für die automatische Tokenisierung ist diese Desambiguierung nicht unproblematisch. So kann es auch Sätze geben, die mit einer Abkürzung enden, bei denen also Abkürzungs- und Satzende punkt zusammenfallen.

(7) Ich wohne in Berlin.

(8) usw., etc., Pkt. (Problem: Ein Satz endet auf eine Abkürzung)

(9) 3. Stock, 9.00 Uhr

Tagging

Der Begriff „Taggen“ bezeichnet i.d.R. die Wortartenzuweisung (part-of-speech = POS), weshalb meist von POS-Tagging die Rede ist, und erfolgt tokenbasiert. Man ordnet also linguistische Information den kleinsten Texteinheiten zu. Die Token erhalten dabei sogenannte Tags. Dabei hat sich das Tag-Inventar (Liste aller verwendeten Wortartenlabel) des Stuttgart-Tübingen-Tagsets (STTS) durchgesetzt, das für eine Vielzahl unterschiedlicher Korpora genutzt wird (vgl. Kapitel zum DWDS-Korpus). Ein Tagset bewegt sich immer zwischen einer möglichst hohen Genauigkeit und Ausdifferenzierung der Label einerseits

und einer angemessenen Handhabbarkeit andererseits. Ein Beispiel für eine getaggte Sequenz ist (6).

- (10) Peter studiert in Berlin,
 Peter/NE studiert/VVFIN in/APPR Berlin/NE ./\$.
 (Token/Tag) NE = Eigennamen; VVFIN = finites Voll-Verb; APPR = Präposition; \$. = satzbeendende Interpunktion

Das Tagging bildet die Grundlage für weitere linguistische Arbeitsschritte (bspw. syntaktische Strukturen) und ermöglicht das Auffinden von Wörtern (Token) in einer bestimmten Position einer bestimmten Sequenz von Wortarten. Eine mögliche Suchanfrage an ein Korpus kann also lauten: alle Sequenzen, in denen drei Adjektive in einem Satz auftreten. Des Weiteren ermöglicht das Tagging die Disambiguierung mehrdeutiger Wortformen, wenn sie verschiedenen Wortarten zugehören (7).

- (11) Sitz gerade!
 Sitz/NN oder VVIMP gerade/ADJD !/\$.
 (Token/Tag) NN = normales Nomen; VVIMP = Imperativ, voll; ADJD = adverbiales oder prädikatives Adjektiv

Zum Einsatz kommen dabei symbolische, stochastische oder hybride Tagger. Symbolische Tagger verwenden Regeln wie: Wenn ein Wort zwischen normalem Nomen und Verb im Imperativ ambig ist, dann verwende das Verb-Tag, wenn das folgende Token ein adverbiales oder prädikatives Adjektiv ist. Stochastische Tagger hingegen werden vorher einem Training unterzogen, bei dem sie anhand eines Trainingskorpus die Frequenzen von Wörtern und Tags zählen und daraus Wahrscheinlichkeiten für Abfolgen errechnen. Hybride Tagger verbinden diese beiden Verfahren.

Lemmatisierung

Bei der Lemmatisierung geht es um flexionsmorphologische Informationen, die für bestimmte Entscheidungen nötig sind. Dabei wird das konkrete Vorkommen eines Wortes, also eine konkrete Wortform, auf eine abstrakte Grundform (Lemma) zurückgeführt. Für die Lemmatisierung wird daher ein Lexikon benötigt, in dem die Lemmata mit ihrer jeweiligen Flexionsklasse aufzufinden sind, und ein Regelwerk, das in der Lage ist, die flektierten Formen auf das jeweilige Lemma zurückzuführen. Ein Beispiel für eine lemmatisierte Sequenz ist (8).

- (12) Die Hunde laufen durch den Park.
 Die/ART/d Hunde/NN/Hund laufen/VVFIN/laufen durch/APPR/durch den/ART/d Park/NN/Park ./\$./*
 (Token/Tag/Lemma) ART = bestimmter Artikel; NN = normales Nomen; VVFIN = finites Voll-Verb; APPR = Präposition; \$. = satzbeendende Interpunktion; d = Lemma für definiten Artikel; * = unbekanntes Lemma

Wie die abstrakten Grundformen benannt werden bzw. welche Labels ihnen zugewiesen werden, ist im Grunde frei wählbar. Ist die Entscheidung für eine Grundform wie *d* für die Ausprägungen des bestimmten Artikels getroffen worden, muss diese jedoch durchgehalten werden. Durch die Lemmatisierung wird es möglich, alle Flexionsformen eines Wortes in einem Korpus zu finden. Im Grunde geht dies auch schon nach der Tokenisierung. Nur muss man dafür jede Form in der Suchanfrage angeben. Sucht man hingegen nach dem Lemma *spielen*, findet man: *spiele, spielst, spielt, spielen, spiel, gespielt* etc. Probleme bei der Lemmatisierung und auch beim POS-Tagging stellen unbekannte Wörter dar.

8.2.2 Annotation

Bei der Annotation sprachlichen Daten ist eine Vielzahl von Annotationsebenen denkbar. So kann nun nach der Vorverarbeitung in einem nächsten Schritt eine syntaktische Annotation der Daten erfolgen, bei der die einzelnen Satzglieder bestimmt und mit definierten Labels belegt werden. Eine Darstellung möglicher Annotationsebenen findet sich unter der fo

Ebene	Annotation
Morphosyntax	Wortarten (POS)
Morphologie Lemma	Flexionsmorphologie Grundform
Syntax	Konstituenten, oft mit syntaktischen Funktionen; topologische Felder
Semantik	Lesarten, thematische Rollen
Pragmatik	Koreferenzen, Informationsstruktur, Diskursstruktur
Weitere	Orthographie, Fehlerannotation, phonetische/phonologische Merkmale, Mimik, Gestik

(Tabelle übernommen nach Lemnitzer/Zinsmeister 2006: 64)

Unabhängig davon, für welche Annotationsebene man sich entscheidet, ist der Kerngedanke dabei die Extraktion von linguistischen Informationen aus den sprachlichen Daten. Der reine Text liefert an der Oberfläche nur weniger dieser Informationen. Erst durch die Interpretation der Daten und ihre Untersuchung in einem bestimmten Kontext entstehen diese linguistischen Informationen.

8.3 Eine Auswahl an bereits bestehenden Korpora

Die Liste deutschsprachiger Korpora ist relativ umfangreich. Eine detaillierte Auflistung findet sich bei Lemnitzer/Zinsmeister 2006: 115-126. An dieser Stelle werden ein paar dieser Korpora, ihre Eigenschaften und ihre Zugangsmöglichkeiten aufgelistet, bevor im folgenden Kapitel genauer auf das DWDS-Korpus eingegangen wird.

Korpus	Zugang	Funktion	Medium	Größe (Token)	Annotation	Verfügbarkeit
Bonner Zeitungskorpus	www.ids-mannheim.de/kt/projekte/korpora/archiv/bzk.html	Teil der IDS-Korpora	geschrieben	3,1 Mio.	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.	gesprochen, geschrieben	100 Mio.	Lemma, Morphosyntax	online
Fehler-	www.linguisti	Untersuchung	geschrieben	im Aufbau	im Aufbau	frei

Annotiertes Lernerkorpus (FALKO)	k.hu-berlin.de/institut/professuren/korpuslinguistik/falko	ngen zu Fehlern von DAF-Lernern			flexible Architektur	
----------------------------------	--	---------------------------------	--	--	----------------------	--

IDS-Korpora	www.ids-mannheim.de/kl/projekte/korpora/	Empirische Basis für ling. Forschung	geschrieben	mehr als 2 Mrd.	Textstruktur, teilweise Morphosyntax	online, teils Verkauf
NEGRA-Korpus	www.coli.uni-saarland.de/projects/sfb378/negra-corpus/negra-corpus.html	Empirische Basis für ling. Forschung und maschinelle Lernen	geschrieben	0,36 Mio.	Morphosyntax, Konstituenten, grammatische Funktionen	Lizenz
TIGER-Korpus	www.ims.uni-stuttgart.de/projekte/TIGER/TIGERCorpus/	ling. und computerling. Forschung	geschrieben	0,9 Mio.	Morphosyntax, Morphologie, Syntax, Semantik	kostenlose Lizenz
Tübinger Baumbank des Deutschen (TüBa-D/Z)	www.sfs.uni-tuebingen.de/de/tuebadz.shtml	Datenbasis für maschinelle Lernen	geschrieben	0,38 Mio.	Morphosyntax, Morphologie, Syntax, topologische Felder, Semantik, Koreferenz	Lizenz

(Tabelle gekürzt übernommen nach Lemnitzer/Zinsmeister 2006: 115-126)

Literatur

Lemnitzer, Lothar/Zinsmeister, Heike (2006): Korpuslinguistik. Eine Einführung. Tübingen. Materialien der korpuslinguistischen Seminare von Prof. Anke Lüdeling (<http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/lehre>)
 Scherer, Carmen (2006): Korpuslinguistik. Heidelberg (Kurze Einführungen in die germanistische Linguistik Bd. 2).

8.4 Das DWDS-Korpus

Grundsätzliches

Das *Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts* ist ein Projekt an der Berlin-Brandenburgischen Akademie der Wissenschaften. Ziel ist ein Wörterbuch auf der Basis großer Textkorpora, wobei auf dem *Wörterbuch der deutschen Gegenwartssprache* (WDG) aufgebaut wird. Es wurde hierzu ein *Kernkorpus* der deutschen Sprache des 20. Jahrhunderts konstruiert, das zeitlich und inhaltlich ausgewogen ist:

- Umfang: > 120 Millionen Wörter aus ca. 80.000 Dokumenten
- Abdeckung: 1900 – 2000 (Erweiterungen sind vorgesehen)
- Jedes Jahrzehnt soll gleich stark vertreten sein (10 Mio Wörter)
- 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%)

Daneben sind weitere Korpora zugänglich, z.B. das *ZEIT*-Korpus, das Korpus der *Berliner Zeitung* und des *Tagesspiegels*, ein *DDR Korpus*, ein *Wendekorpus* und ein *Korpus gesprochener Sprache*.

Zugang: www.dwds.de. Anonyme Verwendung möglich, Registrierung empfohlen: Ermöglicht Zugriff auf mehr Ergebnisse und erweiterte Funktionen.

Suchmöglichkeiten

Lemma-Suche (d.h. mit *lachen* findet man *lacht*, *lachte*, *gelacht* usw.)

Abfrage: lachen #less_by_date[1900-01-01,2000-12-31] #has_field[avail,/MR* OR7W OR3S OR1S/] #cntxt 1 :kern01,kern02,kern03,kern04,kern05,kern06,kern07,kern08,kern09,kern10	
Trefferanzahl: 14670. Davon sind 9586 Treffer aufgrund rechtlicher Nutzungsvereinbarungen anzeigbar.	
Seite: 1 2 3 4 5 6 7 8 9 10 >> 	
1	Ze 1900 ... erdacht, dass die verführte Gesellschaft dabei gelacht hat, scheint mir glaubhaft. Und es reicht ...
2	Ze 1900 ... In Wien wurde schon lange nicht so viel gelacht wie am Tage dieser Enthüllung. - - - - -
3	Ze 1900 ... häufig, wie ich weiß, an der Hand des Lehrbuches lachend kontrolliert und rühmen ihm selbst jene ...
4	Ze 1900 ... wenn der Director fragt: 'Haben die Leute nicht gelacht ? Haben sie nicht applaudiert und sogar den ...
5	Ze 1900 ... zu rüsten, daran dachte man nicht. Uebermütig lachte man gelegentlich über einen Fall muckerischer ...
6	Ze 1900 ... Mitglieder der Verwaltung grinsten und lachten , als wären sie beim » Schwarzen Adler «. Die ...
7	Be 1900 ... war sie unwiderstehlich. » Siehst Du, da lachst Du schon. Es war aber auch höchste Zeit, denn ...

Stringsuche ist natürlich ebenfalls möglich, mit *@lachend* findet man den String:

Trefferanzahl: 1586. Davon sind 1195 Treffer aufgrund rechtlicher Nutzungsvereinbarungen anzeigbar.	
Seite: 1 2 3 4 5 6 7 8 9 10 >> 	
1	Ze 1900 ... häufig, wie ich weiß, an der Hand des Lehrbuches lachend kontrolliert und rühmen ihm selbst jene ...
2	Be 1900 ... nickte er der Alten lachend zu und ging. Während ...
3	Be 1900 ... die Freude dieser Fahrt; lachend schüttelte er die Schneeklumpen ...

Die Korpora sind morphologisch "getaggt", was Fragen nach grammatischen Kategorien ermöglicht. Beispiele:

Arzt *Arzt, Arztes, Ärzte* ... (alle flektierten Formen von Arzt)
 @Arzt *Arzt* (nur die Wortform *Arzt*)
 Arzt* *Arzt, Arztbesuch, Arztberuf, ...*
 *arzt *Sportarzt, Hausarzt, ...*
 "gute Arzt" *guter Arzt, bester Arzt, gute Ärzte, ...*

"das gute Beispiel" *das gute Beispiel, das beste Beispiel, die besseren Beispiele ...*
 "Kanzler #1 Schröder" *Kanzler Schröder, Kanzler Gerhard Schröder,*
... (Kanzler und Schröder im Abstand von höchstens einem Wort)
 Kanzler && Schröder (Sätze, in denen *Kanzler* und *Schröder* zusammen vorkommen)
 Kanzler || Schröder (Sätze, in denen *Kanzler* oder *Schröder* vorkommt)
 Kanzler && !Schröder (Alle Sätze, in denen *Kanzler*, aber nicht *Schröder* vorkommt)
 \$p=NE with Herzog *Roman Herzog, Peter Herzog, ...* (Eigennamen mit Herzog)
 "Ägide #2 \$p=NE" *Ägide Bush, Ägide von Harald Szeemann, Ägide von Vodafone,*
 \$p=NN with *zeit *Weihnachtszeit, Übergangszeit, Halbzeit,*
 "üben #5 aus with \$p=PTKVZ" *... übt er ein Wahlamt aus ...* (Präfixverben)
 "schälen with \$p=VVFIN #5 aus with \$p=PTKVZ" *schälen* als Verb gefolgt von einem
 separablem Präfix *aus*
 "sein with \$p=VVFIN #20 \$p=VVPP #0 @worden": Phrasensuche mit drei Wortformen:
 Auxiliar *sein* gefolgt von einem Partizip und *worden*

Tagset STTS (Stuttgart-Tübinger Tagset)

ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell [er ist] schnell
ADV	Adverb	schon, bald, doch
APPR	Präposition; Zirkumposition links	in [der Stadt], ohne [mich]
APPRART	Präposition mit Artikel	im [Haus], zur [Sache]
APPO	Postposition	[ihm] zufolge, [der Sache] wegen
APZR	Zirkumposition rechts	[von jetzt] an
ART	bestimmter oder unbestimmter Artikel	der, die, das, ein, eine, ...
CARD	Kardinalzahl (Ordinalzahlen sind als ADJA getaggt)	zwei [Männer], [im Jahre] 1994
FM	Fremdsprachliches Material	[Er hat das mit ``] A big fish [`` übersetzt]
ITJ	Interjektion	mhm, ach, tja
KOUI	unterordnende Konjunktion mit ``zu'' und Infinitiv	um [zu leben], anstatt [zu fragen]
KOUS	unterordnende Konjunktion mit Satz	weil, daß, damit, wenn, ob
KON	nebenordnende Konjunktion	und, oder, aber
KOKOM	Vergleichskonjunktion	als, wie
NN	normales Nomen	Tisch, Herr, [das] Reisen
NE	Eigennamen	Hans, Hamburg, HSV
PDS	substituierendes Demonstrativ- pronomen	dieser, jener
PDAT	attribuierendes Demonstrativ- pronomen	jener [Mensch]
PIS	substituierendes Indefinit- pronomen	keiner, viele, man, niemand

PIAT	attribuierendes Indefinit- pronomen ohne Determiner	kein [Mensch], irgendein [Glas]
PIDAT	attribuierendes Indefinit- pronomen mit Determiner	[ein] wenig [Wasser], [die] beiden [Brüder]
PPER	irreflexives Personalpronomen	ich, er, ihm, mich, dir
PPOSS	substituierendes Possessiv- pronomen	meins, deiner
PPOSAT	attribuierendes Possessivpronomen	mein [Buch], deine [Mutter]
PRELS	substituierendes Relativpronomen	[der Hund ,] der
PRELAT	attribuierendes Relativpronomen	[der Mann ,] dessen [Hund]
PRF	reflexives Personalpronomen	sich, einander, dich, mir
PWS	substituierendes Interrogativpronomen	wer, was
PWAT	attribuierendes Interrogativpronomen	welche [Farbe], wessen [Hut]
PWAV	adverbiales Interrogativ- oder Relativpronomen	warum, wo, wann, worüber, wobei
PAV	Pronominaladverb	dafür, dabei, deswegen, trotzdem
PTKZU	``zu'' vor Infinitiv	zu [gehen]
PTKNEG	Negationspartikel	nicht
PTKVZ	abgetrennter Verbzusatz	[er kommt] an, [er fährt] rad
PTKANT	Antwortpartikel	ja, nein, danke, bitte
PTKA	Partikel bei Adjektiv oder Adverb	am [schönsten], zu [schnell]
TRUNC	Kompositions-Erstglied	An- [und Abreise]
VVFIN	finites Verb, voll	[du] gehst, [wir] kommen [an]
VVIMP	Imperativ, voll	komm [!]
VVINF	Infinitiv, voll	gehen, ankommen
VVIZU	Infinitiv mit ``zu'', voll	anzukommen, loszulassen
VVPP	Partizip Perfekt, voll	gegangen, angekommen
VAFIN	finites Verb, aux	[du] bist, [wir] werden
VAIMP	Imperativ, aux	sei [ruhig !]
VAINF	Infinitiv, aux	werden, sein
VAPP	Partizip Perfekt, aux	gewesen
VMFIN	finites Verb, modal	dürfen
VMINF	Infinitiv, modal	wollen
VMPP	Partizip Perfekt, modal	gekonnt, [er hat gehen] können
XY	Nichtwort, Sonderzeichen enthaltend	3:7, H2O, D2XW3
\$,	Komma	,
\$.	Satzbeendende Interpunktion	. ? ! ; :
\$(sonstige Satzzeichen; satzintern	- [,]()

Beispiel: Die Gradpartikel *ausgerechnet*.

Suche nach "@ausgerechnet" erbringt auch das Partizip:

Abfrage: @ausgerechnet #less_by_date[1900-01-01,2000-12-31] #has_field[avail,/MR*|OR7W|OR3S|OR1S/] #cntxt 1 :kern01,kern02,kern03,kern04,kern05,kern06,kern07,kern08,kern09,kern10

Trefferanzahl: 1387. Davon sind 813 Treffer aufgrund rechtlicher Nutzungsvereinbarungen anzeigbar.

Seite: << 8 9 10 11 12 13 14 15 16 17

751	Ze	1999	... "The story of Berlin". Warum wird Geschichte ausgerechnet in diesem Karree angeboten? Es wurde in den ...
752	Be	1999	... in dem ich endlich begriff, warum eigentlich ausgerechnet Autowaschen in meiner gesamten Jugend mein ...
753	Be	1999	... hineingestopft. Ich weiß auch nicht, warum ausgerechnet ich ständig Papiermüll in den Kasten gestopft ...
754	Be	1999	... Der flüchtende Dealer hatte sich wahrscheinlich ausgerechnet , am späteren Abend noch einmal vorbeizuschauen, ...
755	Ze	1999	... Plan für eine Währungsunion mitgearbeitet habe und ausgerechnet Schmidt ihn einst als "unverbesserlichen ...
756	Ze	1999	... völlig Recht. Und er hat schon blitzschnell ausgerechnet , auf welches Sümmchen wir uns freuen können: Bei ...

Einschränkung auf Adverbien durch "@ausgerechnet with \$p=ADV"

Corpus: DWDS Corpus

Abfrage: @ausgerechnet with \$p=ADV #greater_by_date[2000-12-31,1900-1-1] #has_field[avail,/MR*|OR7W|OR3S|OR1S/] #cntxt 1 :kern01,kern02,kern03,kern04,kern05,kern06,kern07,kern08,kern09,kern10

Trefferanzahl: 1336. Davon sind 776 Treffer aufgrund rechtlicher Nutzungsvereinbarungen anzeigbar.

Seite: 1 2 3 4 5 6 7 8 9 10 >>|

1	Ge	1999	... (alles dreht sich um die Erde), hatte er ausgerechnet , daß die Bewegungen der Planeten sich besser ...
2	Be	1999	... gezeigt -, Jochen unterbrechend, sagte: » Daß ausgerechnet du das sagst, der Finanzminister der Familie, der ...
3	Be	1999	... era degno di essere abitato. Warum nun Jochen ausgerechnet diese Kinderkrepelabstellkammer aufgesucht, eins ...
4	Be	1999	... paar neue Läden, den » Billigpreis «, warum der ausgerechnet neben » Conny 's Container « und schräg gegenüber ...
5	Be	1999	... zum Müllkübel. Irgendwelche Leute scheinen ausgerechnet die Wiese hinter Gerd's Grab auserkoren zu haben, ...
6	Be	1999	... meine Oma; sie starb in dem Monat, den sie vorher ausgerechnet hat. Sie hat mir ihre Bücher gegeben, kurz ...

Allerdings ist die Lemmatisierung fehlerhaft, wie das letzte Beispiel zeigt! (Die Lemmatisierung wird weitgehend automatisch erzeugt und ist daher mit Vorsicht zu verwenden).

Darstellungsoptionen

Die Resultate der Suche können in der KWIC-Darstellung (siehe oben) dargestellt werden, aber auch im Satzzusammenhang. Wenn man das Zielwort anklickt, bekommt man den Textkontext angezeigt. Beispiel: *erstgenannte*

12	Wi	1984	... das Grundgesetz der Selbstreferenz, aber für die erstgenannte Gruppe in einem radikaleren, ausschließlicheren ...
13	Wi	1984	... des Unwahrscheinlichen fasziniert sind. Die erstgenannte Variante hat die Tradition für sich, die zweite ...

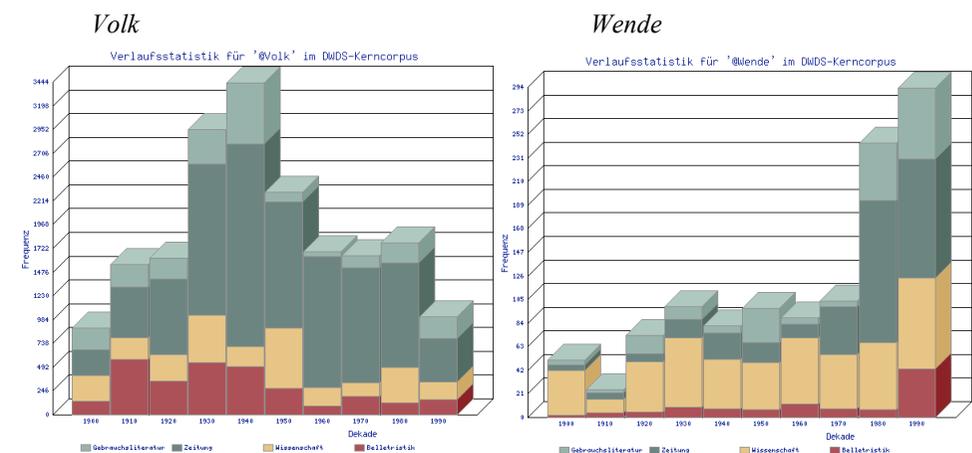
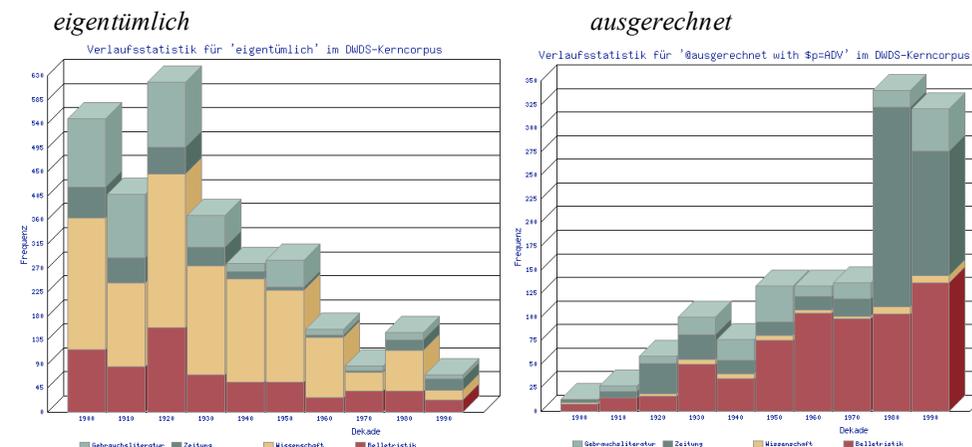
Luhmann, Niklas, Soziale Systeme, Frankfurt a.M.: Suhrkamp 1984, S. 164 [Schließen](#)

Sie sind alle mit Bezug auf Zeitprobleme gearbeitet, und sie beziehen sich alle auf Sonderprobleme bestimmter Funktionssysteme -. In beiden Hinsichten bleiben sie zunächst abhängig von Problemen, die im Umbau des Gesellschaftssystems von einer ständischen Ordnung in ein funktional differenziertes System vordringlich behandelt und in neue semantische Formen gebracht werden müssen -. Auch die Frage, wie soziale Ordnung möglich ist, gewinnt im neuzeitlichen, wissenschaftsspezifischen Reflexionsstil analytische Schärfe -, und speziell die Soziologie beginnt seit etwa hundert Jahren, ihre theoretische Integration an Hand einer solchen Fragestellung zu suchen -. In spezifisch soziologischen Problemstellungen kann man jene beiden Fragemuster ebenfalls unterscheiden, und entsprechend gibt es auch hier nette, hilfsbereite Theorien und solche, die durch das Wahrscheinlichwerden des Unwahrscheinlichen fasziniert sind. Die **erstgenannte** Variante hat die Tradition für sich, die zweite drängt sich auf, sobald man explizit fragt, wie soziale Ordnung möglich ist. Die in der Tradition ganz vorherrschende Auffassung sieht das Problem sozialer Ordnung in der Vermeidung oder Unterdrückung widerwärtigen Verhaltens, feindselliger, störender, schädlicher Aktivitäten, die verhindern, daß andere in sozialen Beziehungen zu ihrem Recht kommen, ihre Bedürfnisse befriedigen, sich wohl fühlen können.

14	Wi	1981	... Yukio erklärt zu dem erstgenannten Beispiel, daß hier eine ...
15	Wi	1979	... in der Kompos. die erstgenannte Darstellungsweise eine erhebliche Rolle ...

Historische Entwicklung

Die Häufigkeiten lassen sich nach Jahrzehnten und nach den Textsorten darstellen. Man kann z.B. zeigen, dass *eigentlich* im Verlauf des 20. Jahrhunderts immer seltener verwendet wurde, dass *ausgerechnet* hingegen zugenommen hat, dass *Volk* im Nationalsozialismus eine Blüte erlebte,



Kollokationen

Häufig auftretende Kombinationen können dargestellt werden. Beispiel: *Volk*

Lemmabasierte Kollokationssuche im DWDS-Kerncorpus	
Suchbegriff: Volk	Submit Query
Optionen	
Statistisches Maß:	Frequenz des Bigrams

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

8.5 Das Internet als Korpus?

8.5.1 Chancen und Probleme

Eine Möglichkeit, an linguistische Daten zu kommen, ist das Internet. Vorteil: Es bietet sehr große Textmengen. Allerdings sollte man sich bewusst sein:

- Es gibt kein allgemeines Prinzip, nach dem die im Internet verfügbaren Texte zusammengestellt sind.
- Die Texte sind von sehr unterschiedlicher Qualität. Sie können von Nicht-Muttersprachlern verfasst worden sein. Sie können regionale Verschiedenheiten repräsentieren. Man sollte sich also die Herkunft der Belege genau ansehen.
- Viele Texte kommen mehrfach vor und werden mehrfach von Suchmaschinen gefunden.
- 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.

Für die absolute statistische Auswertung ist das Internet nicht geeignet, allenfalls für die relative (z.B.: Vergleich der Häufigkeit zweier Formen).

Ferner ist es gut geeignet, um einzelne Beispiele (ohne statistische Auswertung) zu finden.

8.5.2 Google

Die naheliegendste Möglichkeit ist, eine Suchmaschine zu verwenden, z.B. Google. Google liefert Trefferzahlen (allerdings kann es sich um mehrfache Kopien handeln). Man kann Google ferner einstellen nach Domäne (dies korreliert oft mit der Sprache, z.B. *de* für Deutsch). und auch nach Dateityp. (Google advanced).

Wichtig:

- Sie können nach Ketten von Wörtern suchen: “sogar nur“

- Sie können mit Joker (wildcards) für Wörter suchen: “sogar * nur“, “sogar * * nur“;
- Google nimmt in gewissem Grade bereits eine Lemmatisierung vor, d.h. *Auto* findet auch Suchstellen mit *Autos*.

8.5.3 Webcorp

Dieser Dienst will einen Zugang zum Internet bereitstellen, der für linguistische Zwecke versatiler ist.

Verbesserte Internet-Abfragen durch <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 für mehrfache Wörter
- Alternativen in eckigen Klammern: “[ship|boat] sank”, “[ship|boat] s[a]n[k]”; “[r|u]n[n]ing[s]” für *running, runs, run, ranning, rans, ran*

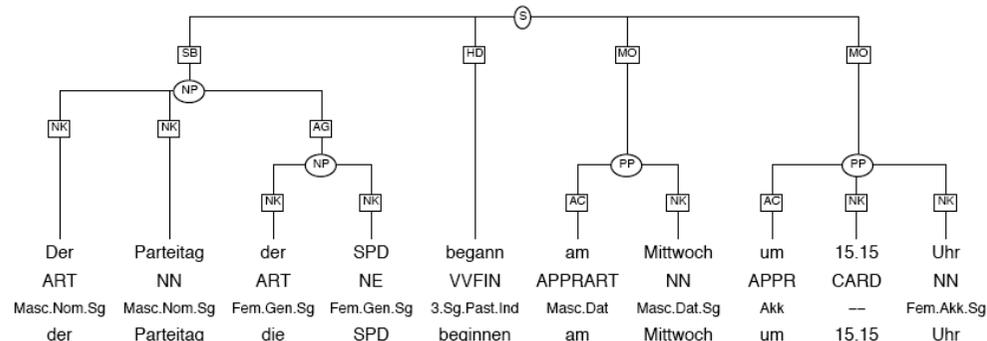
Sie können das Output-Format steuern (verschiedene Formen von Konkordanzen) und die Domäne bestimmen (Sprachen). Sie können angeben, welche Wörter im Text vorkommen müssen bzw. nicht vorkommen dürfen (+/-).

8.6 Ein syntaktisch annotiertes Korpus: TIGER

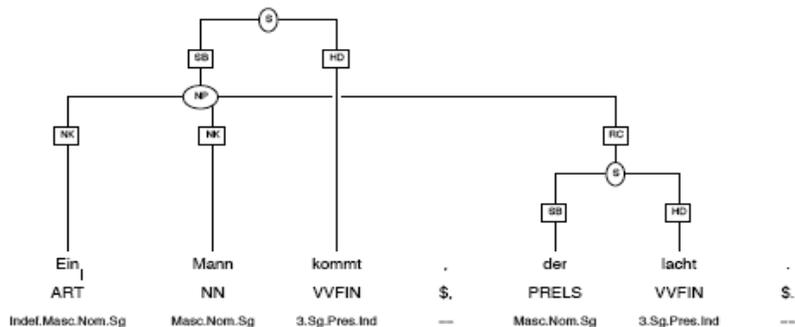
Dieses Korpus enthält mehr als 40,000 Sätze aus Zeitungstexten und ist nicht nur morphologisch annotiert, sondern auch syntaktisch geparkt. Informationen unter:

<http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>

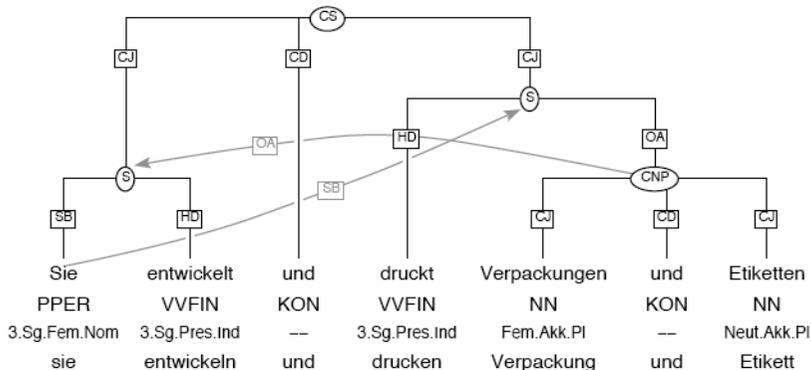
Beispiel einer syntaktischen Annotation: Graphen.



Beispiel mit extraponiertem Relativsatz, kreuzende Kanten:



Beispiel mit Koordinationen, geteilte Konstituenten.



Ziel der syntaktischen Annotation ist nicht die Analyse nach bestimmten syntaktischen Theorien; die Annotation soll in einem gewissen Sinn theorie-neutral sein. Es sollen

vielmehr die Strukturen geschaffen werden, die Abfragen erlauben, die von Interesse für die Syntaxforschung sind.

Für die Suche nach Strukturen wurde eine Abfragesprache entwickelt:

(13) > Dominanzbeziehung, >* verallgemeinerte Dominanz, . Präzedenz, \$ Geschwisterbeziehung

Die folgende Frage identifiziert Sätze mit pronominalen Subjekten:

(14) [cat = "S"] >SB [pos = "PPER"]

Die folgende Frage will solche Sätze mit finitem Verb:

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

8.7 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

8.8 Informationen zu linguistischen Korpora

Korpus-Seite der *Linguist List*: <http://www.linguistlist.org/sp/Texts.html>

Umfangreichere Referenz auf Sprachkorpora: <http://devoted.to/corpora>

Korpora am Institut: Anmeldung unter <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/institutkorpora>

Einführendes Werk:

Lothar Lemnitzer & Heike Zinsmeister (2006) *Korpuslinguistik*. Narr, Tübingen.

8.9 Lexika im Internet

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

8.9.1 Wörterbuch-Portale

Über das Portal <http://germazope.uni-trier.de/Projects/WBB> sind zahlreiche deutsche Wörterbücher erfasst, unter anderem:

- Deutsches Wörterbuch von Jacob und Wilhelm Grimm (!)
 - Mittelhochdeutsche Wörterbücher von Wilhelm Müller und Friedrich Zarncke
 - Diverse Mundarten-Wörterbücher
 - Das Goethe-Wörterbuch der Berlin/Brandenburgischen Akademie der Wissenschaften.
- Das Wörterbuch von Adelung findet man hier: <http://germazope.uni-trier.de/Projects/WBB>

8.9.2 Wortschatz Uni Leipzig: <http://wortschatz.uni-leipzig.de/>

Ein Wörterbuch-Portal zum heutigen Deutsch: Das schließt Hinweise zu aktuellen Wörtern mit ein. Hier finden Sie auch Wörterbüchern, auch Beispiel Wortsuche: *aufheben*

Wort: aufheben
Anzahl: 619
Häufigkeitsklasse: 13 (d.h. *der* ist ca. 2¹³ mal häufiger als das gesuchte Wort)
Morphologie: aufheben
Grammatikangaben: Wortart: Verb
 Partizip II mit haben lautet ab
 transitiv
 Präfix: auf

Relationen zu anderen Wörtern:

- Synonyme: [abschaffen](#), [abschließen](#), [aufbewahren](#), [aufklauben](#), [auflesen](#), [abeseitigen](#), [bewahren](#), [einstellen](#), [erheben](#), [erübrigen](#), [hamstern](#), [heben](#), [hochspeichern](#), [zurückbehalten](#), [zurückhalten](#), [zurücklegen](#)
- vergleiche: [annullieren](#), [beseitigen](#), [streichen](#), [verwahren](#), [widerrufen](#)
- ist Synonym von: [absagen](#), [abschaffen](#), [abzweigen](#), [abzweigen](#), [annullieren](#), [aufnehmen](#), [aufraffen](#), [aufsammeln](#), [aufsparen](#), [aufsuchen](#), [ausgleichen](#), [auskleben](#), [einmotten](#), [einschließen](#), [einsparen](#), [emporheben](#), [erheben](#), [heberklauben](#), [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 Beschränkungen aufheben](#)
- -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](#), [ei](#)

8.9.3 OWID (Online-Wortschatz-Informationssystem Deutsch): <http://www.owid.de/>

Dies ist das Lexikon-Portal des Instituts für Deutsche Sprache (IDS) in Mannheim. Es bietet Zugang zu verschiedenen Lexika und soll in Zukunft weiter ausgebaut werden (Valenzlexika, grammatische Information usw).

Gegenwärtig: [lexikon](#), [Neologismen](#), [Feste Wortverbindungen](#), [Diskurswörterbuch 1945-55](#)
 Beispiel einer Kollokationsanzeige für *aufheben*:

770202 verschiedene Kookkurrenzpartner gespeichert

aufheben als Kookkurrenzpartner

alphabetisch nach Kohäsion nach Wortart

Suchmuster (regex) 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	

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 aufhöbst
	er/sie/es	aufhob	er/sie/es	aufhöbe
	wir	aufhoben	wir	aufhöben
	ihr	aufhobt	ihr	aufhöbet aufhöbt
	sie	aufhoben	sie	aufhöben

8.9.4 Canoonet: <http://www.canoo.net/>

Wörterbuchportal mit zahlreichen grammatischen und idiomatischen Informationen (siehe oben rechts für *aufheben*).

8.9.5 Weitere Wörterbücher:

Hier einige Meta-Webseiten:

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

9. Transkription und Annotationswerkzeuge

9.1 Transkription gesprochener Sprache

9.1.1 Grundlegendes

Wenn man gesprochene Sprache untersuchen will (z.B. in der Konversationsanalyse) ist es zunächst nötig, die Spracherereignisse aufzunehmen (Audio, oft auch Video). Um diese Aufnahmen dann untersuchen zu können, müssen sie in aller Regel **transkribiert** und **annotiert** werden. Das heißt, sie werden verschriftlicht und mit Markierungen versehen, die für die Analyse von Bedeutung sein könnten.

Anders als bei der glättenden Niederschrift etwa von Interviews in Zeitungen geht es oft darum, Eigenschaften des mündlichen Sprachgebrauchs exakt festzuhalten. Es handelt sich dabei um Dinge wie Verschleifungen, Zögerlaute, Wiederholungen, genaue phonetische Realisierungen, prosodische Effekte, Ins-Wort-Fallen usw. Ferner können auch Elemente der Gestik, der Mimik oder der Situation beschrieben und mit den Sprechereignissen korreliert werden.

Es gibt dabei keine "absolute" Transkription, die alles mit maximaler Genauigkeit festhalten würde. Was man festhält, ist zu einem großen Teil von dem Zweck bestimmt, zu dem man etwas transkribieren und annotieren will. Es ist zu einem gewissen Teil auch theorie-geleitet. Es haben sich eine Reihe von Konventionen entwickelt, die für die Analyse gesprochener Sprache eingesetzt werden. Einige der bekannteren lauten:

- GAT ("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

Als Einführung:

Dittmar, Norbert. 2002. *Transkription. Ein Leitfaden mit Aufgaben für Studenten*. Opladen: Leske und Buderich.

9.1.2 Beispiel eines Transkriptionssystems: GAT

Selting e.a.: "Gesprächsanalytisches Transkriptionssystem (GAT)",
<http://www.mediensprache.net/de/medienanalyse/transcription/gat/gat.pdf>

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 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.

Metadaten (Transkriptionskopf)

Darunter versteht man die Angaben über eine bestimmte Aufnahme und deren Transkription. Einige der hier relevanten Aspekte:

- Herkunft, Zugehörigkeit zu einem bestimmten Korpus bzw. Projekt,

Manfred Krifka, Institut für deutsche Sprache und Linguistik, Humboldt Universität zu Berlin -- 5220156 BA Germanistische Linguistik Technische Hilfsmittel Di 12-14, DOR24, 1.102 -- 25.02.2009

- Aufnahme Nummer 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.

Wichtig: Oft ist es aus rechtlichen Gründen nötig, die Personen zu anonymisieren (man wähle Vornamen bei Du-Beziehungen, Vor- und Nachnamen bei Sie-Beziehungen).

Einige allgemeine Konventionen

- Oft werden äquidistante Schriften bevorzugt (Courier).
- Es wird generell klein geschrieben; Großbuchstaben dienen der Notation von Akzenten.
- Transkriptionszeilen werden numeriert: 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, und man kann nach den 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 (siehe Beispiel oben und im folgenden:

```
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.
```

Pausen

werden wie folgt angezeigt: (.) 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.

```
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
```

Weitere segmentale Transkriptionskonventionen

- Dehnungen werden durch :, ::, ::: wiedergegeben, z.B. *so:::*
- Verzögerungssignale durch *äh, öh* etc.
- Glottalverschluss druch ' , 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 für das es unterschiedliche Notationskonventionen gibt. Derzeit 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>

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, - gleichbleigend, ; 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-fallerder, fallend-steigender Akzent auf Silbe: *so`SO, so`SO, so`SO, so^SO, so`SO, so`SO*
- Dabei können Intonations signale 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 von Personen, Distanz von Personen)
-- Kinesik (Körperbewegungen der Personen)
-- Gestik
-- Mimik
-- Blickrichtung
in gesonderten Zeilen, mit Angabe der Dauern.

```
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
```

Für diese Zwecke kann man auch Bilder integrieren, welche die Konversationsteilnehmer zeigen.

Beispiel einer GAT-Analyse: Grobtranskript

Text: <http://www.mediensprache.net/de/medienanalyse/transcription/gat/>

```

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,
08 (--)
09 S1: heute noch-
10 (2.1)
11 s=is der UMbruch.
12 S2: n besonders GUTES beispiel das warn mal unsere NACHbarn.
13 (1.0)
14 ähm (1.0)
15 DREISSig jahre verheiratet, (--)
16 das letzte kind (.) endlich aus m HAUS,
17 zum stuDIERN, (--)
18 WEGgegangen,=ne,
19 nach berLIN, (--)
20 und (.) die ältere tochter is AUCH in berlin gewesen? (1.1)
21 und (.) der KE::RL,
22 das war aber ein peneTRANter: !WI!derling.=also (1.0)
23 der hat (.) äh sein GARTen wie (.) !PIK! AS (--) gePFLEGT,=
24 =ne,
25 !KEIN! BLÄTTchen,=
26 =und NICHTS,=
27 =englischer Rasen, (--)
28 un:dh: bei !JE!der gelegenheit hat er die poliZEI
29 gerufen,
30 und sich mit den NACHbarn ange[legt,=ne, (1.2)
31 S1: [phhh hohoho
32 S2: un wenn da: einmal: jemand zum abschied geHUPT hat,
33 da war der in NULL komma nix draußen;
34 und hat da RUMgeschrien;=
35 =ich hol die poliZEI: und [so-
36 S1: [das GIBS doch wohl nich.
37 S2: ja; V:OLLkommen WIderlich.=ne, (--)
38 un:dh: (--) dann hatte er do son (.) son KLEInen BA:RT
39 hier,=ne, (.)
40 und ham wir immer gesagt HIT[ler;=ne,
41 S1: [HITler;
42 S1: h[m,

```

Beispiel einer GAT-Transkription: Feintranskript

```

01 S1: ja;; (.) die lVIERziger 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: =<<dim> oder hschEiden lassen hÜBERhaupt.>
07 S2: hm,
08 (--)
09 S1: <<pp> heute noch- >
10 (2.1)
11 <<p> s=is der ↑UMbruch.>
12 S2: n besonders ↑GUTES beispiel das warn mal unsere ↑NACHbarn.
13 (1.0)
14 ähm (1.0)
15 hDREISSig jahre verhEiratet, (--)
16 das letzte kind (.) hEndlich aus m hHAUS,
17 zum stuhDIERN, (--)
18 hWEGgegangen,=hne, .h
19 nach berhLIN, (--)
20 und (.) die hältere tochter is hAUCH in berlin gewesen? .hh
21 und (.) der hKE::RL,
22 <<t> das war aber ein penehTRANter: h!WI!derling.=also .hh
23 der hat (.) äh sein ↑GARTen wie (.) ↑!PIK! hAS (--)
24 gehPFLEGT,=hne,
25 h!KEIN! hBLÄTTchen,=
26 =und hNICHTS,=
27 =hEnglischer hRasen, .hh
28 un:dh: bei <<all>↑!JE!der gehlEgenhheit hat er die poli↑ZEI
29 gehrufen,>
30 <<all> und sich mit den ↑NACHbarn ange[hlegt,=hne,> (1.2)
31 S1: [hpp>pphhh hohoho>
32 S2: un hWenn da: hEinmal: hjEmand zum hAbschied gehHUPT hhat,
33 da hwar der in <<all> hNULL komma nix> hdraußen;
34 und hat da hRUMgeschrien;=
35 =ich hol die poli↑ZEI: und [so-
36 S1: [hpp> das ↑GIBS doch wohl nich.>
37 S2: ja; hV:OLLkommen hWIderlich.=hne, .h
38 un:dh: (--) dann hatte er do son (.) son hKLEInen hBA:RT
39 hier,=hne, (.)
40 <<all> und ham wir immer gesagt hHIT[ler;=hne,
41 S1: [hHITler;
42 S1: hh[m,

```

9.2 Methoden zur Erstellung von Transkriptionen

9.2.1 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,

9.2.2 Partitur-Editoren: EXMARaLDA

Es gibt jedoch auch spezialisierte Werkzeuge für das Erstellen von Transkriptionen, sogenannte Partitur-Editoren, welche die zeitalinierte Transkription und Annotation auf mehreren Zeilen erlaubt. Ein solches Werkzeug ist der Editor von EXMARaLDA, der vom SFB für Mehrsprachigkeit in Hamburg entwickelt wurde, weit eingesetzt wird und kostenlos erhältlich ist (Webseite:

Grundaufbau

Der Partitureditor ist ein System zur Eingabe von Transkriptionen und zur Ausgabe in verschiedenen Formaten, insbesondere in pdf, html und rtf. Auf diese Weise können Transkriptionen in Textverarbeitungs-Dateien eingebunden werden. Es gibt auch die Möglichkeit, Transkriptionen mit Audio- oder Video-Dateien zu verknüpfen und zu synchronisieren.

Zusätzlich zum Editor gibt es auch noch Programme zum Durchsuchen von Dateien und zum Aufbau und zur Pflege von Korpora.

Prinzipien eines Partitureditors

Mit einem Partitureditor kann man teilweise gleichzeitig ablaufende Ereignisse (wie in einer Konversation) erfassen und mit einer Zeitlinie korrelieren. Es gibt auch die Möglichkeit, neben der eigentlichen Transkription auf eigenen Zeilen weitere Phänomene (Prosodie, Gestik, Mimik, evtl. auch Glossierungen, idiomatische Übersetzungen usw.) festzuhalten.

Im folgenden soll das Anlegen einer einfachen Transkription gezeigt werden.

Beispiel der Arbeit mit EXMARaLDA

- Aufruf einer neuen Datei: File → New
- Angabe der Metadaten: File → Meta information
- Angabe des Speaker-Tables: File → Speakertable; gebe zwei Sprecher ein: S1, S2
- Angabe der Partiturzeilen (Tier): Add Tier für die beiden Sprecher, evtl. je ein weiteres Tier für zusätzliche Annotationen.

- Eingabe von Text; automatische Alinierung.
- Speicherung und Ausgabe in rtf-Datei oder html-Datei

9.2.3 Das Annotationswerkzeug CLAN

Dies ist kein Partitureditor, aber erlaubt ebenfalls das einfache Erstellen von Transkriptionen. Es ist für die CHILDES-Datenbank nötig und auch dort erhältlich:

<http://chilides.psy.cmu.edu>

Man kann mit CLAN einfache textstatistische Untersuchungen vornehmen und auch Ton- und Bilddateien mit einbinden. Der Editor erlaubt verschiedene Modi, z.B. für die Transkription und die Kodierung.

9.2.4 Das Annotationswerkzeug ELAN

Dieses Werkzeug wurde vom MPI Nijmegen entwickelt, vor allem für die Annotation von Bild- und Tondateien. Das Programm ist erhältlich mit ausführlicher Dokumentation unter www.mpi.nl/tools/

9.3 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.

9.3.1 XML: Extensible Markup Language

Dies ist ein allgemeines Format zur Erfassung und Annotation von hierarchisch strukturierten Daten. Es hat eine sehr weite Verbreitung gefunden.

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).

9.3.2 Die Text Encoding Initiative (TEI)

TEI ist eine internationale Arbeitsgruppe, die Vorschläge für die Dateiformate von Sprachdaten entwickelt hat; siehe <http://www.tei-c.org/index.xml> und den Wikipedia-Eintrag zu TEI: http://de.wikipedia.org/wiki/Text_Encoding_Initiative

10. Phonetische Software: Praat

Für die Untersuchung gesprochener Sprache gibt es ein frei zugängliches Software-Paket, das Analysen erlaubt, wie sie bis in die 80er Jahre hinein nur in gut ausgestatteten Phonetiklaboren möglich waren: Praat, entwickelt vor allem vom Paul Boursma.

Die Software kann heruntergeladen werden von <http://www.fon.hum.uva.nl/praat>

10.1 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.

10.1.1 Einlesen von Sound-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 dann zur Weiterverarbeitung im PRAAT OBJECTS-Fenster als ein OBJECT.

10.1.2 Aufnahmen

Akustische Daten können auch direkt über ein Mikrofon aufgenommen werden. (Aufnahmen 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 unter einem zu gebenden Namen zur Weiterverarbeitung.
- CLOSE schließt das Fenster SOUND RECORDING.

Die Datei erscheint dann als ein Objekt im PRAAT OBJECTS-Fenster.

10.1.3 Darstellung von Sound-Dateien

Durch Auswählen eines Objekts und EDIT erhält man das SOUND EDITOR-Fenster. Hier sieht man zwei Felder:

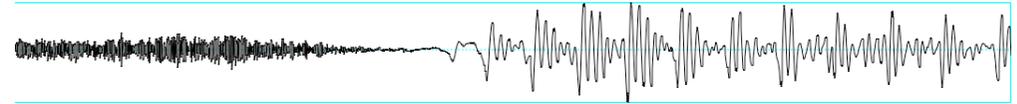
- Oben eine grafische Darstellung der Schallwelle (schwarz). Die Amplitude der Welle (die vertikale Ausdehnung) gibt die Lautstärkeintensität wieder.



- Unten (je nach Einstellung) ein Feld, in dem Analysen der Sounddatei vorgenommen werden. Die Einstellungen kann man in dem Fenster auswählen; davon später.

Die Größe des Fensters ist veränderbar (Ziehen des Cursors in der rechten unteren Ecke).

Man kann in die grafische Darstellung hineinzoomen und die Welle genau betrachten. Das folgende Beispiel ist eine Realisierung von [sa], insgesamt 0,2 Sekunden. Man sieht am Beginn die unregelmäßigen, vielfach überlagerten Schwingungen (Rauschen), die für Frikative wie das [s] typisch sind, dann die für Vokale charakteristischen periodischen Schwingungen.



Abspielen von Sound-Dateien

Man kann die Datei abspielen und kann dann am Cursor das akustische Ereignis grafisch nachverfolgen. Man drückt hierzu auf die unterste Leiste, welche die Gesamtänge der Aufnahme (in Sekunden) darstellt. Die Leiste darüber stellt die Länge dar, die grafisch dargestellt wird.

Man kann den Cursor auf beliebige Abschnitte setzen, um von dort an abzuspielen.

Man kann Abschnitte der Datei markieren (Cursor bei gedrückter Taste ziehen). Auf diese Weise kann man sich einzelne Abschnitte gezielt anhören (sehr geeignet für detaillierte Transkriptionen). Die oberste Leiste am Fußende gibt den jeweils markierten Abschnitt wieder.

10.1.4 Schneiden von Sound-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).

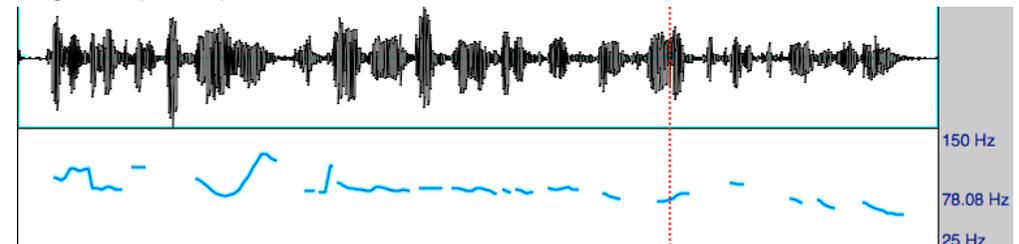
Man kann ferner durch EDIT → CUT/COPY/PASTE Teile von Sound-Dateien kopieren, ausschneiden und woanders wieder einsetzen, ganz ähnlich wie bei Textdateien.

10.2 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).

10.2.1 Stimmton (Pitch)

Mit PITCH → SHOW PITCH wird die Grundfrequenz der Stimmlippen (der Stimmton) dargestellt (in Blau).



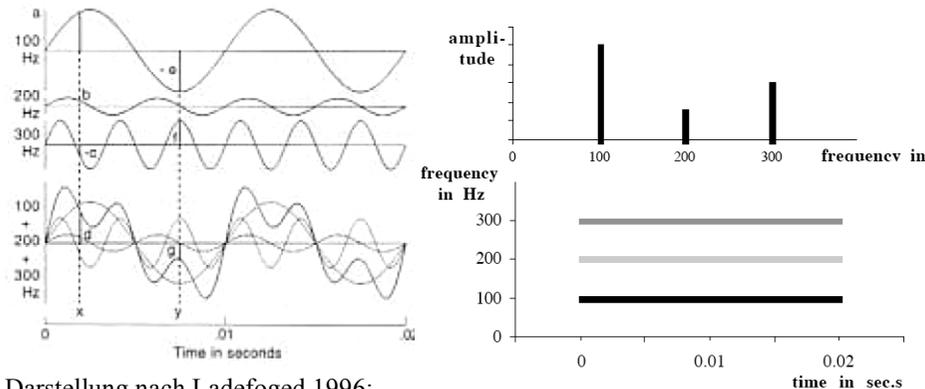
Rechts sieht man eine Anzeige der Frequenz in Hertz; wenn man den Cursor setzt, sieht man die genaue Anzeige der Frequenz an dieser Position in Hertz.

Die Stimmton-Analyse ist wichtig zur Darstellung der Intonation (z.B. Position und Art des Satzakkzents, des satzfinalen Akzents usw.; eine Standard-Darstellung hierfür ist die ToBI-Analyse, vgl. #).

Die hier dargestellte Stimmtonanalyse ist bereits geglättet und erfasst nicht alle stimmhaften Teile; die Parameter der Analyse können für den Forschungszweck verändert werden.

10.2.2 Spektrogramm

Unter einem Spektrogramm versteht man die Zerlegung einer Welle in Sinuskurven. Nach dem Satz von Fourier kann jede periodische Welle als eine Überlagerung von Sinuskurven dargestellt werden.



Darstellung nach Ladefoged 1996:

- Darstellung einer periodischen Schwingung durch Überlagerung von drei Sinuskurven;
- Darstellung durch das Spektrum (Amplitude auf der y-Achse, Frequenz auf der y-Achse),
- Darstellung durch ein Spektrogramm (Frequenz auf y-Achse, zeitlicher Verlauf auf 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.

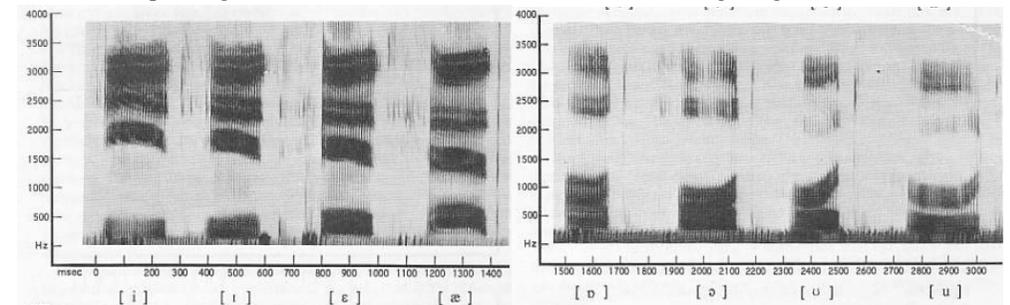
Formanten Vokale

Faustregel:

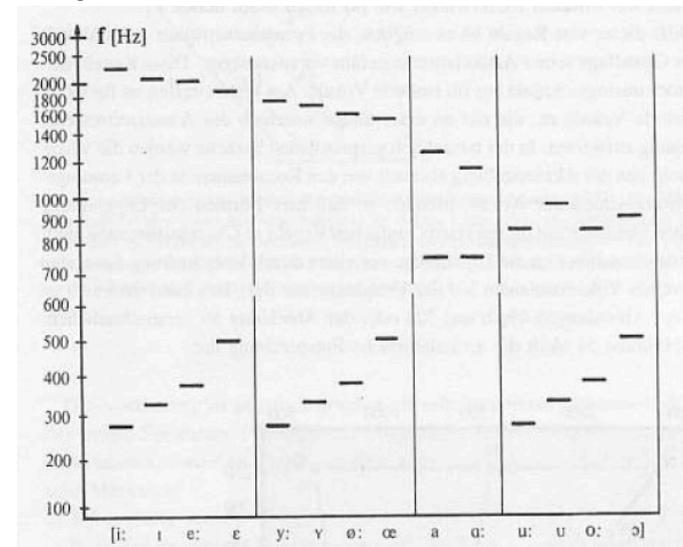
- Je höher die Zungenposition, desto tiefer der erste Formant.
- Je vorderer die Zungenposition, desto höher der zweite Formant.

Mit Praat kann man diese Regeln anhand von künstlich erzeugten Vokalen überprüfen (siehe NEW → SOUND → CREATE SOUND FROM VOWEL EDITOR).

Vokale im Spektrogramm; hier werden auch höhere Formanten angezeigt.



Beispiel: Formanten deutscher Vokale, nach 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.

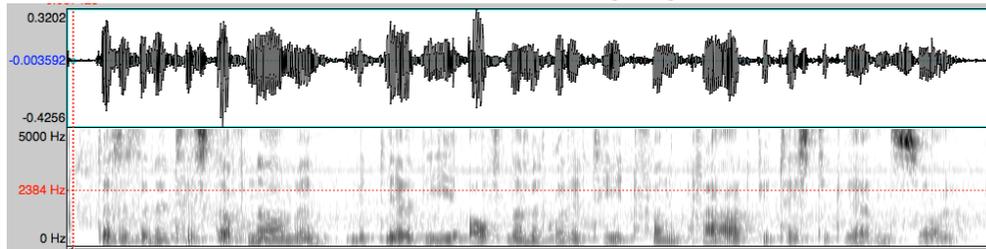
Nasale: Formanten ähnlich wie bei Vokalen, aber mit nasalen Formanten bei 250, 2500 und 3250 Hz.

Laterale: Formanten ähnlich wie bei vokalischen, aber mit Formanten um 250, 1200 und 2400 Hz; höhere Formanten sind reduziert.

Approximaten: Formanten ähnlich wie bei Vokalen, die sich aber über den Verlauf des Vokals verändern.

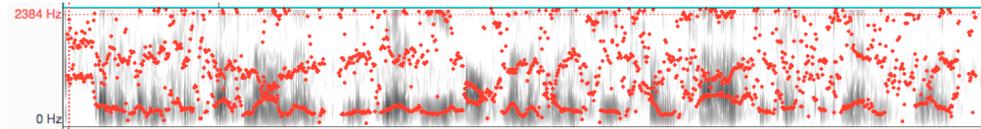
Spektrogramm im Praat

Dies kann über SPECTROGRAM → SHOW SPECTROGRAM gezeigt werden.

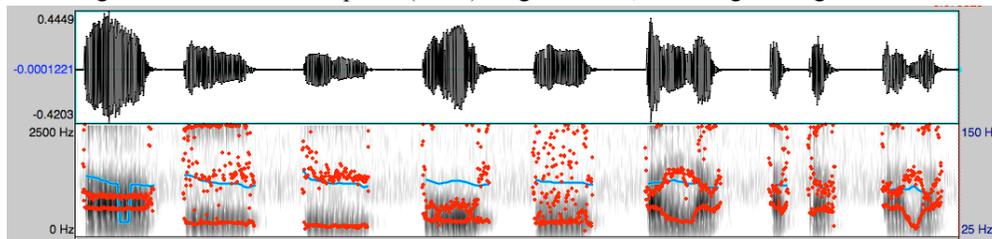


Standardmäßig wird das Spektrum bis 5000 Hz gezeigt, dies kann man jedoch auch verändern.

Man kann auch die Formanten anzeigen lassen (FORMANTS → SHOW FORMANTS); hier ist das mit dem Spektrum für den Bereich bis 2500 Hz durchgeführt. Die Formanten werden rot angezeigt.

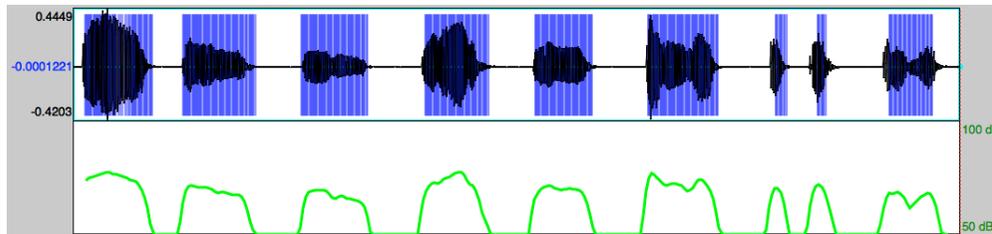


Hier eine Anzeige des Spektrogramms der fünf Vokale a, e, i, o, u, ferner [ala], [ata], [ada]. Es ist zugleich auch die F0-Frequenz (Pitch) eingezeichnet, die weitgehend gleich bleibt.



Intensität, Pulses

Weitere Anzeigemöglichkeiten: Die Intensität der Schallwelle und die “Pulses”, d.h. die Vibrationen der Stimmlippe.



10.3 Annotation

Praat enthält auch einen Partitur-Editor, der eine Annotation des akustischen Materials erlaubt.

10.3.1 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.

10.3.2 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.

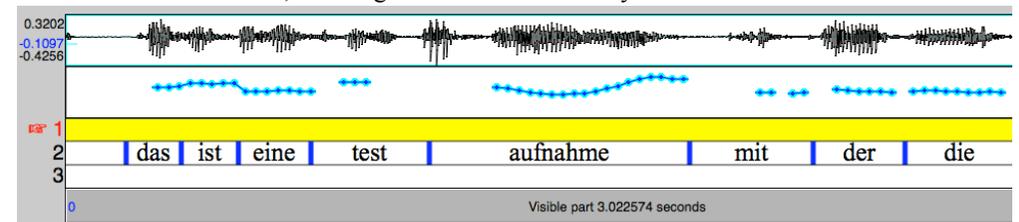
Grenzen können wieder entfernt werden: BOUNDARY → REMOVE.

Grenzen können verschoben werden: Klicken mit Taste und 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 gut an der Darstellung 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.

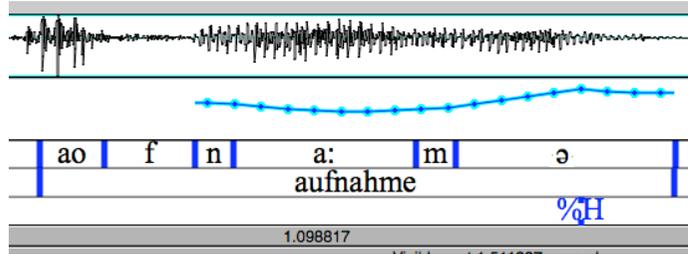


Die entstandene Annotationsdatei muss aus dem Objekts-Fenster heraus gespeichert werden: WRITE → WRITE TO SHORT TEXT FILE.

Wenn man auf ein Intervall klickt, sieht man die Länge dieses Intervalls, man kann also Längenmessungen von akustischen Einheiten vornehmen.

Annotation: Laute, Akzente

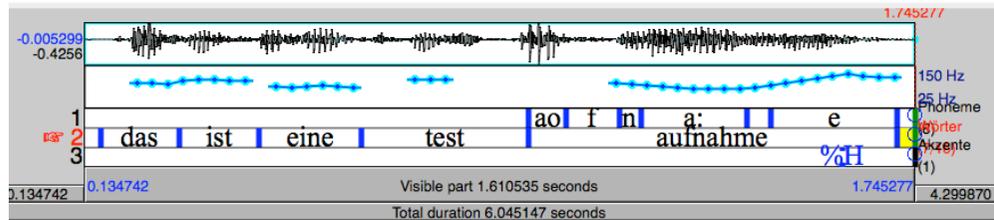
Wie Wörter, so kann man auch einzelne Phoneme annotieren. Auch die Annotation von Akzenten ist möglich (z.B. %H für hohen Grenzton); hierfür eignen sich punktuelle Annotationsebenen. Die Verwendung phonetischer Umschrift ist möglich



10.4 Ausdrucken und 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.

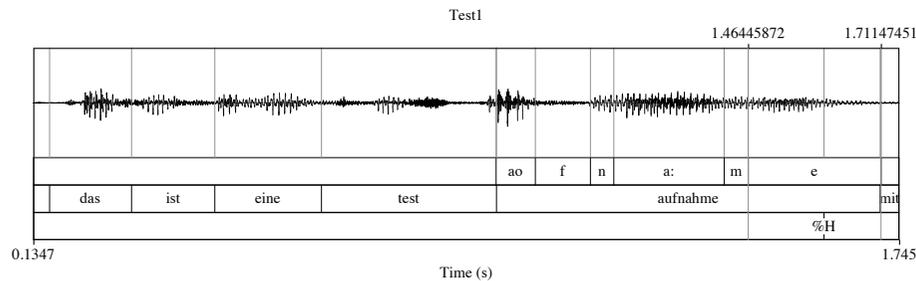
Beispiel: Markieren eines Abschnitts im Editor; VIEW → ZOOM TO SELECTION



Im Praat Picture Fenster wird ein Feld mit der intendierten Größe gezogen.

Im Editor-Fenster wird angewählt: EDIT → DRAW VISIBLE SOUND AND TEXT GRID

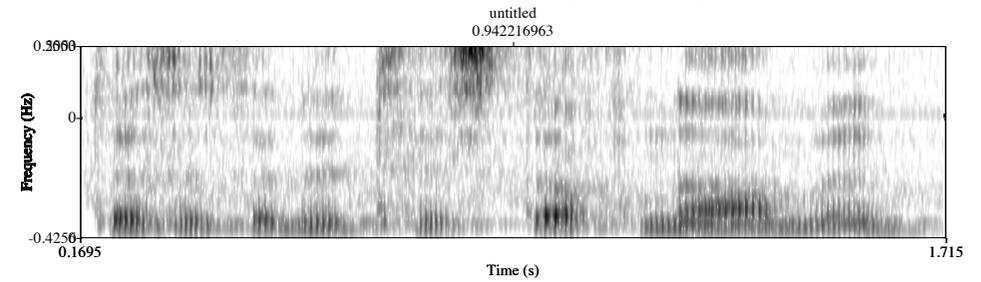
Man sieht dann die Schallwelle und die Transkription. Diese können z.B. über das Clipboard in eine Textdatei eingefügt werden:



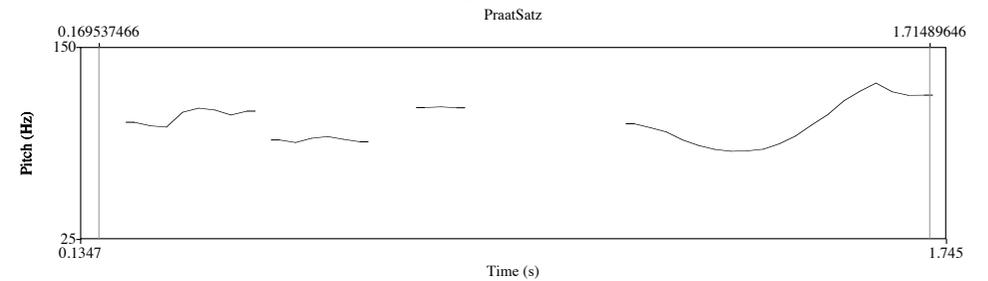
Wenn man auch Spektrogramme und Pitchverläufe darstellen will, muss man dies mit eigenen Fenstern bewerkstelligen.

Erster Schritt: In EDITOR-Fenster: SPECTROGRAM → EXTRACT VISIBLE SPECTROGRAM (hier muss man sicherstellen, dass das sichtbare Spektrogramm mit der Annotatin übereinstimmt).

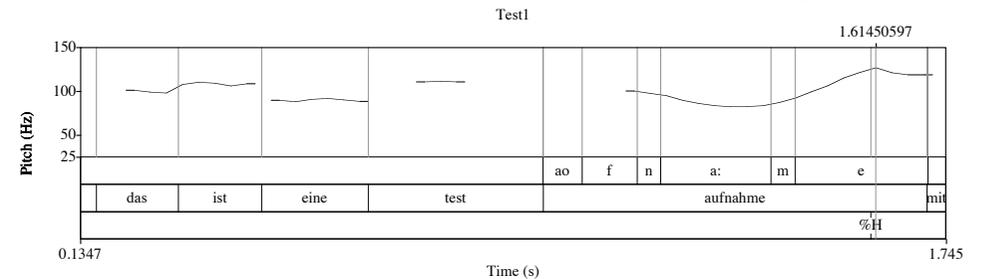
Zweiter Schritt: Anklicken des neu entstandenen Objekts im Objekt-Fenster; PAINT



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:



Weitere Information zur Verwendung von Praat findet man auf der Webseite.

11. Empirische Untersuchungen

Einge gute, knappe erste Einführung: R. Albert, C. J. Koster (2002), *Empirie in Linguistik und Sprachlehrforschung*, Narr.

11.1 Zugänge zu Daten: Beobachtung, Befragung, Experiment

Die Beobachtung

Wir haben bereits eine wichtige empirische Datenquelle für die Sprachwissenschaft kennengelernt, nämlich linguistische Korpora. Dies sind in der Regel tatsächliche sprachliche Äußerungen (sowohl schriftsprachliche als auch sprechsprachliche), die unabhängig von der wissenschaftlichen Fragestellung entstanden sind und die nach bestimmten Kriterien gesammelt werden. Es handelt sich hier im weitesten Sinne um Beobachtung.

Der Beobachter sollte so wenig wie möglich Einfluss auf das linguistische Material nehmen. Dies ist oftmals sehr gut möglich, z.B. bei einem Korpus von Zeitungstexten, oder bei Mitschnitten von Interviews im Radio. Manchmal ist eine Einflussnahme aber auch schwer zu vermeiden, z.B. bei Ton- und Filmaufnahmen, bei denen Sprecher sich gerade, weil sie sich beobachtet wissen, anders verhalten. Man spricht hier vom sog. **Beobachter-Paradox**. Dies gilt es zu minimieren, z.B., indem man die beobachteten Personen an den Aufnahme-prozess gewöhnt. Die verdeckte Aufnahme ist aus ethischen und rechtlichen Gründen nicht zulässig.

Die Befragung

Eine zweite Datenquelle sind 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 bestimmten Satzes hinsichtlich Herkunft oder sozialen Status einschätzen.

Ein Problem der Befragung ist, dass Personen – auch ungewollt – unehrliche Auskunft geben. Der Grund liegt darin, dass sie von sich ein sozial angemessenes Bild geben wollen.

- (1) *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.”

Ein Weg, dem zu begegnen, sind anonymisierte Befragungen; damit verliert man aber eventuell wichtige Daten über die befragte Person.

Ein zweites Problem ist, dass viele sprachliche Prozesse der bewussten Einsicht nicht zugänglich sind. Mit der Befragung erfassen wir immer das sekundäre Phänomen der Einsicht, nicht die Rohdaten der Sprachverarbeitung selbst. Dies kann stark auseinanderklaffen. Sprachwissenschaftliche Laien sind z.B. oft überzeugt, dass *Rad* und *Rat* unterschiedlich ausgesprochen werden.

Das Experiment

Im Experiment wird die Sprachverarbeitung selbst an vorbereitetem sprachlichen Material überprüft. Man kann das Material so auswählen oder konstruieren, dass die Reaktion von Versuchspersonen zwischen zwei Hypothesen unterscheidet: Wenn sich Versuchspersonen in der Weise X verhalten, ist das mit der Hypothese H besser verträglich als mit der Alternativhypothese H', also sollte Hypothese H angenommen und Hypothese H' verworfen werden. (Solche Hypothesenbildungen gibt es natürlich auch bei der Befragung. Bei der

Beobachtung gibt es sie ebenfalls, allerdings ist es oft nur schwer möglich, in der Masse der beobachteten Phänomene solche herauszufinden, die zwischen zwei Hypothesen entscheiden.

11.2 Gütekriterien

Empirische Untersuchungen sind in der Regel hypothesengeleitet (wobei sich manchmal die Hypothesen erst nach der Untersuchung ergeben – zur Problematik dieses Verfahrens im Statistik-Teil).

Eine entscheidende Frage ist: Wie gut stützt oder widerlegt eine empirische Untersuchung eine Hypothese? Es gibt hier unterschiedliche Gütekriterien.

11.2.1 Verlässlichkeit (Reliabilität)

Dieses Kriterium beschäftigt sich mit der Frage: Wie gut misst das Messverfahren dasjenige, was es misst?

Ein Messverfahren gilt als verlässlich, wenn es bei Wiederholungen im wesentlichen dasselbe Resultat liefert. Daraus ergeben sich auch die Möglichkeiten, die Verlässlichkeit zu testen:

- Testwiederholung. Derselbe Test wird noch einmal durchgeführt. Wenn man dies bei denselben Personen macht, muss allerdings sichergestellt werden, dass der erste Test den zweiten nicht beeinflusst (Lerneffekt).
- Paralleltest: Man untersucht dieselben Versuchspersonen mit einer zweiten Version des Texts. 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.

11.2.2 Geltungsbereich

Wenn wir eine bestimmte Gruppe von Personen untersuchen – z.B. wenn die Studenten einer Vorlesung an einem Experiment teilnehmen – stellt sich die Frage, inwieweit die Ergebnisse relevant sind für eine größere Gruppe. Eine ähnliche Frage stellt sich bei Korpusuntersuchungen: Wie aussagekräftig sind Beobachtungen an einem bestimmten Korpus für die Sprachverwendung allgemein? Oder: In einer Dialektuntersuchung werden Daten aus einigen wenigen Orten eines Gebietes gewonnen; die Frage ist: Kann man diese Daten für das Gebiet verallgemeinern?

Es geht hier um die Frage, inwieweit die Eigenschaften der gewählten Stichprobe die Eigenschaften insgesamt widerspiegeln. Die wichtigste Möglichkeit, dies zu überprüfen, besteht darin, verschiedene Stichproben zu ziehen und zu sehen, ob diese ähnliche Messerte liefern.

11.2.3 Gültigkeit (Validität)

Sprachliche Fähigkeiten sind in der Regel nicht direkt messbar. Wir messen Phänomene, die damit 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, und empirische Untersuchungen sind hier am ehesten angreifbar. Es hängt oft wiederum von theoretischen Voraussetzungen ab, ob die Resultate eines Messverfahrens mit den Phänomenen, die wir untersuchen wollen, tatsächlich in Verbindung stehen. Beispiel: Intelligenztests.

Ein Hinweis auf Gültigkeit ist, dass unterschiedliche Testverfahren für das zugrundeliegende Phänomen ähnliche Ergebnisse liefern.

11.3 Beobachtung

11.3.1 Grundsätzliches

Die Beobachtung von nicht elizitierten Daten ist besonders aufwändig:

- Wir wissen oft nicht, ob sich in den beobachtbaren Daten überhaupt hypothesen-relevante Daten finden, 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: Sprachwerbung) 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, Chats, veröffentlichte Texte.

11.3.2 Datenklassifikation

Ein Problem bei Beobachtungen ist die Klassifikation von Phänomenen. Beispiele: sprachbegleitende Gesten, Sprechakt-Typen in Konversationen, inhaltliche Beziehungen zwischen Textteilen. Hierbei kann es zu recht 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.

11.3.3 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.

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 superioren 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 Sprechakttyp 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.*

11.4 Befragung

11.4.1 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.

Es ist zu unterscheiden zwischen **mündlichen** und **schriftlichen** 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.

11.4.2 Die Stichprobe

Eine Stichprobe soll Aussagen über die Grundgesamtheit liefern. Damit sie dies tun kann, muss sie bestimmte Anforderungen erfüllen.

Auswahl

Hinsichtlich der Auswahl muss man zu vermeiden suchen, dass sie bestimmte Arten von Individuen der Grundgesamtheit überrepräsentiert. 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.

Größe

Wie groß soll eine Stichprobe sein, damit sie verlässliche Aussagen über die Grundgesamtheit liefert? Dies hängt von der Fragestellung und von den verwendeten statistischen Tests ab. Wenn die Daten relativ klar ausgeprägt sind (Beispiel: Verwendung des Präteritums vs. Perfekts als Erzähltempus im norddeutschen vs. süddeutschen Raum) genügt eine kleinere Stichprobe. Die notwendige Größe von Stichproben kann errechnet werden durch einen **Sample Size Calculator**, z.B.

Determine Sample Size

Confidence Level: 95% 99%

Confidence Interval:

Population:

Sample size needed:

<http://www.surveysystem.com/sscalc.htm>: 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 die Prozentzahl zu $\pm 5\%$ kennen (das **Konfidenzintervall**). Der Rechner sagt Ihnen: Sie brauchen ein Sample von 384 Personen (wenn wir eine Sprecherzahl von 100 Millionen annehmen, das spielt aber keine große Rolle). Die Personen des Samples sollten natürlich zufällig ausgewählt sein.

Wieder gilt: In Abschlussarbeiten wird selten (bei Befragungen) eine Stichprobengröße erreicht, mit der man statistisch signifikante Resultate erwarten kann.

11.4.3 Befragungsarten

Es ist zu unterscheiden zwischen

- Offenen Interviews: Der Interviewer kann flexibel reagieren. Phänomene, an die gar nicht gedacht wurden, können ins Blickfeld rücken. Nachteil: Kategorisierung der Auswertung. Diese Form eignet sich eher für qualitative als für quantitative Forschung.
- Festgelegte Fragen, oft in einem Fragebogen. Vorteil: Bessere Auswertmöglichkeit, Verminderung des Interviewer-Effekts.

Manchmal ist es günstig, erst offene Interviews durchzuführen und daraufhin einen Fragebogen zu entwickeln.

11.4.4 Fragebogen

Häufig bestehen Fragebögen aus Fragen oder Aussagen, zu denen Zustimmung oder Ablehnungsreaktionen verlangt werden. Hier kann man sich verschiedene Antwortarten vorstellen:

- Ja/Nein: Nur zwei Alternativen.
- Mehrere Alternativen:
 - Ja/Nein/Weiß nicht.
 - Bei Grammatikbeurteilung: richtig – möglich – seltsam – kaum möglich – falsch
 - Bei Häufigkeitsbeurteilungen: immer / häufig / gelegentlich / selten / nie
 - sehr einverstanden 1 2 3 4 5 6 7 überhaupt nicht einverstanden

Es hat sich bewährt, bei abgestuften Antworten fünf oder sieben Kategorien anzubieten. 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.

Es gibt weitere Fragebogentechniken, die in der Sprachwissenschaft eingesetzt werden, z.B. Satzkomplettierungsaufgaben oder Aufgaben, sprachliche Reaktionen in einem Kontext anzugeben. Beispiele:

- (1) *Peter zweifelt, /Peter bezweifelt, ...* (Frage: Komplementiererselektion *ob* vs. *dass*)
- (2) Sie haben sich von einem Freund ein (nicht besonders wertvolles) Buch ausgeliehen und können es nicht mehr finden. Sie wollen sich bei Ihrem Freund entschuldigen. Was sagen Sie?

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.*

11.4.5 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).

In aller Regel müssen Sie den Versuchspersonen zusichern, die Daten anonym zu behandeln.

11.5 Experimente

11.5.1 Grundsätzlich

Durch Experimente können Sie – wie bei Fragebögen – auf die Daten so Einfluss nehmen, dass sie für Ihre Hypothesen relevant sind. Anders als es in der Regel bei Fragebögen der Fall ist untersuchen Sie aber nicht die bewusste Einstellung (bzw. die Mitteilung dieser Einstellung) von Versuchspersonen, sondern deren unmittelbares Verhalten. Das unter-

suchbare Verhalten steht allerdings oft nur indirekt mit den sprachlichen Fähigkeiten in Verbindung, an denen Sie eigentlich interessiert sind.

11.5.2 Abhängige und unabhängige Variable

Die logische Struktur eines Experiments (und Fragebögen sind hier ein Sonderfall von Experimenten) ist die folgende:

- 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 nachzusprechen; Akzeptanz der Verlaufsform des Typs *Er ist den Tee am Trinken* und regionale Herkunft des Sprechers.
- Mindestens eine dieser Eigenschaftsausprägungen ist experimentell manipulierbar. Dies ist die **unabhängige Variable**. Diese wird im Experiment variiert.
Beispiel: Versuchspersonen werden zu unterschiedlichen Tageszeiten aufgefordert, Sätze nachzusprechen; Versuchspersonen aus unterschiedlichen Regionen Deutschlands 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 zusammengestellt. 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. Untersucht man z.B. bei der Verlaufsform nur Sprecher aus Köln und aus München, oder deckt man eine größere regionale Spannweite ab? Untersucht man bei dem Tageszeitexperiment zu jeder Stunde, oder nur am Vormittag, Nachmittag und Abend? Berücksichtigt man dabei den Schlaftyp der Versuchsperson, wie lang sie in der Nacht zuvor geschlafen hat, wann sie zuletzt etwas gegessen hat – und was?

Manchmal werden zwei oder mehr unabhängige Variablen untersucht. Z.B. kann man bei der Untersuchung der Verlaufsform sowohl die regionale Herkunft als auch das Alter der Versuchspersonen als unabhängige Variable wählen. Wir sprechen hier von einem **mehrfaktoriellen Design** des Experiments. In der Regel sind bei solchen Experimenten mehr Versuchspersonen nötig, um für die Grundgesamtheit gültige Aussagen zu treffen.

11.5.3 Labor- vs. Feldexperimente

Im **Laborexperiment** löst man ein Phänomen aus der natürlichen Umgebung heraus und untersucht es isoliert. Der Vorteil solcher Untersuchungen ist, dass man die unabhängige Variable leichter manipulieren kann und die Umstände, in denen das Experiment durchgeführt wird, besser kontrollieren und konstanthalten kann. Problem: Gerade dadurch kann man die Aussagekraft eines Experiments beeinträchtigen.

Im **Feldexperiment** untersucht man ein Phänomen in seiner natürlichen Umgebung. Zum Beispiel kann man an Wegbeschreibungen interessiert sein, sich als Tourist geben und Passanten nach dem Weg zu bestimmten Orten fragen. Die unabhängige Variable könnte dabei das Geschlecht, das geschätzte Alter oder die Schwierigkeit der Wegbeschreibung sein. 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.

11.5.4 Typen von Experimenten (wird noch weiter ausgeführt)

- **Verhaltensexperimente: Reaktionszeit.**
Beispiele: Priming-Experimente, Speeded Grammaticality Judgement, Self-Paced Reading
- **Verhaltensexperimente: Augenbewegung**
Beispiele: Beobachtung von Lesebewegungen, Beobachtung von Augenbewegungen beim Lösen von Aufgaben.
- **Physiologische Experimente: Hautwiderstand, Pupillometrie, Ereignis-korrelierte Potentiale (EEG/EKP), funktionale Magnetoresonanztomographie (fMRT).**

Lektüre: Albert & Koster S. 11-67.

12. Beschreibende Statistik

In diesem Abschnitt geht es um einige wesentliche Grundzüge der **deskriptiven Statistik**, d.h. um die Beschreibung und Darstellung von Daten.

12.1 Grundgesamtheiten und Stichproben

Wir beschreiben eine Menge von sog. **statistischen Einheiten**, z.B.

- (1) a. die Menge der Wörter in einem Text
- b. die Menge der Sätze in einem Text
- c. die Reaktionen von Sprechern einer Sprache auf eine Menge von Fragen
- d. die Menge der grammatisch wohlgeformten Sätze einer Sprache
- e. Eigenschaften von Grammatiken von Sprachen.

Die Menge der statistischen Einheiten nennt man **Grundgesamtheit** ("population"). Diese kann endlich oder unendlich sein (z.B. die Menge der wohlgeformten Sätze einer Sprache, die Menge der möglichen menschlichen Sprachen).

Oft können wir nicht alle statistischen Einheiten untersuchen, weil die Grundgesamtheit zu groß ist oder nicht alle statistischen Einheiten bekannt sind. Wir untersuchen daher Teilgesamtheiten, sog. **Stichproben**. Wenn diese nach bestimmten Prinzipien ausgewählt werden, kann man Rückschlüsse auf die Grundgesamtheit ziehen.

Beispiele:

- (2) Die Reaktionen von Sprechern des Deutschen in einem psycholinguistischen Experiment werden anhand einer Stichprobe von 50 Versuchspersonen abgeschätzt
- (3) Die durchschnittliche Silbenzahl von Wortvorkommnissen der deutschen Sprache wird anhand von Wörtvorkommnissen in einer Stichprobe von Texten festgestellt.

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 Rechtstexte, werden wir möglicherweise zu falschen Rückschlüssen kommen). Sie oben, Stichprobenrechner, und unten, #.

12.2 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:

- (4) a. Die Länge der Wörter (z.B. 1, 2, 3, 4, ... Grapheme, Phoneme, Silben, Morpheme)
- b. Die syntaktische Form der Sätze (z.B. Verb-Erst, Verb-Zweit, Verb-Letzt)
- c. Die Reaktionsarten von Sprechern auf Fragen (z.B. Ja/Nein/Weiss nicht)
- d. Die Reaktionszeiten von Antworten auf eine Frage (z.B. 100ms, 200ms, 300ms, ...)
- e. 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 werden von bestimmten beobachtbaren Größen bestimmt, den **Einflussgrößen** oder **Faktoren**. Beispiele:

- (5) a. Textsorte
- b. Entstehungszeit des Textes
- c. Alter der Sprecher
- d. Sprachfamilie, zu der eine Sprache gehört

Daneben muss man auch mit dem Einfluss von nicht beobachtbaren Größen rechnen, den **Störgrößen**.

12.3 Merkmalstypen: Diskret und Stetig

Die Ausprägungen von Merkmalen können von verschiedenen Typen sein, was wesentlich ihre Darstellung bestimmt.

12.3.1 Diskrete Merkmale

haben klar voneinander unterscheidbare Ausprägungen. Beispiele:

- (6) a. Reaktionen auf eine Frage in einem Experiment (Ja/Nein/Weiss nicht)
- b. Grundwortstellung einer Sprache (SVO, SOV, VSO, OVS, OSV, VOS)
- c. Syntaktische Form der Sätze des Deutschen (V1, V2, VL)
- d. Zahl der Phoneme in den Wörtern eines Korpus (1, 2, 3, 4, ...)

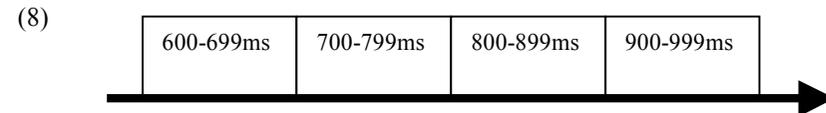
Bermerke: Es gibt unendlich viele mögliche Ausprägungen im letzteren Fall, dennoch ist das Merkmal diskret. (Man sagt, es gibt "abzählbar unendlich" viele Ausprägungen).

12.3.2 Stetige Merkmale

Hier können die Ausprägungen beliebig nah beieinander liegen. Beispiele:

- (7) a. Reaktionszeiten auf eine Frage (800ms, 801ms, 800,5ms, 800,7ms usw.)
- b. 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.



12.4 Merkmalstypen: Nominal-, Ordinal- und Kardinalskalen

Darüber hinaus unterscheiden wir Merkmale danach, welche Struktur ihren Ausprägungen unterliegt – die sog **Skalentypen**.

12.4.1 Nominalskalen

Bei Nominalskalen gibt es keine bestimmte Ordnung zwischen den einzelnen Ausprägungen.

- (9) a. Grundwortstellung einer Sprache (SVO, SOV, VSO, OVS, OSV, VOS)
- b. Syntaktische Form der Sätze des Deutschen (V1, V2, VL)

12.4.2 Ordinalskalen

Es gibt eine natürliche Anordnung zwischen den Ausprägungen, allerdings ist es nicht sinnvoll, von bestimmten Entfernungen von Ausprägungen zu sprechen.

- (10) a. Reaktionen auf eine Frage in einem Experiment (Ja – Weiss nicht – Nein)
- b. Einschätzung der Akzeptabilität einer Wortkette durch Sprecher einer Sprache: (5-Punkte-Skala: Normal – Akzeptabel – Grenzwertig – Schlecht – Sehr schlecht)

Bei Nominalskalen und Ordinalskalen mit endlich vielen Ausprägungen sprechen wir von **qualitativen** Skalen.

12.4.3 Kardinalskalen

Die Ausprägungen sind natürlich angeordnet, und man kann von verschiedenen Entfernungen sprechen. Wir sprechen von **quantitativen** Skalen.

- Zahl der Phoneme in den Wörtern einer Sprache (1, 2, 3, ...)
- Reaktionszeiten auf Fragen in einem Experiment (800ms, 900ms, 1000ms, ...)

12.5 Statistikprogramme

Je nach der Art der Ausprägungen von Merkmalen können diese verschiedenartig grafisch dargestellt werden. Dies erlaubt eine bessere Einschätzung als z.B. die Präsentation von Daten in einer Tabelle.

Es gibt verschiedene Hilfsmittel, um Daten darzustellen:

12.5.1 Tabellenkalkulationsprogramme

Das am weitesten verbreitete Tabellenkalkulationsprogramm ist Microsoft Excel; es ist ein Teil von Microsoft Office. Excel bietet eine Reihe von Statistikfunktionen und auch Möglichkeiten der grafischen Darstellung statistischer Information.

12.5.2 Statistikprogramme: R

Das Programm R ist ein frei verfügbares, plattform-unabhängiges Statistikpaket, das wesentlich mehr erlaubt als Excel. Man kann sich R von der Webseite <http://www.r-project.org/> auf Ihren Computer und installieren sie dieses. Vergessen Sie nicht, auch das Manual *An Introduction to R* mit herunterzuladen.

12.5.3 Statistikprogramme: SPSS

Dies ist das Standard-Statistikprogramm. Ein kommerzielles Programm in vielen Varianten, das für Studenten (mit zeitlich beschränkter Lizenz) preisgünstig angeboten wird, für andere Anwender aber sehr teuer ist.

Es gibt natürlich noch zahlreiche weitere Statistikprogramme und auch Online-Ressourcen.

12.6 Die Darstellung von qualitativen (nominalen) Daten

Nach dem *World Atlas of Language Structure*, siehe <http://wals.info/>, verhält es sich mit den Häufigkeiten der Grundwortstellung von Subjekt, Verb und Objekt in transitiven Sätzen so:

Stellung	SOV	SVO	VSO	VOS	OVS	OSV	keine dominante St.
Anzahl Sprachen	497	435	85	26	9	4	172

Es handelt sich hierbei um eine Nominalskala. Wir können solche Daten gut in einem **Säulendiagramm** (Balkendiagramm, “bar chart”) darstellen.

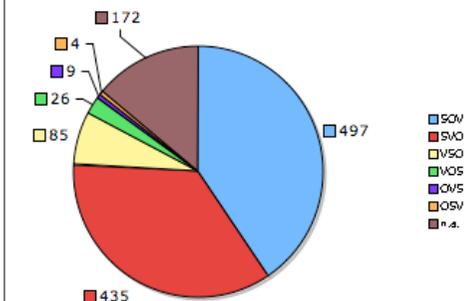
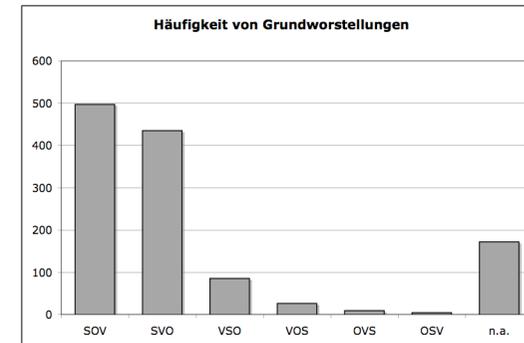
12.6.1 Darstellung in Microsoft Excel

Wir tragen die Daten in einer Zeile eines Excel-Arbeitsblattes ein:

	A	B	C	D	E	F	G	H
1	SOV	SVO	VSO	VOS	OVS	OSV	n.a.	
2		497	435	85	26	9	4	172
3								
4								

Wir markieren die Daten und drücken auf die Schaltfläche für Grafik, wobei wir viele mögliche Darstellungen zur Auswahl haben. Die meisten angebotenen Darstellungen sind leider unseriös und irreführend. Vor allem sollte man dreidimensionale Darstellungen nicht verwenden.

Für unsere Daten bietet sich ein Säulendiagramm, Balkendiagramm oder ein Kreisdiagramm (“Kuchendiagramm”) an.



Wichtig: Für die Einfügung in Artikel müssen Sie die Grafiken unbedingt bearbeiten, da die vorgegebenen Grundeinstellungen nicht optimal sind. Es gibt hier viele Optionen.

12.6.2 Darstellung in R

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 = der Variablen Grundwortstellung zugeordnet. Das Kompositionszeichen ist c, wir führen hier jeweils die Ausprägung des Merkmals und seine Häufigkeit an. NA steht für “not applicable”; die Sprachen ohne dominante Grundwortstellung.

Den Inhalt dieses Vektors kann man sich anschauen:

```
> Grundwortstellung
```

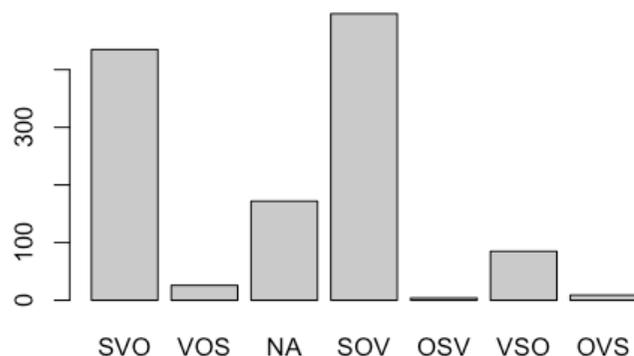
Wir erhalten eine tabellarische Darstellung der Daten:

```
SVO VOS NA SOV OSV VSO OVS
435 26 172 497 4 85 9
```

Wir können diese Daten mithilfe eines Säulendiagramm mithilfe der Funktion `barplot` darstellen:

```
> barplot(Grundwortstellung)
```

In einem eigenen Fenster wird folgendes Diagramm angezeigt:



Wir wollen nun zur besseren Darstellung die Daten nach absteigender Größe sortieren. Dies können wir, da wir es ja mit einer Nominalskala zu tun haben. Das heißt, die Reihenfolge, in der die Grundwortstellungstypen aufgeführt werden, ist inhaltlich belanglos Ein Versuch:

```
> sort(Grundwortstellung)
```

```
OSV OVS VOS VSO NA SVO SOV
4 9 26 85 172 435 497
```

Das ist offensichtlich fehlgeschlagen, die Sortier-Reihenfolge ist aufsteigend. Wir können uns die Optionen der Funktion `sort` näher anschauen:

```
> ?sort
```

Wir werden über ein Hilfe-Fenster darüber informiert, dass wir eingeben müssen:

```
> sort(Grundwortstellung, decreasing=TRUE)
```

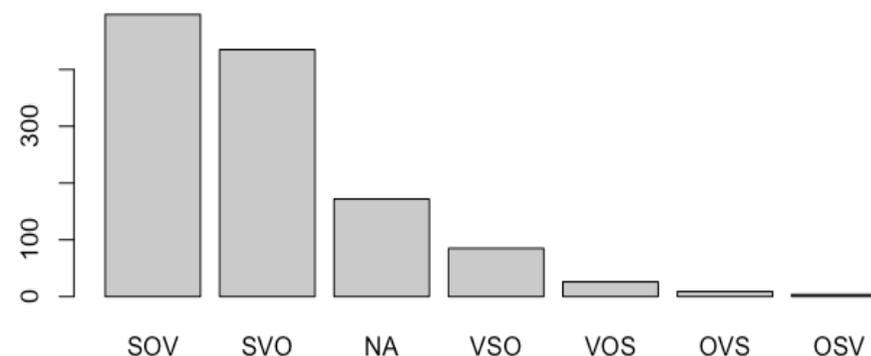
```
SOV SVO NA VSO VOS OVS OSV
497 435 172 85 26 9 4
```

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)
```



Es gibt viele verschiedene Möglichkeiten, diese Darstellung zu modifizieren (durch Beschriftung der Achsen, verschiedene Farben für die Kolumnen usw.); man erhält hierüber über `?barplot` Auskunft.

Bitte führen Sie die R-Befehle wie beschrieben durch. Bitte 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.)

Mit R gibt es ebenfalls verschiedene Möglichkeiten der grafischen Darstellung.

12.7 Die Darstellung von quantitativen (intervallskalierten) 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.*).

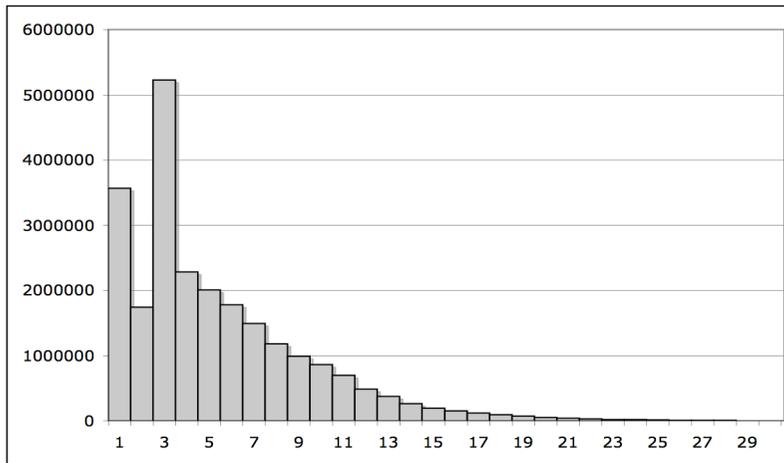
1	3564430	11	693824	21	35344
2	1742674	12	484361	22	26020
3	5226684	13	373814	23	17548
4	2282737	14	257697	24	13768
5	2004860	15	192147	25	9134
6	1775106	16	147870	26	5727
7	1489365	17	115291	27	3928
8	1175359	18	90033	28	2717
9	985117	19	67285	29	1744
10	861835	20	50051	30	1266

12.7.1 Darstellung mithilfe von Excel

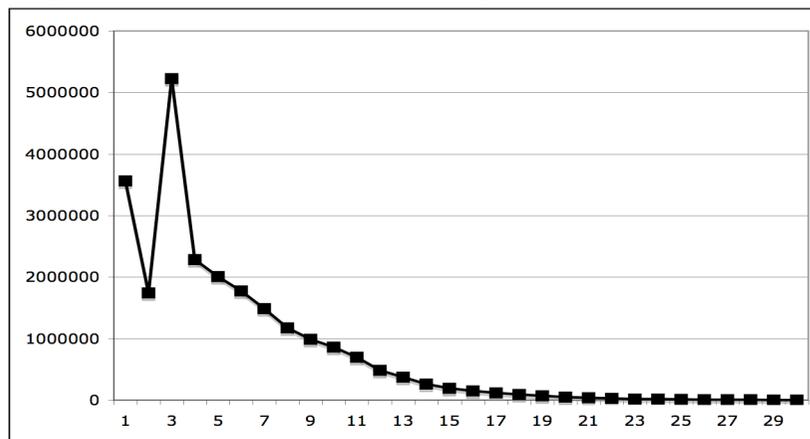
Wir geben wiederum die Daten in ein Arbeitsblatt ein:

	A	B	C	D	E	F	G	H	I
1	1	2	3	4	5	6	7	8	9
2	3564430	1742674	5226684	2282737	2004860	1775106	1489365	1175359	985117
3									

Wir können daraus wieder ein Säulendiagramm gewinnen. Da es sich bei den Merkmalsausprägungen (Wortlänge) jetzt um quantitative Daten handelt, werden die Säulen ohne Zwischenraum dargestellt (das muss man eigens festlegen). Man nennt solche Diagramme auch **Histogramme**. Genauer sind H. allerdings Darstellungen, in denen die Fläche, nicht die Höhe jedes Balkens die Größe der Daten angibt; bei gleich breiten Säulen sind allerdings Höhe und Fläche gleichwertig.



Bei quantitativen Daten sind auch **Liniendiagramme** sinnvoll:



12.7.2 Darstellung mithilfe von R

Wir können diese Daten am R-Prompt eingeben. Mit der folgenden Eingabe 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) einfach `\n` eingeben.

Die hier verwendete `scan`-Funktion kann auch dazu verwendet werden, um Daten aus einer Datei einzulesen. Wir geben mit einem einfachen Texteditor die folgende Reihe von Zahlen in einer Zeile ein und speichern sie unter dem Namen `tokenlength` ab.

```
3564430; 1742674; ... 1744; 1266
```

Wir können nun mit einem der beiden Befehle diese Datei einlesen:

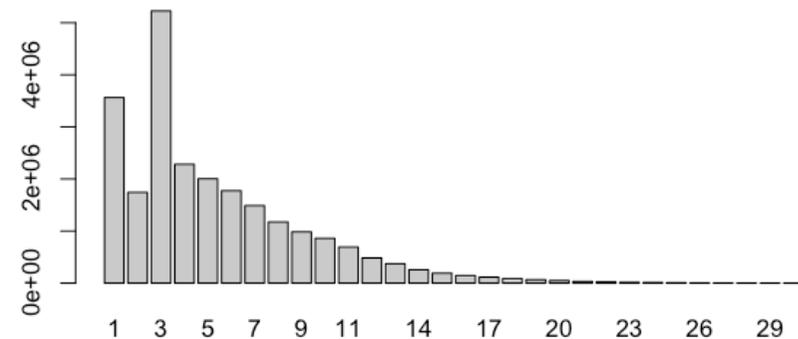
```
> t1 = scan("Dateipfad+Dateiname", sep=";")
> t1 = scan(file.choose(), sep=";")\nt
```

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.

Wir können den eben auf die Variable `t1` eingelesenen Vektor natürlich mit einem Säulendiagramm darstellen. Dies bewerkstelligt der folgende Befehl, wobei die Angabe `names.arg=1:30` besagt, dass als Säulennamen die Zahlen von 1 bis 30 verwendet werden sollen. (Genauer gesagt bewirkt der Ausdruck `1:30` die Konstruktion eines Vektors, der dann als Vektor von Namen für die Säulen verwendet wird).

```
> barplot(t1, names.arg=1:30)
```

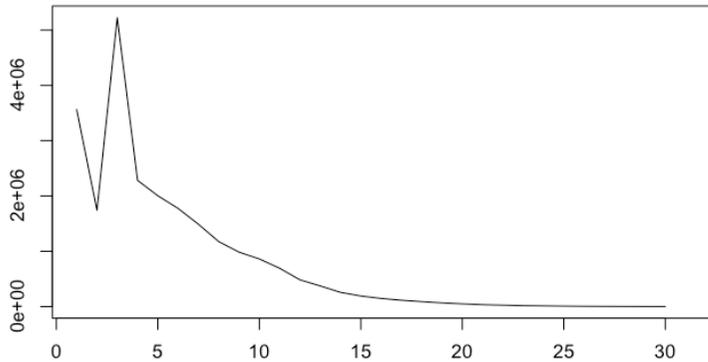
Dies erzeugt das folgende Säulendiagramm. Die Anzahlen sind hier in der Exponentenschreibweise angegeben ($4e+06$ steht für 4 Millionen, also $4 \cdot 10^6$).



Ein Liniendiagramm kann wie folgt erzeugt werden:

```
> plot(w1, type="l")
```

Für einige weitere Diagramm-Möglichkeiten siehe `?plot`.



12.8 Mittelwert, Streuung, Standardabweichung

Nehmen wir an, wir testen die Behaltungsleistung von Vokabeln. Wir haben dreißig Versuchspersonen, die jeweils zwanzig Vokabeln lernen und nach einer Stunde abgefragt werden. Ergebnisse (tabellarische Darstellung, Beispiel für Excel-Eintrag):

(11) Verhaltensleistung von gelernten Vokabeln

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

Wir können die Verteilung nun auf verschiedene Weise charakterisieren:

12.8.1 Das arithmetische Mittel (Mittelwert, eng. Average)

Auch arithmetischer Mittelwert (μ) genannt; dies ist die Summe aller behaltenen 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: $\mu_X = \Sigma X / n$.

Wenn man die Daten wie angegeben in ein Excel-Arbeitsblatt eingegeben hat, kann man sie leicht berechnen. Im folgenden nehmen wir an, dass die Zahl der behaltenen Vokabeln in den Feldern B2 bis B31 angezeigt ist.

(12) a. = SUMME(B2..B31)

b. = MITTELWERT(B2..B31)

12.8.2 Der Median (Zentralwert) und Quartile

Dies ist der Wert, für den gilt: die eine Hälfte der Versuchspersonen hat Werte erzielt, die gleich sind oder darunter liegen, die andere hat Werte erzielt, die gleich sind oder darüber liegen. In unserem Beispiel ist der Median 12,5; d.h. 15 Werte liegen über und 15 Werte unter 12,5. In Excel können wir den Median berechnen mit:

(13) = MEDIAN(B2..B31)

Der Median lässt sich auch auf Ordinaldaten anwenden und ist weniger beeinflussbar durch Extremwerte ("Ausreißer"). Damit verwandt sind die Quartile: Das 25%-Quartil ist derjenige Datenpunkt, sodass gilt: Unerhalb dieses liegen 25% der Daten. Entsprechend das 75%-Quartil. Das 50%-Quartil ist der Median selbst.

12.8.3 Der Modalwert

Damit bezeichnet man den häufigsten Wert; in unserem Beispiel ist das 14. Er kann im Excel mit MODALWERT berechnet werden.

12.8.4 Die Streuung

Darunter versteht man die Differenz zwischen dem höchsten Wert (hier 19) und dem niedrigsten (hier 3), in unserem Beispiel also 16. In Excel berechenbar mit:

(14) = MAX(B2..B31) – MIN(B2..B31)

12.8.5 Die Standardabweichung

Die Streuung gibt nur ein grobes Maß darüber, wie dicht die Werte beieinander liegen; ein einzelner "Ausreißer" wie die Versuchsperson 3 affiziert die Streuung gewaltig. 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.

Mit Excel kann man die Standardabweichung direkt berechnen:

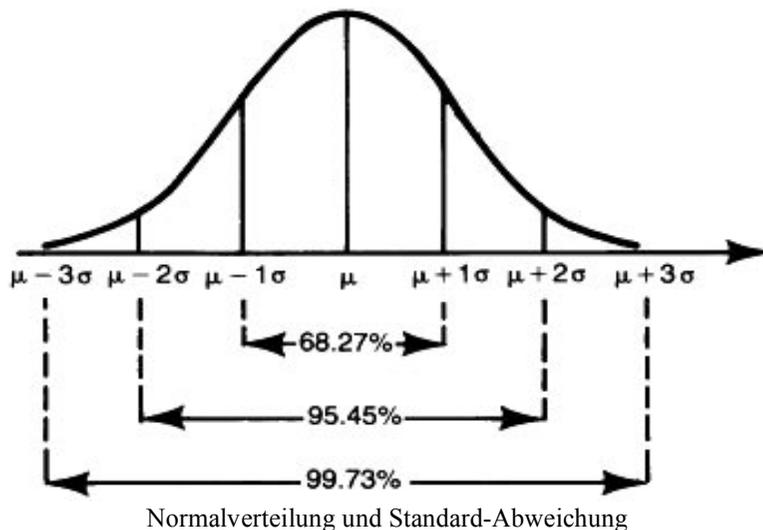
(15) = STABW(B2..B31)

	I	J	K
Vokabeln	Abw.	Mittwert	Quadrat
9	3,7	13,69	
12	0,7	0,49	
19	-6,3	39,69	
13	-0,3	0,09	
16	-3,3	10,89	
14	-1,3	1,69	
10	2,7	7,29	
11	1,7	2,89	
12	0,7	0,49	
12	0,7	0,49	
14	-1,3	1,69	
17	-4,3	18,49	
12	0,7	0,49	
15	-2,3	5,29	
13	-0,3	0,09	
17	-4,3	18,49	
14	-1,3	1,69	
11	1,7	2,89	
16	-3,3	10,89	
16	-3,3	10,89	
14	-1,3	1,69	
14	-1,3	1,69	
9	3,7	13,69	
10	2,7	7,29	
12	0,7	0,49	
3	9,7	94,09	
10	2,7	7,29	
14	-1,3	1,69	
11	1,7	2,89	
11	1,7	2,89	
Summe		282,3	
/ (N - 1)		9,73448	
Wurzel		3,12001	

Ein hoher Wert deutet an: Die Werte streuen stark.

Standardabweichung und Normalverteilung

Viele Zufallsgrößen sind normalverteilt (d.h. sie folgen der Gauß-Kurve). Wenn das so ist, dann ist die Standardabweichung so zu interpretieren: Ca. 68% der Werte liegen innerhalb der Standardabweichung, in unserem Beispiel: 68% der Werte liegen innerhalb von 12,7 \pm 3,12.



12.8.6 Die Varianz

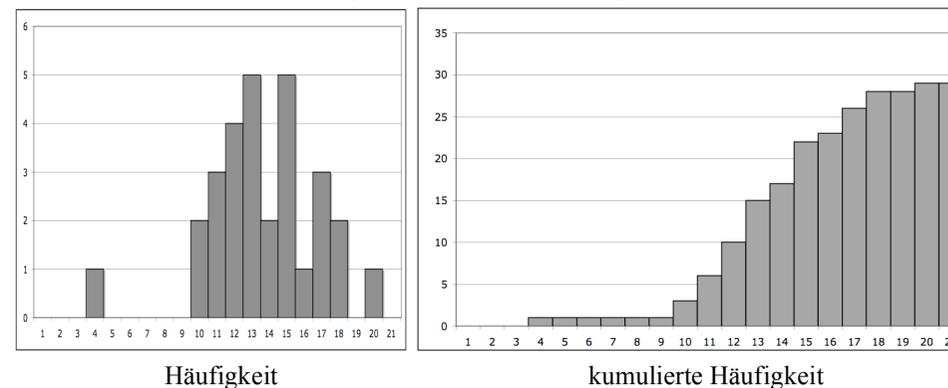
ist das Quadrat der Standardabweichung.

12.9 Häufigkeitsdiagramm

Die Daten von (11) kann man wie folgt ordnen und daraus ein **Häufigkeitsdiagramm** erstellen, in welchem festgehalten wird, wie oft eine bestimmte Anzahl von Vokabeln gelernt wurde. Wir unterscheiden zwischen einfacher und kumulativer Häufigkeit.

Vokabeln	Häufigkeit	Kum. H.	Vokabeln	Häufigkeit	Kum. H.
0	0	0	11	4	10
1	0	0	12	5	15
2	0	0	13	2	17
3	1	1	14	5	22
4	0	1	15	1	23
5	0	1	16	3	26
6	0	1	17	2	28
7	0	1	18	0	28
8	0	1	19	1	29
9	2	3	20	0	29
10	3	6			

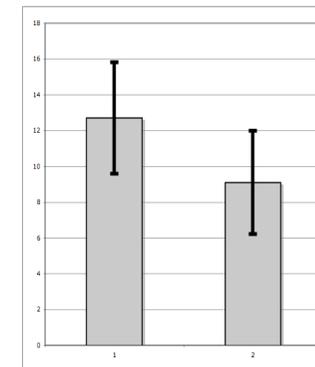
Diese Häufigkeiten können in einem Säulendiagramm dargestellt werden; bei der Häufigkeit finden wir hier eine Annäherung an die Normalverteilung.



12.10 Angabe der Standardabweichung in Säulendiagrammen

Nehmen wir an, die erste Vokabeluntersuchung geschah mit konkreten Wörtern. Eine zweite mit abstrakten Wörtern führte zu dem folgenden Resultat:

VP	Vokabeln	VP	Vokabeln	VP	Vokabeln
1	5	11	9	21	9
2	8	12	15	22	8
3	16	13	7	23	6
4	9	14	16	24	8
5	10	15	9	25	9
6	11	16	13	26	4
7	7	17	8	27	7
8	8	18	7	28	11
9	8	19	11	29	8
10	7	20	10	30	9



Wir erhalten nun als Mittelwert 9,1 und als Standard-Abweichung 2,88. Die beiden Daten können in einen Säulendiagramm verglichen werden, wobei die Standardabweichung durch einen Strich angegeben wird.

12.10.1 Darstellung In Excel

- Berechne die Mittelwerte und die Standardabweichung im Arbeitsblatt.
- Stelle die Mittelwerte in einem Säulendiagramm dar.
- Auf eine Säule klicken, dann auf "Fehlerindikator" klicken, auf "Benutzerdefiniert" klicken, die Standardabweichungen im Arbeitsblatt markieren und klicken.

Nun erscheinen die Standardabweichungen über den Säulen.

12.10.2 Darstellung in R: Boxplot-Diagramm

Wir lesen zunächst die zwei Datensätze (insgesamt 60 Daten) ein, wobei die einzelnen Werte durch Semikolon separiert sind. Dann wird diese Reihe durch den dim-Befehl als ein Array von zwei mal dreißig Elementen umformatiert. Der Anfang dieses Arrays wird gezeigt.

```
> vokabeln = scan(file.choose(), sep=";")
Read 60 items
> dim(vokabeln) = c(30,2)
> vokabeln
      [,1] [,2]
[1,]    9    5
[2,]   12    8
[3,]   19   16
[4,]   13    9
[5,]   16   10
-- -- -- --
```

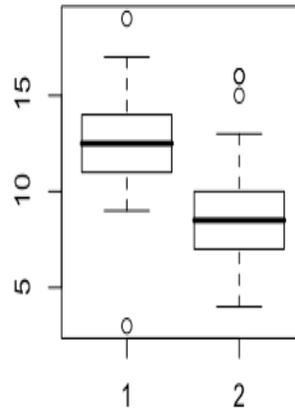
Auf die erste Datenspalte kann man sich mit `vokabeln[, 1]` beziehen, auf die zweite mit `vokabeln[, 2]`.

Eine moderne, sehr informative Darstellung der Daten sind sog. **Boxplot-Diagramme**. Sie können wie folgt erzeugt werden:

```
> 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.



12.11 Korrelationen

12.11.1 Darstellung von Korrelationen in Diagrammen

Wir stellen uns die Frage: Gibt es eine Korrelation zwischen der Erlernbarkeit der beiden Klassen von Vokabeln? Lernen Versuchspersonen, die bei der einen Klasse von Vokabeln gut sind, auch die andere gut?

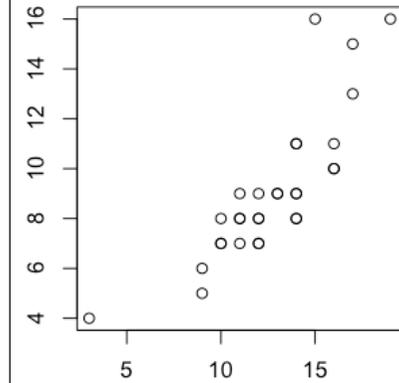
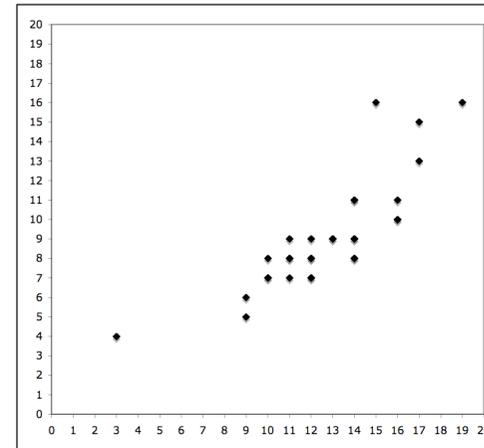
die Frage, ob es **Korrelationen** zwischen den beiden Datensätzen gibt. Dies kann man zunächst abschätzen, indem man die Zusammenhänge grafisch darstellt.

Mit R genügt hier der einfache Befehl:

```
> plot(vokabeln)
```

um das linksstehende Diagramm zu erzeugen, in welchem für jede Versuchsperson die Leistung in beiden Tests abgetragen ist. Offensichtlich besteht ein starker positiver Zusammenhang: Wenn jemand gut bei Vokabeln I ist, ist er auch gut bei Vokabeln II (die vertikale Achse).

Mit Excel kann man ein solches Punktediagramm erzeugen, indem man die zwei Datenreihen markiert, auf die Schaltfläche für Grafiken drückt und den Diagrammtyp Punkt(X,Y) wählt.



Dies ist

Bei **positiver Korrelation** (wie hier) kann man die Verteilung durch eine ansteigende Gerade annähern, bei **negativer Korrelation** durch eine fallende, bei Fehlen einer Korrelation gibt es keine solche Annäherung. Man spricht hier von **linearer Korrelation**.

12.11.2 Kovarianz

Wir wollen ein Maß für die Korrelation von zwei Datensätzen angeben. Ein erster Schritt hierfür ist die **Kovarianz**. Dies ist ein Maß dafür, wie stark sich zwei Datensätze parallel verändern. Die Kovarianz ist hoch, wenn gilt: Wenn ein Wert des einen Datensatzes über dem Mittelwert liegt, liegt der Wert des anderen in möglichst gleichem Maße ebenfalls darüber, und umgekehrt.

Für ein bestimmtes Element des Datensatzes (z.B. dem Resultat einer Versuchsperson bei den beiden Datentests) gehen wir so vor:

$$(16) (x_i - \mu_X)(y_i - \mu_Y)$$

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:

- Wenn x_i und y_i von den Mittelwerten in die gleiche Richtung abweichen, ergibt sich ein positiver Wert.
- Wenn x_i und y_i in unterschiedliche Richtungen abweichen, ergibt sich ein 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).

$$(17) \text{COV}(X,Y) = \frac{\sum (X - \mu_X)(Y - \mu_Y)}{n - 1}$$

Die Kovarianz in unserem Beispiel beträgt 7,51379. Man kann die Kovarianz zwischen zwei Datensätzen auch direkt berechnen durch die Excel-Formel `KOVAR(X; Y)`, wobei X und Y die Angabe von zwei Datensätzen sind, z.B. `KOVAR(O2:O31; P2P31)`.

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.

12.11.3 Der Korrelationskoeffizient

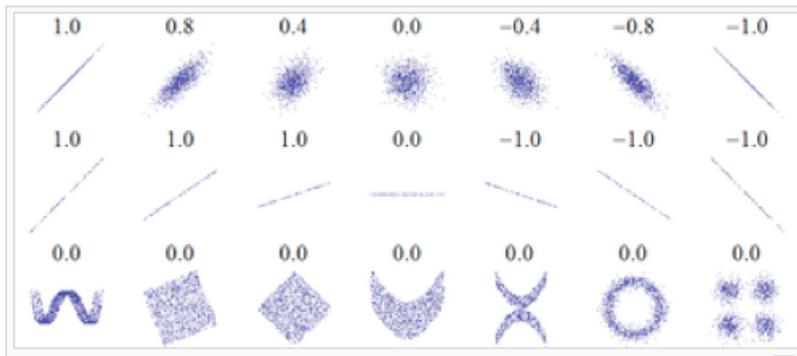
Abhilfe schafft hier der **Korrelationskoeffizient** (hier: der PM-Korrelationskoeffizient, die *product-moment correlation* nach Pearson. Hierzu teilen wir die Kovarianz durch das Produkt der Standardabweichung von X und der Standardabweichung von Y:

$$(18) \quad \rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

In unserem Beispiel erhalten wir den Wert 0,836. Dies ist ein dimensionsloser Wert, d.h. er bleibt auch gleich, wenn wir eine richtig beantwortete Fragen mit 10 Punkten messen. Er hat die folgenden Eigenschaften:

- $\rho_{X,Y} = 1$: stärkstmögliche positive lineare Korrelation
- $\rho_{X,Y} = -1$: stärkstmögliche negative lineare Korrelation
- $\rho_{X,Y} = 0$: keine lineare Korrelation

Beispiele für Punktediagramme ("scatterplots") und zugehörige Korrelationskoeffizienten. Das mittlere Beispiel hat hier den Koeffizienten 0, der Koeffizient ist aber eigentlich nicht definiert (Division durch 0). Die untere Zeile zeigt intuitiv klar bestehende Relationen auf, die aber nicht linear sind und daher vom Pearson'schen Korrelationskoeffizienten nicht erfasst werden.



12.11.4 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 sagt, dass die Veränderung einer Größe die Veränderung einer zweiten nach sich zieht und dafür auch ein Grund angibt, dann ist dies ein Argument für diese Theorie und macht die Theorie wahrscheinlicher.

12.11.5 Erklärte Varianz

Die Höhe des Korrelationskoeffizienten gibt an, wie viel der Varianz der abhängigen Variablen durch die unabhängige Variable erklärt werden kann.

(19) ρ^2 ist das Maß, in dem die Wahl der unabhängigen Variable die Wahl der abhängigen bestimmt.

In unserem Beispiel erhielten wir den Wert 0,836. Nach Quadrierung erhalten wir den angenäherten Wert 0.7. Wir können also ca. 70% des Verhaltens einer Versuchsperson in einem Test voraussagen, wenn wir die Versuchsperson und ihre Eigenschaften kennen.

13. Schließende Statistik

Mit der Korrelation haben wir uns von der rein beschreibenden Statistik der schließenden (inferentiellen) Statistik genähert. Die beschreibende Statistik sagt nur etwas über die erhobenen Daten (die Stichprobe) aus. Die schließende Statistik überprüft, ob diese erhobenen Daten eventuell nur dem Zufall der Auswahl zuzuschreiben sind, oder ob sie Rückschlüsse auf die Verhältnisse in der Grundgesamtheit erlauben.

13.1 Der Chi-Quadrat-Test

Dieser Test eignet sich für Nominaldaten, in denen die Merkmalsausprägungen keiner bestimmten Ordnung unterliegen.

13.1.1 Das Testverfahren

Ein Beispiel: In dem Sample von 30 mehr oder weniger zufällig ausgewählten Sprachen 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), wobei hier jeweils noch die Gesamtsummen von Grundwortstellungstypen und Adpositionsstellungstypen angegeben sind.

(1)

	SOV	SVO	VSO	
P-NP	0	10	6	16
NP-P	11	3	0	14
	11	13	6	30

Es scheint eine klare Tendenz zu geben: VSO-Sprachen haben Präpositionen, SOV-Sprachen haben Postpositionen, SVO-Sprachen liegen in der Mitte, mit einer Tendenz zu Postpositionen. Wie sicher können wir aber sein, dass die in dem Sample beobachteten Tendenzen auch für die Grundgesamtheit zutreffen?

Dazu gibt der χ^2 -Test (der **Chi-Quadrat-Test**) eine Schätzung. Er beantwortet die Frage:

Wie hoch ist die Wahrscheinlichkeit, dass die in der Stichprobe zu sehende Korrelation nicht zufällig ist, sondern eine Tendenz in der Grundgesamtheit widerspiegelt?

Die Überlegungen, die hinter dem Test stehen, sind die folgenden. Wenn es keine Korrelation zwischen den Merkmalsausprägungen gäbe, würden wir folgendes Resultat als

Mittelwert von vielen Stichproben erwarten (die sog. **Erwartungswerte**). Sie ergeben sich jeweils als Produkt von Zeilen- und Spaltensumme geteilt durch die Gesamtzahl der Sprachen, 30.

(2)

	SOV	SVO	VSO
P-NP	5,866666667	6,933333333	3,2
NP-P	5,133333333	6,066666667	2,8

Die 11 SOV-Sprachen, 13 SVO-Sprachen und 6 VSO-Sprachen werden hier im Verhältnis 16:14 auf die Typen P-NP und NP-P aufgeteilt.

Die tatsächlichen Werte der Stichprobe weichen wie folgt ab, wobei wir jeweils von dem beobachteten Wert den Erwartungswert abziehen:

(3)

	SOV	SVO	VSO
P-NP	-5,866666667	3,066666667	2,8
NP-P	5,866666667	-3,066666667	-2,8

(Wir sehen, dass die übereinanderstehenden Zellen jeweils entsprechende Werte haben; dies ist bei nur zwei Zeilen notwendig).

Wir haben hier Plus- und Minus-Werte. Um einheitlich positive Zahlen zu erhalten, quadrieren wir jede Zelle:

(4)

	SOV	SVO	VSO
P-NP	34,41777778	9,404444444	7,84
NP-P	34,41777778	9,404444444	7,84

Diese Zahlen hängen davon ab, wie hoch der bei zufälliger Verteilung erwartete Wert wäre (siehe (2)). Wir teilen die Zellen daher durch diesen und erhalten die folgende Matrix:

(5)

	SOV	SVO	VSO
P-NP	5,866666667	1,356410256	2,45
NP-P	6,704761905	1,55018315	2,8

Bemerkung: Wenn eine Zelle nicht vom zu erwartenden Wert abweicht, stünde hier 0.

Wir summieren nun die Werte und erhalten 20,728.

Zusammenfassung: Der χ^2 -Wert wird also wie folgt errechnet.

$$(6) \quad \chi^2 = \sum \frac{(\text{beobachteter} - \text{erwarteter Wert})^2}{\text{erwarteter Wert}}$$

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). Wir sehen, dass bei 2 Freiheitsgraden ein χ^2 -Wert von 20,782 nur mit einer (viel) geringeren Wahrscheinlichkeit als 0,1% zu erwarten ist. Dass es zwischen den zwei Wortstellungseigenschaften 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

13.1.2 Bemerkungen

Zu unserem Beispiel

- In unserem Beispiel arbeiteten wir mit zu erwartenden Werten unter der Annahme, dass es keine Interaktion von Grundwortstellung und Adpositionstyp gibt. Die Annahme des

Fehlens von Interaktionen wird **Nullhypothese** genannt. Wir könnten auch Werte eingeben, die nach einer bestimmten Theorie zu erwarten wären, und überprüfen, wie gut die beobachteten Werte den theoretisch vorausgesagten entsprechen.

- 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 hier bei den VSO-Sprachen nicht der Fall ist.
- Wenn wir nur einen einzigen Freiheitsgrad haben, gibt es eine Sonderregelung (die Yates-Korrektur).

$$(7) \quad \chi^2 = \sum \frac{(|\text{beobachteter} - \text{erwarteter Wert}| - 0,5)^2}{\text{erwarteter Wert}}$$

13.1.3 Berechnung von Chi-Quadrat mit Excel

In Excel gibt es die Funktion CHITEST, welche bei Eingabe einer Matrix von beobachteten Werten und von zu erwartenden Werten das Signifikanzniveau liefert.

(8) CHITEST(Beobachtete-Werte; Erwartete-Werte).

Für unser Beispiel erhalten wir 3,155E-05, d.h. $3,155 \times 10^{-5}$, also 0,00003155. Dies ist das Maß für die Wahrscheinlichkeit, dass diese Stichprobe gezogen wurde, obwohl die Daten zufallsverteilt waren; diese Wahrscheinlichkeit ist sehr klein und kann daher zurückgewiesen werden.

13.1.4 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 werden die Wahrscheinlichkeiten nach sog. Signifikanzniveaus α wiedergegeben. Dabei gilt:

- $p \geq 0.05$: **nicht signifikant**; in einer von mind. 20 Stichproben ist das Ergebnis auch 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 manchmal mit einem, zwei oder drei Sternchen angeben; nicht signifikante Beziehungen durch "n.s."

13.1.5 Berechnung von Chi-Square mit R

Wir geben mit `scan` die Daten ein und konfigurieren sie mit `dim` als 3×2 -Matrix:

```
> 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
```

Darauf führen wir den Chi-Quadrat-Test durch und erhalten das Ergebnis:

```
> chisq.test(wortstellung)

Pearson's Chi-squared test

data: wortstellung
X-squared = 20.728, df = 2, p-value = 3.155e-05

Warning message:
In chisq.test(wortstellung) : Chi-Quadrat-Approximation kann inkorrekt sein
```

Die Warnung bezieht sich auf die kleinen Zahlenwerte.

13.1.6 Web Chi Square Calculator

Es gibt auch die Möglichkeit, χ^2 auf Statistik-Webseiten berechnen zu lassen. Man sucht solche Seiten mit "chi square calculator".

13.1.7 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 ($\chi^2 = 20.728$, $df = 2$, $p < 0.0001$).

13.1.8 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.

$$(9) \quad \text{Cramérs V} = \text{Quadratwurzel aus } \frac{\chi^2}{N(k-1)}$$

In unserem Beispiel ergibt sich der Wert: Quadratwurzel aus $430 / 30 \cdot 2 = 2,68$. Dies ist ein hoher Effekt. Wir haben im allgemeinen:

- (10) 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

13.2 Der t-Test

Dieser Test untersucht, ob die Korrelation zwischen intervallskalierten Daten in einer Stichprobe als zufällig angesehen werden kann, oder ob sie auf eine entsprechende Korrelation in der Grundgesamtheit zurückgeht. Der Test nimmt an, dass die Daten der Normalverteilung folgen.¹

13.2.1 Beispiel für t-Test: Abhängige Stichproben

Wir kommen auf unser Beispiel von Kapitel 12.10 zurück (Lernen von Vokabeln, konkrete und abstrakte Wörter). Unsere Frage ist: Wie wahrscheinlich ist es, dass die Korrelation in dem Sample auftreten würde, wenn es in der Grundgesamtheit diese Korrelation nicht gibt?

Es gibt den t-Test für **abhängige** und für **unabhängige Stichproben**. In unserem Test haben wir eine abhängige Stichprobe, da wir jeweils dieselbe Versuchsperson mit konkreten und mit abstrakten Wörtern getestet haben.

Schritt 1: Berechnen der Differenz und des Quadrats der Differenz der Messwerte

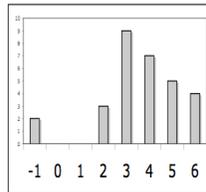
(11)

VPn	Vokab. I	Vokab. II	Differenz	Differenz ²
1	9	5	4	16
2	12	8	4	16
3	19	16	3	9
4	13	9	4	16
...
29	29	8	3	9
30	30	9	2	4
Summe	381	273	108	476
Mittelwert	12,7	9,1		

Schritt 2: Berechnung der Standardabweichungen der Differenzen

Wir berechnen nach Kap. 12.8.5 die Standardabweichung der Differenzen (SDD). Dies erfasst, ob die Differenzen relativ eng oder weit gestreut sind. Das Berechnungsverfahren haben wir bereits erläutert; hier erhalten wir den folgenden Wert und die angegebene Verteilung.

(12) SDD = 1,734



Schritt 3: Berechnung des Standardfehlers des Mittelwerts

Der Standardfehler des Mittelwerts (SED, *standard error of deviation*) relativiert die SDD hinsichtlich der Größe der Stichprobe n. Er wird wie folgt berechnet:

$$(13) \quad SED = \frac{SDD}{\sqrt{n}}$$

In unserem Beispiel haben wir mithin:

$$(14) \quad SED = 1,734/\sqrt{30} = 0,3166$$

Man beachte: SED wird größer, wenn die Unterschiede zwischen den beiden Reihen stärker schwanken, und kleiner, wenn die Stichprobe größer wird.

Schritt 4: Berechnen des t-Wertes

$$(15) \quad t = \frac{\text{Mittelwert}(X) - \text{Mittelwert}(Y)}{SED}$$

In unserem Beispiel erhalten wir

$$(16) \quad t = 0,3166 / 5,477 = 11,3711$$

Man beachte: Der t-Wert wird größer, wenn die Mittelwerte weiter auseinanderliegen, und kleiner, wenn der SED-Wert größer wird, dh. wenn die Unterschiede zwischen den beiden Reihen stärker schwanken oder die Stichprobe kleiner wird.

Schritt 5: Bestimmen der Freiheitsgrade

Die Freiheitsgrade ergeben sich, indem man von der Zahl der Differenzen von Messwerten 1 subtrahiert: Wenn es (n-1) Differenzen gibt und die Summe der Referenzen bekannt ist, liegt die letzte Differenz ebenfalls fest. In unserem Beispiel ist df = 29.

Schritt 6: Signifikanzniveau: Nachschlagen in einer Tabelle

Im allgemeinen verfährt man nach dem zweiseitigen Test; nur wenn es gute Gründe gibt, dass die Abweichung in eine Richtung erfolgt, nach dem einseitigen.

Wir schlagen bei 29 Freiheitsgraden nach und erhalten und sehen, dass der nächste Wert (bei 29 Freiheitsgraden) 3,659 ist, was einer Wahrscheinlichkeit von 0,001 entspricht. Der t-Wert von 11,371 liegt weit darüber, also ist es höchst wahrscheinlich, dass bereits die Grundgesamtheit die Korrelation zeigt.

13.2.2 Berechnen des t-Wertes durch Excel

In Excel kann man unter EXTRAS → ANALYSEFUNKTIONEN den t-Test durchführen. Wir können das genaue Signifikanzniveau feststellen (hier: 1,65⁻¹²).

Zweistichproben t-Test bei abhängigen Stichproben (Paarvergleichstest)		
	Variable 1	Variable 2
Mittelwert	12,7	9,1
Varianz	9,73448276	8,3
Beobachtung	30	30
Pearson Korr	0,83591813	
Hypothetisch	0	
Freiheitsgrad	29	
t-Statistik	11,3711368	
P(T<=t) einsei	1,6558E-12	
Kritischer t-W	1,699127	
P(T<=t) zwei	3,3116E-12	
Kritischer t-W	2,04522961	

¹ Der Test wurde in der Guinness-Brauerei in Dublin 1908 zur Überwachung der Bierqualität von dem Mathematikstudenten W. Gosset erfunden, der dort einen Ferienjob machte. Die Brauerei erlaubte es nicht, dass er ihn unter eigenem Namen veröffentlichte; seitdem ist er unter *Student's t-Test* bekannt.

13.2.3 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

```
> t.test(vokabeln[,1], vokabeln[,2], paired=TRUE)

Paired t-test

data: vokabeln[, 1] and vokabeln[, 2]
t = 11.3711, df = 29, p-value = 3.312e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.952499 4.247501
sample estimates:
mean of the differences
                3.6
```

13.2.4 Tabelle t-Test

Tail Probabilities								
One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
D 1	3.078	6.314	12.71	31.82	63.66	318.3	637	
E 2	1.886	2.920	4.303	6.965	9.925	22.330	31.6	
G 3	1.638	2.353	3.182	4.541	5.841	10.210	12.92	
R 4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	
E 5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	
E 6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	
S 7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	
O 9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	
F 10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	
F 12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	
R 13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	
E 14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	
E 15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	
D 16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	
O 17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	
M 18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	

13.2.5 Der t-Test für unabhängige Stichproben

Nehmen wir an, wir hätten nicht die gleichen Versuchspersonen zwei Mal (für konkrete und für abstrakte Wörter) getestet, sondern zwei Gruppen von Versuchspersonen: Eine für konkrete, eine für abstrakte Wörter. Wir haben also **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 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.

Beispiel

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:

VPn	20-jährige	40-jährige
1	9	5
2	12	8
3	19	16
...
30	11	9

Wir gehen hier auf das Berechnungsverfahren der verschiedenen *t*-Tests für unabhängige Stichproben nicht ein. In unserem Fall sind die beiden Stichproben gleich groß, und wir nehmen an, dass die Varianzen in den beiden Populationen unterschiedlich sein können.

Berechnung durch Excel

Wie zuvor, mit EXTRA → ANALYSEFUNKTIONEN, dann auswählen: ZWEISTICHPROBEN UNTERSCHIEDLICHER VARIANZEN. Wir nehmen an, dass die Differenz der Mittelwerte nach der zurückzuweisenden Hypothese 0 beträgt, und geben als Signifikanzniveau 0,005 an. Wir erhalten das folgende Resultat:

Zweistichproben t-Test unter der Annahme unterschiedlicher Varianzen		
	Variable 1	Variable 2
Mittelwert	12,7	9,1
Varianz	9,734482759	8,3
Beobachtungen	30	30
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	58	
t-Statistik	4,643134696	
P(T<=t) einseitig	1,00603E-05	
Kritischer t-Wert bei einseitigem t-Test	2,66328694	
P(T<=t) zweiseitig	2,01206E-05	
Kritischer t-Wert bei zweiseitigem t-Test	2,91839445	

Die Untersuchung lieferte ein positives Resultat: Die Nullhypothese (keine Differenz zwischen den Populationen) konnte zu $p = 0.00002$ zurückgewiesen werden.

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 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 = 4.6431, df = 57.635, p-value = 2.030e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.047783 5.152217
sample estimates:
mean of x mean of y
 12.7      9.1
```

Auch hier konnte die Nullhypothese mit einer Wahrscheinlichkeit von 0,00002 zurückgewiesen werden.

13.3 Die Varianzanalyse (ANOVA)

Der *t*-Test kann nur verwendet werden bei zwei Meßreihen. Die Varianzanalyse (ANOVA, für "Analysis of Variance") kann auch verwendet werden, wenn mehr Eigenschaften gemessen werden. Wir unterscheiden dabei:

- Die **einfaktorielle V.**: Es liegt eine unabhängige Variable vor;
 - Die **mehrfaktorielle V.**: Es liegt mehr als eine unabhängige Variable vor.
- Eine wichtige Voraussetzung ist wiederum, dass die Eigenschaftsausprägungen in der Grundgesamtheit normalverteilt sind.

13.3.1 Einfaktorielle Varianzanalyse mit Excel

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 Antezedensausdrücke ersetzt wurden.

Es wurde die Lesezeit der Texte (natürlich bei verschiedenen Versuchspersonen, jeweils 16, insgesamt 48) gemessen.

Wir erhalten die dargestellten Messwerte, die wir wie angegeben in ein Excel-Arbeitsblatt übertragen.

Wir führen eine einfache Varianzanalyse (mit EXTRAS → ANAYSEFUNKTIONEN → EINFAKTORIELLE VARIANZANALYSE) durch.

Die Eingabemaske wird wie nebenstehend ausgefüllt, wenn die

	A	B	C
1	Text A	Text B	Text C
2	61	58	61
3	59	59	58
4	66	49	62
5	68	63	57
6	51	52	62
7	70	48	59
8	79	72	51
9	62	56	58
10	66	55	76
11	74	66	65
12	69	64	70
13	63	57	60
14	65	60	61
15	62	59	66
16	64	56	68
17	66	57	62
18			

Resultate an der Position E1 (linke obere Ecke) erscheinen soll.

Die Daten sind nach Spalten geordnet und enthalten oben Beschriftungen.

Als Signifikanzniveau (Alpha) wählen wir 0.05.



E	F	G	H	I	J	K
Anova: Einfaktorielle Varianzanalyse						
ZUSAMMENFASSUNG						
Gruppen	Anzahl	Summe	Mittelwert	Varianz		
Text A	16	1045	65,3125	39,9625		
Text B	16	931	58,1875	37,4958333		
Text C	16	996	62,25	34,2		
ANOVA						
Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
Unterschiede zwischen den Gruppen	408,791667	2	204,395833	5,49164117	0,00734461	3,20431729
Innerhalb der Gruppen	1674,875	45	37,2194444			
Gesamt	2083,66667	47				

In der Zusammenfassung der Daten sehen wir zunächst die Mittelwerte und die Varianz für die Texte A, B, und C.

In dem ANOVA-Teil werden die Unterschiede zwischen den Gruppen A, B und C und innerhalb der Gruppen berechnet (mithilfe der mittleren Quadratsumme der Abweichung vom jeweiligen Mittelwert). Die Prüfgröße *F* bezieht sich auf die Varianz, der kritische *F*-Wert auf den Wert, ab dem das Ergebnis zum angegebenen Niveau ($p = 0.05$) signifikant ist. Die Prüfgröße *F* liegt über diesem Wert, daher ist der Unterschied zwischen den Gruppen tatsächlich signifikant. Der genaue *p*-Wert ist ebenfalls angegeben, er beträgt ca. 0,007 und liegt damit noch beträchtlich unter dem Wert 0.05 und erreicht tatsächlich das nächste Signifikanzniveau, 0.01.

Das Ergebnis der ANOVA würde berichtet mit: Einfaktorielle Varianzanalyse, $F(2, 45) = 5,49$, $p < 0.01$. Damit wird mitgeteilt: Es handelt sich um einen Test mit 2 Freiheitsgraden zwischen den Gruppen (wir haben drei Gruppen) und 45 Freiheitsgraden unter den gemessenen Werten (jeweils 16 Werte pro Kategorie). Der zugehörige *F*-Wert ist (gerundet) 5,49, was dem angegebenen Signifikanzniveau entspricht.

Das Ergebnis sagt lediglich aus, dass die Gruppe mit dem höchsten Mittelwert (Text A) von der Gruppe mit dem niedrigsten (Text B) signifikant verschieden ist. Ob nun im Vergleich auch Text A von C und Text B von C signifikant verschieden ist, muss gesondert nachgeprüft werden.

13.3.2 Zweifaktorielle Varianzanalyse

Nehmen wir nun an, die oberen acht Versuchspersonen Männer und die unteren acht Frauen waren. Wir sind nun auch interessiert an mögliche Einflüsse des Geschlechts. Dies führt zu einer **zweifaktoriellen** Analyse, da es **zwei unabhängige Variablen** gibt: Geschlecht (M, W) und Texttype (A, B, C).

Im Folgenden wird gezeigt die Excel-Tabelle, die Eingemaske für die zweifaktorielle Varianzanalyse mit Messwiederholung (Messwiederholung, weil wir jede Kategorie mehrfach, nämlich je acht Mal, messen).

	A	B	C	D
1		Text A	Text B	Text C
2	Männlich	61	58	61
3		59	59	58
4		66	49	62
5		68	63	57
6		51	52	62
7		70	48	59
8		79	72	51
9		62	56	58
10	Weiblich	66	55	76
11		74	66	65
12		69	64	70
13		63	57	60
14		65	60	61
15		62	59	66
16		64	56	68
17		66	57	62

Zweifaktorielle Varianzanalyse mit Messwiederholung

Eingabe

Eingabebereich: SAS1:SDS17

Zeilen je Stichprobe: 8

Alpha: 0,05

Ausgabe

Ausgabebereich: SFS1

F	G	H	I	J	K	L
Anova: Zweifaktorielle Varianzanalyse mit Messwiederholung						
ZUSAMMENFASSUNG	Text A	Text B	Text C	Gesamt		
<i>Männlich</i>						
Anzahl	8	8	8	24		
Summe	516	457	468	1441		
Mittelwert	64,5	57,125	58,5	60,0416667		
Varianz	69,4285714	62,4107143	12,8571429	54,7373188		
<i>Weiblich</i>						
Anzahl	8	8	8	24		
Summe	529	474	528	1531		
Mittelwert	66,125	59,25	66	63,7916667		
Varianz	14,6964286	15,3571429	28,2857143	28,5199275		
<i>Gesamt</i>						
Anzahl	16	16	16			
Summe	1045	931	996			
Mittelwert	65,3125	58,1875	62,25			
Varianz	39,9625	37,4958333	34,2			
ANOVA						
Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
Stichprobe	168,75	1	168,75	4,98680739	0,03092632	4,07265366
Spalten	408,791667	2	204,395833	6,04019349	0,00494766	3,21994229
Wechselwirkung	84,875	2	42,4375	1,25408971	0,29580034	3,21994229
Fehler	1421,25	42	33,8392857			
Gesamt	2083,66667	47				

Wir erhalten das nebenstehende Ergebnis.

In der Zusammenfassung erhalten wir zunächst die Mittelwerte und Varianzen der männlichen, weiblichen und aller Versuchspersonen zusammen, die Mittelwerte und Varianzen der drei Texte und die Mittelwerte und Varianzen von jeder Kombination von Text und Geschlecht. Es fällt bei diesen hypothetischen Daten auf: Männer zeigen einen etwas kleineren Mittelwert als Frauen, dafür aber haben sie eine wesentlich höhere Varianz. Der Mittelwert-Unterschied zwischen Männern und Frauen ist vor allem dem Text C zuzuschreiben (58,5 vs. 66).

Die ANOVA zeigt, dass es zwischen den zwei Versuchspersonengruppen (Männern und Frauen) signifikante Unterschiede gibt ($p = 0,03$, d.h. $p < 0,05$) und dass es zwischen den drei Satzarten hoch signifikante Unterschiede gibt ($p = 0,0049$, d.h. $p < 0,01$). Eine Wechselwirkung zwischen den Kategorien konnte nicht als signifikant bestimmt werden ($p = 0,295$).

13.3.3 Varianzanalyse in R

Das Statistikpaket R erlaubt eine wesentlich detailliertere Varianzanalyse als Excel. Unter anderem kann man auch bestimmen, inwieweit die Mittelwerte einzelner Kategorien signifikant voneinander abweichen (z.B. Männer bei Text A und Frauen bei Text A). Einfache paarweise Tests (t -Tests) kann man hier nicht durchführen, da eine Vielzahl von Kombinationen getestet werden kann und dabei es wahrscheinlich ist, dass bei den Stichproben Signifikanzniveau erreicht wird, obwohl dem die Verhältnisse in der Grundgesamtheit nicht entsprechen.

13.4 Schlußbemerkungen zur schließenden Statistik

13.4.1 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.

13.4.2 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.

14. Werkzeuge der Modellierung: Mengen und Relationen

Abschließend wenden wir uns wesentlichen theoretischen Werkzeugen für die Beschreibung sprachlicher Phänomene zu. Sie dienen dazu, formale **Modelle** zu entwickeln, die bestimmte Beobachtungen erfassen und die selbst wiederum Voraussagen machen, die man überprüfen kann. Wir beschränken uns hier auf Werkzeuge der **diskreten** Mathematik; solche der numerischen Mathematik (z.B. Differentialgleichungen, Fourier-Analyse), wie sie z.B. für die Phonetik nötig sind, sind für diese Einführung zu komplex.

14.1 Mengen

14.1.1 Was ist eine Menge?

Die Mengenlehre wurde gegen Ende des 19. Jahrhunderts von dem Mathematiker Georg Cantor als theoretische Basis der Mathematik entwickelt. Der Grundgedanke war, eine elementare, einfache und konsistente Theorie zu schaffen, auf deren Grundlage sich die gesamte Mathematik aufbauen ließe. Nachfolgend die klassische Definition einer **Menge** (englisch: **set**).

- (1) 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. Des Weiteren handelt es 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).

Wann sind zwei Mengen gleich? Genau dann, wenn sie dieselben Elemente enthalten:

- (2) Definition: $A = B$ gdw. für alle x gilt: $x \in A$ gdw. $x \in B$

14.1.2 Spezifikation von Mengen: Aufzählung und Abstraktion

Es gibt verschiedene Weisen, wie man Mengen angeben kann. Hier wollen wir zwei Methoden einführen: die Aufzählung und die Abstraktion.

Aufzählung

Wir können die Elemente einer Menge einfach **aufzählen**. Man verwendet dazu geschweifte Klammern sowie Kommata, um die Elemente voneinander zu trennen. Beispiele:

- (3) a. $\{a, e, i, o, u\}$ (die fünf Grundvokale)
b. $\{a, e, i, \{o, u\}\}$ (diese Menge enthält nur 4 Elemente, eines ist selbst eine Menge)
c. $\{a\}$ (eine Einermenge)
d. $\{\}$ (die leere Menge, auch \emptyset geschrieben)

Hierbei ist zu beachten, dass die Reihenfolge der Aufzählung irrelevant ist: $\{a, b, c, d, e\} = \{b, c, d, a, e\}$. Ferner spielt auch das mehrfache Vorkommen eines Elementes keine Rolle: $\{a, b, c, d, e\} = \{a, b, c, c, d, e\}$.

Abstraktion

Die Abstraktion, auch **Prädikatsnotation** genannt, beschreibt die Elemente, die zu der Menge gehören, in einer Sprache – auf Deutsch beispielsweise oder in einer mathematischen Sprache. Alle Objekte, auf welche die Beschreibung zutrifft, und nur diese Objekte, sind dann in der Menge enthalten. Die folgende Schreibweise ist gebräuchlich:

- (4) $\{\text{Variable} \mid \text{Beschreibung der Variablen}\}$

Typischerweise verwenden wir x, y, z als Buchstaben für Variablen. Einige Beispiele:

- (5) a. $\{x \mid x \text{ ist ein Grundvokal}\}$
b. $\{x \mid x \text{ ist eine natürliche Zahl und } 1 \leq x \leq 1000\}$
c. $\{x \mid x \text{ ist eine natürliche Zahl}\}$.

Beispiel (5.a) liest man “die Menge aller x so, dass x ein Grundvokal ist”.

14.1.3 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:

- (6) Definition: $A \subseteq B$ gdw. gilt: Für alle x , wenn $x \in A$, dann $x \in B$.

Zwei Beispiele für Teilmengenbeziehungen:

- (7) a. $\{a, e, i\} \subseteq \{a, e, i, o, u\}$
b. $\emptyset \subseteq \{a, e, i\}$

Die leere Menge ist Teilmenge jeder anderen Menge. Die Teilmengenbeziehung gehorcht drei wichtigen Gesetzen:

- (8) a. Reflexivität: Für jede Menge A gilt: $A \subseteq A$
b. Transitivität: Wenn $A \subseteq B$ und $B \subseteq C$, dann gilt auch: $A \subseteq C$
b. Antisymmetrie: Wenn $A \subseteq B$ und $B \subseteq A$, dann gilt: $A = B$

Das Gesetz der Reflexivität besagt, dass laut Definition jede Menge eine Teilmenge von sich selbst ist. Wenn man das ausschließen will, kann man zu der Relation der **echten Teilmenge** greifen, für die wir die Schreibweise $A \subset B$ verwenden:

- (9) Definition: $A \subset B$ gdw. $A \subseteq B$, aber nicht: $B \subseteq A$.

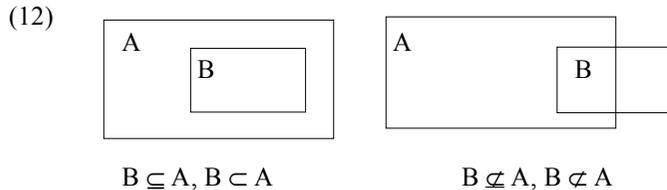
14.1.4 Kardinalität von Mengen

Wir wollen manchmal wissen, wie viele Elemente eine Menge hat. Dies wird die Kardinalität einer Menge genannt. Die Kardinalität der Menge A wird durch $\text{card}(A)$ oder $\#(A)$ angegeben, wobei die Klammern auch weggelassen werden können, wenn die Menge durch eine Mengenklammer gegeben ist.

- (10) $\text{card}(A) = \#(A) =$ die Zahl der Elemente in der Menge A
- (11) a. $\#\{a, b, c\} = 3$ c. $\#\{a\} = 1$
 b. $\#\{a, b, b, c\} = 3$ d. $\#\{\emptyset\} = 0$

14.1.5 Darstellung von Mengen durch Venn-Diagramme

Eine nützliche Methode zur Darstellung mengentheoretischer Beziehungen bilden die sogenannten **Venn-Diagramme** (nach dem Mathematiker John Venn), auch "Euler-Kreise" genannt (nach dem im 18. Jahrhundert lebenden Mathematiker Leonhard Euler, der sie als didaktisches Hilfsmittel in seinen im Jahre 1768 in St. Petersburg veröffentlichten *Lettres à une Princesse d'Allemagne* einführte). Wir haben im vorausgegangenen Kapitel schon so etwas wie Venn-Diagramme benutzt. 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:



14.1.6 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.

- (13) Definition: $A \cup B = \{x \mid x \in A \text{ oder } x \in B\}$
- (14) a. $\{a, e, i\} \cup \{i, o, u\} = \{a, e, i, o, u\}$
 b. $\{a, e, i\} \cup \{o, u\} = \{a, e, i, o, u\}$
 c. $\{a, e, i\} \cup \emptyset = \{a, e, i\}$

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:

- (15) Definition: $A \cap B = \{x \mid x \in A \text{ und } x \in B\}$
- (16) a. $\{a, e, i\} \cap \{i, o, u\} = \{i\}$
 b. $\{a, e, i\} \cap \{o, u\} = \emptyset$
 c. $\{a, e, i\} \cap \emptyset = \emptyset$

Die mengentheoretische **Differenz** (englisch: **subtraction**) $A \setminus B$ ist diejenige Menge, welche genau die Elemente aus A enthält, die nicht in B enthalten sind:

- (17) Definition: $A \setminus B = \{x \mid x \in A \text{ und } x \notin B\}$
- (18) a. $\{a, e, i\} \setminus \{i, o, u\} = \{a, e\}$
 b. $\{a, e, i\} \setminus \{o, u\} = \{a, e, i\}$
 c. $\{a, e, i\} \setminus \emptyset = \{a, e, i\}$

Man beachte, dass die mengentheoretischen Operationen \cup , \cap und \setminus grundsätzlich von der Teilmengenbeziehung verschieden sind. Wenn wir $A \cup B$ schreiben, so bezeichnen wir damit eine neue Menge. Schreiben wir stattdessen $A \subseteq B$, so erhalten wir keine neue Menge, sondern eine Behauptung, die wahr oder falsch sein kann.

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 (wobei $A \subseteq U$), geschrieben A' , mittels

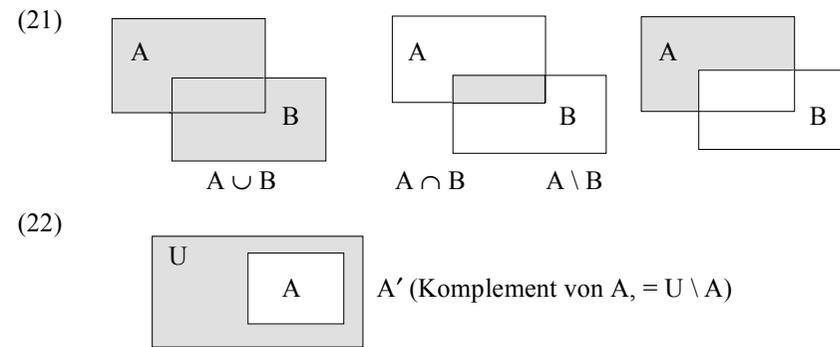
- (19) Definition Komplement: $A' =_{\text{def}} U \setminus A$.

Zum Beispiel gilt bezüglich der Grundvokale als Universum

- (20) a. $\{a, e, i\}' = \{o, u\}$

14.1.7 Darstellung von Mengenoperationen durch Venn-Diagramme

Wir können die eben eingeführten Operationen wie folgt durch Venn-Diagramme darstellen:



14.1.8 Mengentheoretische Gesetze

Die mengentheoretischen Begriffe wie Vereinigung, Durchschnitt, Teilmenge, Komplement etc. sind durch strukturelle Beziehungen, die wir in Gleichungen ausdrücken können, miteinander verbunden. Im Folgenden seien A, B, C beliebige Mengen. Die jeweiligen Gesetze lassen sich mit Hilfe von Venn-Diagrammen darstellen.

- (23) a. Idempotenz: $[A \cap A] = A$
 $[A \cup A] = A$
- b. Kommutativität: $[A \cap B] = [B \cap A]$
 $[A \cup B] = [B \cup A]$
- c. Assoziativität: $[A \cap [B \cap C]] = [[A \cap B] \cap C]$
 $[A \cup [B \cup C]] = [[A \cup B] \cup C]$
- d. Distributivität: $[A \cap [B \cup C]] = [[A \cap B] \cup [A \cap C]]$
 $[A \cup [B \cap C]] = [[A \cup B] \cap [A \cup C]]$
- e. De Morgan: $[A \cap B]' = [A' \cup B']$
 $[A \cup B]' = [A' \cap B']$

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:

- (24) a. $A \subseteq B$ gdw. $A \cup B = B$
b. $A \subseteq B$ gdw. $A \cap B = A$

Die leere Menge \emptyset und das Universum U spielen die Rolle von Kontradiktion und Tautologie; sie gehorchen den folgenden Gesetzmäßigkeiten:

- (25) $A \cup A' = U$ $U' = \emptyset$ $A'' = A$
 $A \cap A' = \emptyset$ $\emptyset' = U$

Wie in der Aussagenlogik, so kann man auch in der Mengenlehre diese Gesetzmäßigkeiten zum Nachweis von bestimmten Beziehungen zwischen Mengen heranziehen.

14.2 Relationen

14.2.1 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:

- (26) $\{\langle a, b \rangle, \langle b, c \rangle, \langle c, d \rangle, \langle e, f \rangle, \dots, \langle x, y \rangle, \langle y, z \rangle\}$

Wie stets bei Mengen, kommt es hierbei auf die Reihenfolge der geordneten Paare nicht an.

14.2.2 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 $A \times B$.

- (27) $A \times B = \{\langle x, y \rangle \mid x \in A \text{ und } y \in B\}$

Beispiel: $\{a, b, c\} \times \{1, 2\} = \{\langle a, 1 \rangle, \langle b, 1 \rangle, \langle c, 1 \rangle, \langle a, 2 \rangle, \langle b, 2 \rangle, \langle c, 2 \rangle\}$

Unter dem **Feld** einer Relation versteht man die Menge aller Objekte, die zu irgendeinem Objekt in der Relation stehen.

- (28) das Feld von $R = \{x \mid \text{es gibt ein } y \text{ sodass gilt: } \langle x, y \rangle \in R \text{ oder } \langle y, x \rangle \in R\}$

Das Feld der Relation $\{\langle a, 1 \rangle, \langle b, 2 \rangle, \langle c, 2 \rangle\}$ ist beispielsweise die Menge $\{a, b, c, 1, 2\}$.

14.2.3 Eigenschaften von Relationen

Viele Relationen weisen bestimmte Eigenschaften auf; die wichtigsten davon werden genannt.

- (29) R ist **reflexiv** wenn gilt: Wenn $\langle x, y \rangle \in R$, dann gilt: $\langle x, x \rangle \in R$ und $\langle y, y \rangle \in R$

- (30) R ist **transitiv** wenn gilt: Wenn $\langle x, y \rangle \in R$ und $\langle y, z \rangle \in R$, dann gilt: $\langle x, z \rangle \in R$

- (31) R ist **symmetrisch** wenn gilt:

- (32) 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.

14.2.4 Wichtige Relationstypen

Mit den eben definierten Eigenschaften können wir wichtige Typen von Relationen definieren.

Ein Typ von Relation sind die **Äquivalenzrelationen**. Sie 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:

- (33) 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**. Wenn Hans, Maria und Peter in Berlin geboren sind, Otto und Sophia in Lübeck und Karl in Buxtehude, dann erhalten wir folgende Partition:

- (34) $\{\{Hans, Maria, Peter\}, \{Otto, Sophia\}, \{Karl\}\}$

Ein weiterer wichtiger Typ von Relation ist die der **Präordnung**. Beispiele sind "ist größer als" (für Personen). Präordnungen haben die folgenden Eigenschaften:

- (35) 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".

Eine spezifischere Relation ist die der **Halbordnung**; sie ist wie folgt definiert:

- (36) R ist eine **Halbordnung** (engl. partial order) gdw:

- a. R ist reflexiv
b. R ist transitiv
b. R ist antisymmetrisch

Ein Beispiel für eine Halbordnung ist die Teilmengenbeziehung. Sie ist reflexiv: Für alle Mengen A gilt: $A \subseteq A$. Sie ist transitiv: Wenn $A \subseteq B$ und $B \subseteq C$, dann $A \subseteq C$. Und sie ist

antisymmetrisch: Wenn $A \subseteq B$ und $B \subseteq A$, dann gilt auch: $A = B$. Wir haben eben den Begriff der allgemeinen (nicht-strengen) Halbordnung definiert; bei strengen Halbordnungen werden reflexive Paare der Art $\langle x, x \rangle$ entfernt. Ein Beispiel ist die Relation der echten Teilmenge, \subset .

Abschließend führen wir noch den Begriff der **totalen Ordnung** (auch einfach **Ordnung**) ein. Dies ist eine Halbordnung, die auch die Eigenschaft der Totalität erfüllt:

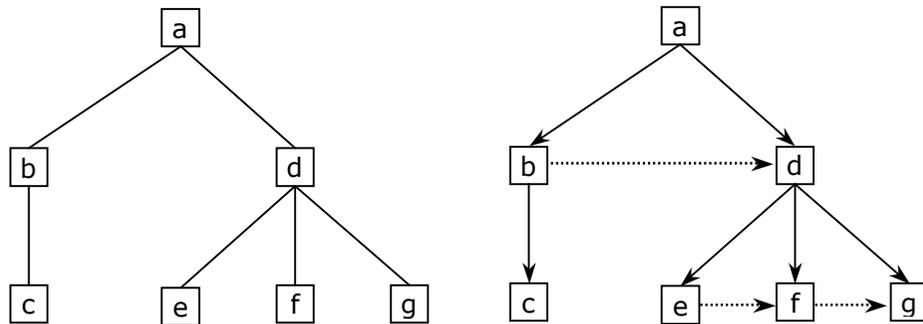
- (37) R ist eine **totale Ordnung** gdw.
- R ist eine Halbordnung
 - R ist total, d.h. für alle x, y im Feld von R gilt: $\langle x, y \rangle \in R$ oder $\langle y, x \rangle \in R$.

Dies trifft zu z.B. für die Relation \leq für natürliche Zahlen: Diese Relation erfüllt die Gesetze der 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 und m identisch sind).

14.3 Anwendungen: Definition von syntaktischen Bäumen

Ein wichtiges Instrument zur Beschreibung von syntaktischen Strukturen sind **Bäume**. Dies ist ein Untertyp von **Graphen**.

- (38) Beispiel eines Baumes in normaler Darstellung und mit expliziten Dominanz- und Präzedenzrelationen.



Graphen bestehen im allgemeinen bestehen aus **Knoten** und diese Knoten verbindende **Kanten**. Die Kanten, die jeweils zwei Knoten verbinden, können als (zweistellige) **Relation** dargestellt werden, die die **Dominanzrelation**. Per Konvention stehen die dominierenden Knoten über den dominierten. Zum Beispiel dominiert der Knoten a den Knoten d . 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. $\langle a, d \rangle \in D$.

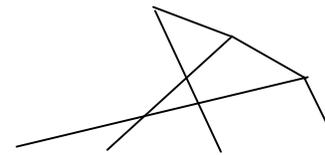
- Zwischen Knoten gibt es ferner die Relation der Reihenfolge oder **Präzedenz** (von links nach rechts; diese Relation wird in der Regel nicht eigens durch Kanten dargestellt). Diese Relation ist ebenfalls eine **Halbordnung**. Sie ist daher transitiv. Zum Beispiel geht der Knoten e dem Knoten f voran, 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 Dominanz und Präzedenz gibt es einen Zusammenhang, der verhindert, dass sich Kanten kreuzen:
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 relativ wortstellungsfreien Sprachen wie dem Deutschen keine Analysen der folgenden Art vornehmen darf:

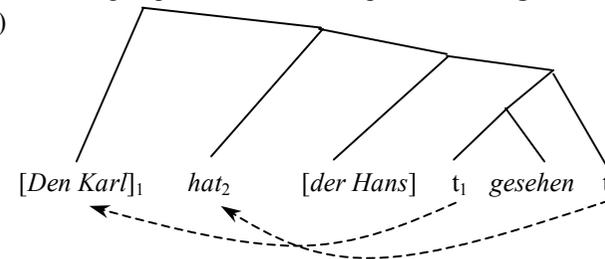
- (39) *Den Karl hat der Hans gesehen.*



[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.

- (40)



15. Werkzeuge der Modellierung: Funktionen

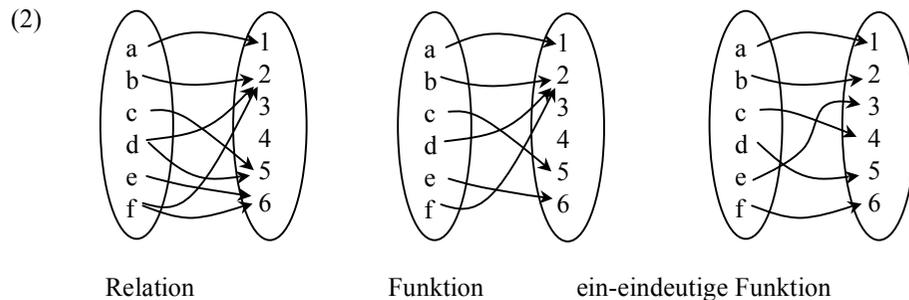
15.1 Funktionen

15.1.1 Funktionen als rechtseindeutige Relationen

Unter den Relationen gibt es einen bestimmten Typ, der für Mathematiker besonderes Interesse besitzt, die **Funktionen**. Es handelt sich dabei im wesentlichen um Relationen, die **rechtseindeutig** sind. Damit ist gemeint, dass ein Element x immer nur mit genau einem Element y in der Verbindung $\langle x, y \rangle$ stehen darf. Diese Eigenschaft kann man wie folgt ausdrücken:

- (1) Eine Funktion R ist eine Funktion, wenn für alle x, y, z gilt:
Wenn $\langle x, y \rangle \in R$ und $\langle x, z \rangle \in R$, dann gilt: $y = z$

Man kann Relationen allgemein darstellen, indem man Elemente zweier Mengen durch Pfeile verbindet, wobei " $x \rightarrow y$ " dafür steht, dass $\langle x, y \rangle$ sich in der Relation befinden. Betrachten wir die folgenden Relationen zwischen einer Menge $\{a, b, c, d, e, f\}$ und einer Menge $\{1, 2, 3, 4, 5\}$. Das erste Beispiel ist eine Relation, die keine Funktion ist. Das zweite Beispiel ist eine Relation, die auch eine Funktion ist, da keinem Element in der ersten Menge mehr als ein Element in der zweiten zugewiesen wird. Das dritte Beispiel ist eine sogenannte **ein-eindeutige Funktion**, da darüber hinaus auch keinem Element der zweiten Menge mehr als ein Element in der ersten Menge entspricht.



Offenbar ist die Bedeutung von *Vater* rechtseindeutig, da jeder Mensch genau einen (biologischen) Vater hat:

- (3) Für alle x, y, z : Wenn $\langle x, y \rangle \in \llbracket \text{Vater} \rrbracket$ und $\langle x, z \rangle \in \llbracket \text{Vater} \rrbracket$, dann $y = z$.

Hingegen ist die Bedeutung von *Tochter* nicht rechtseindeutig, da eine Elternperson (z.B. ein Vater) mehr als nur eine Tochter haben kann.

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:

- (4) $y = f(x)$.

Wir sagen, dass x das **Argument** und y der **Wert** ist. Wir sprechen davon, dass F auf x **angewendet** wird und dass f x auf y **abbildet**. Häufig verwenden wir für Funktionen. Man beachte, dass die in (4) illustrierte Schreibweise nicht zulässig ist für Relationen, die keine

Funktionen sind. Wir können sagen: der Vater von Jakob IST Isaak, aber wir können unmöglich sagen: der Sohn von Isaak IST Jakob – schließlich hatte Isaak einen zweiten Sohn, nämlich Esau.

Ein weiteres wichtiges Begriffspaar sind der **Definitionsbereich**, englisch **domain** und der **Wertebereich**, englisch **range** einer Funktion f :

- (5) a. Definitionsbereich einer Funktion f , $\text{DOM}(f)$: $\{x \mid \text{es gibt } y \text{ so, dass } \langle x, y \rangle \in f\}$
b. Wertebereich 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 . Der Definitionsbereich von $\llbracket \text{Vater} \rrbracket$ ist die Menge der Personen (jeder hat einen Vater), der Wertebereich von $\llbracket \text{Vater} \rrbracket$ ist die Menge der Väter.

15.1.2 Beschreibung von Funktionen

Für die Beschreibung von Funktionen haben sich verschiedene Schreibweisen eingebürgert. Wenn die Funktion klein ist, kann man die Paare einfach aufzählen. Das folgende Beispiel zeigt auch die Notationskonvention, Bedeutungen von Ausdrücken durch Doppelklammern wiederzugeben.

- (6) $\llbracket \text{Vater} \rrbracket = \left[\begin{array}{l} \text{Isaak} \rightarrow \text{Abraham} \\ \text{Jakob} \rightarrow \text{Isaak} \\ \text{Esau} \rightarrow \text{Isaak} \\ \dots \end{array} \right]$

Dies macht deutlich, dass es sich bei Funktionen im wesentlichen um **Zuweisungsvorschriften** oder **Abbildungen** (englisch **mappings**) handelt. Beispielsweise weist die Funktion $\llbracket \text{Vater} \rrbracket$ dem Isaak den Abraham zu, dem Jakob den Isaak usw.

Eine weitere Schreibweise ist die folgende:

- (7) $\llbracket \text{Vater} \rrbracket \quad v: \text{Väter} \rightarrow \text{Personen},$
 $x \mapsto \text{die Vater von } x.$

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.

In der Logik und auch in der linguistischen Semantik hat sich die **Lambda-Notation** durchgesetzt. Die *Vater*-Funktion wird hier wie folgt dargestellt:

- (8) $\llbracket \text{Vater} \rrbracket = \lambda x[\text{Vater von } x]$

Wir nennen solche Ausdrücke **Lambda-Terme**. Sie haben die folgende Struktur:

- (9) λ Variable [Beschreibung des Wertes der Variablen]

Dieser Ausdruck steht für diejenige Funktion, welche jedem Objekt, für das 'Variable' stehen kann, den Wert der Variablen gemäß der Beschreibung im sogenannten Rumpf des Lambda-Terms zuordnet. Die Bildung eines Lambda-Terms aus einer Beschreibung, die eine Variable enthält, wie z.B. "Mutter von x " oder " $x^2 + x + 1$ ", heißt (Lambda) **Abstraktion**.

Wie schon angedeutet, ist eine nette Eigenschaft der Lambda-Notation die, dass den Funktionen ihre Definition auf der Stirn geschrieben steht. Das macht es möglich, den Wert einer Funktion angewendet auf ein Argument anzugeben, indem man einfach die Lambda-Variablen durch das Argument ersetzt. Dieser Vorgang heißt (Lambda) **Konversion** bzw.

Reduktion:

- (10) a. $\lambda x[\text{Vater von } x](\text{Isaak})$
 = Vater von Isaak
 = Abraham
- b. $\lambda x[x^2 + x + 1](3)$
 = $3^2 + 3 + 1$
 = $9 + 3 + 1$
 = 13

Die folgenden Beispiele zeigen, dass man auch Funktionen definieren kann, die Funktionen als Werte liefern können, und solche, die Funktionen als Argumente erwarten:

- (11) a. $\lambda x[\lambda y[2 \cdot x + y]](4)(5)$
 = $\lambda y[2 \cdot 4 + y](5)$
 = $[2 \cdot 4 + 5]$
 = 13
- b. $\lambda f[f(2) + f(3)](\lambda x[x^2 + 1])$
 = $[\lambda x[x^2 + 1]](2) + \lambda x[x^2 + 1](3)$
 = $[2^2 + 1 + 3^2 + 1]$
 = 15

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 λx und Vorkommen von Variablen $[\dots x \dots]$. 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, wie in dem folgenden Beispiel:

$$(12) \lambda x[\lambda f[f(x) + 1](\lambda x[2x + 1])]$$

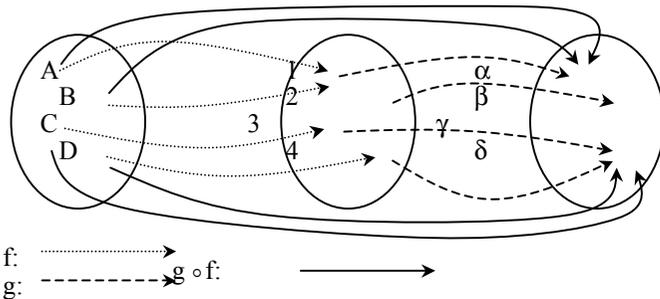
In solchen Fällen ist es besser, eine Umbenennung der Variablen vorzunehmen, z.B.:

$$(13) \lambda x[\lambda f[f(x) + 1](\lambda y[2y + 1])]$$

15.1.3 Verknüpfung von Funktionen

Wir können Funktionen nicht nur auf Argumente anwenden; wir können Funktionen (und allgemeiner Relationen) auch miteinander verknüpfen. Beispiel:

(14)



Wir schreiben $g \circ f$ für die Verknüpfung der Funktion f mit der Funktion g . Diese Operation ist wie folgt definiert:

$$(15) \text{ Für alle } x \text{ im Definitionsbereich von } f: g(f(x)) = g \circ f(x)$$

Es gibt Phänomene in der natürlichen Sprache, in denen solche Verknüpfungen eine Rolle spielen. Sehen wir uns die folgenden schwedischen Verwandtschaftstermini an:

- (16) a. *farfar*: Vater des Vaters c. *mormor*: Mutter der Mutter
 b. *farmor*: Mutter des Vaters d. *morfar*: Vater der Mutter

Diese Formen reflektieren die Verknüpfung der Funktionen *far* 'Vater' und *mor* 'Mutter':

$$(17) \llbracket \text{morfar} \rrbracket = \llbracket \text{far} \rrbracket \circ \llbracket \text{mor} \rrbracket$$

Nach der Definition der funktionalen Komposition gilt dann:

$$(18) \langle x, y \rangle \in \llbracket \text{morfar} \rrbracket \text{ gdw. es gibt ein } z \text{ mit } \langle x, z \rangle \in \llbracket \text{mor} \rrbracket \text{ und } \langle z, y \rangle \in \llbracket \text{far} \rrbracket.$$

Wir erhalten also den Großvater mütterlicherseits, den Vater der Mutter.

15.2 Charakteristische Funktionen

Mengen haben wir als Zusammenfassungen von Elementen (eines gegebenen Universums) definiert. Eine Menge zu kennen bedeutet, in der Lage zu sein, die Elemente jener Menge zu identifizieren. Wir müssen von jedem Objekt wissen, ob es in der Menge enthalten ist oder nicht. Diese Information kann als eine Funktion gegeben sein, als diejenige Funktion, die jedem Objekt in dem Universum genau einen von zwei möglichen Werten 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, weil sie die jeweilige Menge "charakterisieren". Wir schreiben χ_A ("chi-A") für die charakteristische Funktion der Menge A. Sei U ein Universum und A eine Menge mit $A \subseteq U$, dann haben wir die folgende Definition für die charakteristische Funktion von A:

Beispiel:

(19) Sei das Universum U die Menge der Vokale {a, e, i, o, u}. Dann gilt:

$$\chi_{\{e, i\}} = \{\langle a, \underline{0} \rangle, \langle e, \underline{1} \rangle, \langle i, \underline{1} \rangle, \langle o, \underline{0} \rangle, \langle u, \underline{0} \rangle\}$$

Man kann also eine Menge A immer auch als eine charakteristische Funktion χ_A angeben. Die Elementschaftsbeziehung kann man dann durch die Anwendung dieser Funktion ausdrücken:

$$(20) x \in A \text{ gdw. } \chi_A(x) = \underline{1}$$

Eine Anwendung dieser Idee, eine Menge durch ihre charakteristische Funktion anzugeben, ist die Verwendung der Lambda-Notation zur Angabe von Mengen. Die beiden folgenden Schreibweisen sind äquivalent:

- (21) a. $\{x \mid x \text{ ist rot}\}$
 b. $\lambda x[x \text{ ist rot}]$

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 $\underline{1}$ zuweist, falls x rot ist, und den Wert $\underline{0}$, falls dies nicht der Fall ist.

Wir können auch die mengentheoretischen Relationen und Operationen, wie die Teilmengenbeziehung \subseteq und die Schnittbildung \cap , für charakteristische Funktionen definieren. Dies erlaubt uns dann, Beziehungen wie die folgenden aufzustellen:

$$(22) [\lambda x[x \text{ ist rot}]] \cap [\lambda x[x \text{ ist rund}]] \subseteq [\lambda x[x \text{ ist rot}]]$$

15.3 Aufgaben: Mengen, Relationen, Funktionen, semantische Beziehungen

- Wir haben gesehen, dass die Hyponymie mithilfe des Begriffs der Teilmenge modelliert werden kann. Zeigen Sie, wie mit mengentheoretischen Begriffen die Beziehung von konträren Begriffen (Beispiel: *reich* und *arm*) und von komplementären Begriffen (Beispiel: *reich* und *nicht reich*) modellieren werden können. Gehen Sie dabei von der Grundmenge der Menge der Menschen aus.
- Es sei $A = \{1, 2, 3, 4\}$, $B = \{4, 5, 6, 7\}$, U (Universum) = $\{x \mid 0 \leq x \leq 10\}$
Was ist $A \cap B$, $A \cup B$, $A \setminus B$, A' , $A \cup (B')$?
- Definieren Sie “ \cap ”, mithilfe der Operationen \cup und $'$.
D.h., geben Sie eine Gleichung $A \cap B = \dots$, wobei “ \dots ” das “ \cap ”-Zeichen nicht enthält.
- Die mengentheoretischen Regeln für \cup und \cap scheinen den Regeln der Addition $+$ und Multiplikation \cdot ähnlich zu sein. Beispielsweise ist $+$ kommutativ, da wir $a+b = b+a$ haben. Vergleichen Sie die mengentheoretischen Gesetze (Kommutativität, Assoziativität, Distributivität, Idempotenz) mit den arithmetischen und stellen Sie Ähnlichkeiten und Unterschiede fest.
- Welche der folgenden Paarmengen ist eine Funktion mit Argumentbereich $\{1,2,3,4\}$?
i. $\{\langle 1, a \rangle, \langle 2, b \rangle, \langle 3, c \rangle, \langle 4, d \rangle\}$ ii. $\{\langle 1, a \rangle, \langle 2, b \rangle, \langle 3, c \rangle, \langle 3, d \rangle, \langle 4, d \rangle\}$
iii. $\{\langle 1, a \rangle, \langle 2, b \rangle, \langle 3, c \rangle, \langle 4, c \rangle\}$ iv. $\{\langle 1, a \rangle, \langle 2, b \rangle, \langle 3, c \rangle\}$
- Reduzieren Sie die folgenden Lambda-Ausdrücke so weit wie möglich:
a) $\lambda x[2x + 3](7)$
b) $\lambda x[\lambda y[2x + y](3)](4)$ (auch als $\lambda x\lambda y[2x + y](3)(4)$ darstellbar)
c) $\lambda f[f(7) + 4](\lambda x[2x + 3])$
d) $\lambda f\lambda y[f(7) + y](4)(\lambda x[2x + 3])$
e) $\lambda f[f(7) + f(6)](\lambda x[2x + 3])$
f) $\lambda f[f(f(7))](\lambda x[2x + 3])$
- Geben Sie die charakteristischen Funktion χ_{\emptyset} , $\chi_{\{a, c\}}$, $\chi_{\{c, d\}}$, and $\chi_{\{a, b, c, d\}}$ an, d.h. die charakteristischen Funktionen der Mengen \emptyset , $\{a, c\}$, $\{c, d\}$ and $\{a, b, c, d\}$, mit der Menge $\{a, b, c, d\}$ als Universum.
- Wir können den Begriff *Onkel* unter Bezugnahme auf die Begriffe *Elternteil* und *Bruder* wie folgt definieren:
 $\llbracket \text{Onkel} \rrbracket$ definiert durch $\llbracket \text{Elternteil} \rrbracket$ und $\llbracket \text{Bruder} \rrbracket$:
 $\lambda x\lambda y[\text{es gibt ein } z \text{ mit } \llbracket \text{Elternteil} \rrbracket(x)(z) \text{ und } \llbracket \text{Bruder} \rrbracket(y)(z)]$,
 $(y \text{ ist Onkel von } x \text{ gdw. es ein } z \text{ gibt sodass } z \text{ Elternteil von } x \text{ und } y \text{ Bruder von } z \text{ ist}).$
 Es gilt z.B. $\llbracket \text{Onkel} \rrbracket(\llbracket \text{Maria} \rrbracket)(\llbracket \text{Hans} \rrbracket)$, wenn Hans der Onkel von Maria ist, wenn gilt:
 $\lambda x\lambda y[\text{es gibt ein } z \text{ mit } \llbracket \text{Elternteil} \rrbracket(x)(z) \text{ und } \llbracket \text{Bruder} \rrbracket(z)(y)](\llbracket \text{Maria} \rrbracket)(\llbracket \text{Hans} \rrbracket)$
 $= \lambda y[\text{es gibt ein } z \text{ mit } \llbracket \text{Elternteil} \rrbracket(\llbracket \text{Maria} \rrbracket)(z) \text{ und } \llbracket \text{Bruder} \rrbracket(z)(y)](\llbracket \text{Hans} \rrbracket)$
 $= [\text{es gibt ein } z \text{ mit } \llbracket \text{Elternteil} \rrbracket(\llbracket \text{Maria} \rrbracket)(z) \text{ und } \llbracket \text{Bruder} \rrbracket(z)(\llbracket \text{Hans} \rrbracket)]$
 $= [\text{es gibt ein } z, \text{ wobei } z \text{ ein Elternteil von Maria ist, und Hans ein Bruder von } z \text{ ist}]$
 Definieren Sie nach derselben Methode die folgenden Bedeutungen:
 a) $\llbracket \text{Nichte} \rrbracket$, durch $\llbracket \text{Geschwister} \rrbracket$, $\llbracket \text{Tochter} \rrbracket$
 b) $\llbracket \text{Kusine} \rrbracket$, durch $\llbracket \text{Elternteil} \rrbracket$, $\llbracket \text{Geschwister} \rrbracket$, $\llbracket \text{Tochter} \rrbracket$
 c) $\llbracket \text{Enkelin} \rrbracket$, durch $\llbracket \text{Elternteil} \rrbracket$ und $\llbracket \text{weiblich} \rrbracket$.
 d) $\llbracket \text{Schwiegermutter} \rrbracket$, durch $\llbracket \text{Ehegatte} \rrbracket$ und $\llbracket \text{Mutter} \rrbracket$.

sei nach Italien **und** Spanien gefahren, kann der Hörer schließen, dass diese Aussage nicht wahr wäre. Er kann also die wörtliche Bedeutung des Satzes (inklusive Negation) durch Implikatur verstärken zur exklusiven Negation. Dass es sich dabei um eine Implikatur handelt, sieht man daran, dass diese Bedeutungskomponente wieder zurückgenommen werden kann:

(30) *Maria ist nach Italien gefahren, oder nach Spanien, vielleicht sogar in beide Länder.*

Auf den DM-Scheinen war zu lesen:

(31) *Wer Banknoten fälscht oder gefälschte Banknoten in den Verkehr bringt, wird mit Zuchthaus nicht unter zwei Jahren bestraft.*

Wenn wir das *oder* als semantisch exklusiv verstehen würden, dann hätte ein Banknoten-fälscher, der seine gefälschten Banknoten auch in den Verkehr gebracht hat, argumentieren können, dass das Gesetz auf ihn gar nicht anwendbar ist.

(32) *Wer*

(33) **Materiale Implikation, Konditional**

Φ	Ψ	$[\Phi \rightarrow \Psi]$
0	0	1
0	1	1
1	0	0
1	1	1

Das Konditional $[\Phi \rightarrow \Psi]$ besagt: Wenn der Vordersatz Φ wahr ist, dann ist auch der Folgesatz Ψ wahr. (Vorder- und Folgesatz heißen auch **Antezedens** und **Konsequens**). Daher ist die Implikation falsch, wenn das Antezedens Φ wahr und das Konsequens Ψ falsch ist. Wenn das Antezedens nicht wahr ist, dann ist jeder Wahrheitswert des Konsequens zugelassen. Die materiale Implikation ist in diesem Fall sicherlich nicht falsch, und da es nur einen anderen Wahrheitswert gibt, ist sie also wahr.

Dies entspricht nur sehr grob den Konditionalsätzen in der natürlichen Sprache. Demnach wäre ein Satz wie

(34) *Wenn es regnet, bin ich Millionär*

automatisch wahr, wenn es nicht regnet. Und der folgende Satz wäre eine Tautologie:

(35) *Wenn $2+2 = 5$ ist, bin ich Millionär.*

Dies ist natürlich nicht der Fall. Konditionalsätze gehören zu den semantisch komplexesten Konstruktionen, die die natürliche Sprache kennt, und die einfache materiale Implikation kann diese nicht erfassen.

(36) **Materiale Äquivalenz, Bikonditional**

Φ	Ψ	$[\Phi \leftrightarrow \Psi]$
0	0	1
0	1	0
1	0	0
1	1	1

Das Bikonditikonale $[\Phi \leftrightarrow \Psi]$ ist also genau dann wahr, wenn beide Sätze den gleichen Wahrheitswert haben.

Mithilfe dieser Regeln kann man nun den Wahrheitswert eines komplexen Satzes bestimmen, wenn die Wahrheitswerte der einfachen Sätze bekannt sind. Wir verdeutlichen

dies mit dem Beispiel (24.e), unter der Annahme, dass p_1 wahr, p_2 falsch und p_3 wahr ist. In dem folgenden Beispiel wird der Wahrheitswert des Gesamtsatzes Schritt für Schritt berechnet.

(37)

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

Eine andere Möglichkeit, die Berechnung des Wahrheitswerts graphisch darzustellen, ist die folgende:

$$(38) \begin{array}{ccccccc} p_1 \wedge & [[& p_3 & \rightarrow & \neg & [& \neg p_1 & \vee & p_2]]] \\ 1 & & 1 & & & 1 & & 0 & \\ & & & & & & & 0 & \\ & & & & & & & & 1 & \\ & & & & & & & & & 1 & \\ & & & & & & & & & & 1 \end{array}$$

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:

$$(39) \begin{array}{ccc} \neg [p_1 \wedge p_2] & & [\neg p_1 \wedge p_2] \\ 1 & 0 & 0 \\ 0 & & 0 \\ 1 & & 0 \end{array}$$

Der linke Satz erhält den Wahrheitswert wahr, der rechte falsch; die beiden Sätzen bestehen zwar aus denselben Zeichen, sie sind jedoch unterschiedlich geklammert.

16.3 Tautologien und Kontradiktionen

Bei (38) handelte es sich um einen kontingenten Satz. Unter einer anderen Zuweisung von Wahrheitswerten kann es sein, dass der Gesamtsatz falsch ist – etwa wenn p_1 falsch, p_2 wahr und p_3 wahr ist, dann ist der Gesamtsatz falsch. (Nachrechnen!)

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.

$$(40) p_1 \vee [[p_3 \rightarrow \neg [p_1 \wedge p_2]]]$$

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, wie die folgende Betrachtung aller Fälle zeigt.

(41)

p ₁	p ₂	p ₃	[p ₁ ∧p ₂]	¬[p ₁ ∧p ₂]	[p ₃ →¬[p ₁ ∧p ₂]]	[p ₁ ∨[p ₃ →¬[p ₁ ∧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.

16.4 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 Φ wahr ist, ist auch Ψ 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.

(42)

p ₁	p ₂	[p ₂ → p ₁]	¬p ₁	[¬p ₁ ∧ [p ₂ → p ₁]]	⇒	¬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:

- (43) a. Die logische Folgerung $\Phi \Rightarrow \Psi$ besteht gdw. $[\Phi \rightarrow \Psi]$ eine Tautologie ist.
 b. 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:

(44)

p ₁	p ₂	[p ₂ → p ₁]	¬p ₁	[¬p ₁ ∧ [p ₂ → p ₁]]	¬p ₂	[[¬p ₁ ∧ [p ₂ → p ₁]] → ¬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:

(45) Zeige: $\neg[\neg p_1 \wedge p_2] \Leftrightarrow [\neg p_2 \vee p_1]$

Beweis: Wir zeigen: $[\neg[\neg p_1 \wedge p_2] \leftrightarrow]$ ist eine Tautologie.

p ₁	p ₂	¬p ₁	[¬p ₁ ∧p ₂]	¬[¬p ₁ ∧p ₂]	¬p ₂	[¬p ₂ ∨p ₁]	[¬[¬p ₁ ∧p ₂]↔ [¬p ₂ ∨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 Φ eine Kontradiktion ist, d.h. immer falsch ist, dann ist der Satz $[\Phi \rightarrow \Psi]$ immer wahr, egal welche Wahrheitswerte Ψ annimmt. Das folgt aus der Wahrheitstafel für das Konditional, \rightarrow . Dann aber ist nach (43) 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.

16.5 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:

- (46) 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 eigentlich mit weniger Satzkonnectoren auskommt, als gemeinhin anzunehmen: 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:

- (47) 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 (45) kann nun wie folgt gelöst werden:

- (48) $\neg[\neg p_1 \wedge p_2]$
 $\Leftrightarrow [\neg\neg p_1 \vee \neg p_2]$ (Gesetz von de Morgan)
 $\Leftrightarrow [p_1 \vee \neg p_2]$ (Doppelte Negation)
 $\Leftrightarrow [\neg p_2 \vee p_1]$ (Kommutativität)

16.6 Aussagenlogik und Bedeutung von Aussagesätzen

Die Aussagenlogik sagt nichts über die Bedeutung von Konstituenten unterhalb der Satzebene, also über Wörter, Nominalphrasen usw. Sie kann also kein Modell für die Bedeutung von Ausdrücken der natürlichen Sprache sein. Kann sie aber zumindest ein Modell für die Bedeutung von Aussagesätzen sein?

Ein erster Versuch ist, zu sagen: Die Bedeutung eines Aussagesatzes ist sein Wahrheitswert. Das kann aber nicht sein: Wir haben nur zwei Wahrheitswerte, und damit hätten wir nur zwei mögliche Satzbedeutungen.

Ein zweiter Versuch ist raffinierter: Er besagt, dass alle Sätze, die logisch äquivalent sind, die gleiche Bedeutung haben. Ein Beispiel:

(49) *Peter kaufte ein Auto von Maria.* \Leftrightarrow *Maria verkaufte ein Auto an Peter.*

Es ist sicherlich so, dass wenn immer der erste Satz wahr ist, dann auch der zweite Satz wahr ist, und umgekehrt, sofern überhaupt die normalen Interpretationsregeln des Deutschen gelten.

Das Kriterium der logischen Äquivalenz für Gleichheit der Bedeutung erfasst die Idee der Wahrheitsbedingungen: Zwei Sätze sind logisch äquivalent, wenn sie dieselben Wahrheitsbedingungen haben. Erfasst es aber auch den intuitiven Begriff der Bedeutung? Das ist nicht der Fall; zwei logisch äquivalente Sätze können durchaus als bedeutungsverschieden angesehen werden. Dies ist besonders deutlich bei Tautologien und Kontradiktionen: Der Satz *Es regnet oder es regnet nicht* hat sicher nicht dieselbe Bedeutung wie der Satz *Zwei plus zwei ist vier*. Dies weist auf eine weitere Begrenzung des logischen Zugangs zur Bedeutung, und ganz allgemein der Zugrundelegung von Wahrheitsbedingungen für die Semantik, hin.

16.7 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 “ Φ und Ψ sind äquivalent”, “ Φ und Ψ sind konträr”, und “ Φ und Ψ 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: $p_1: 0, p_2: 1, p_3: 0, p_4: 1$
 $[\neg [[p_1 \vee p_2] \wedge \neg p_4] \rightarrow [p_1 \vee \neg p_3]]$
- Welche der folgenden aussagenlogischen Formeln sind Tautologien, Kontradiktionen oder kontingente Sätze?
 - $[[p_1 \wedge p_2] \rightarrow \neg p_2]$
 - $[[p_1 \vee p_2] \rightarrow \neg p_2]$
 - $[p_1 \wedge \neg[p_1 \vee p_2]]$
 - $[[p_1 \vee p_2] \wedge [p_2 \rightarrow p_1]]$
 - $[\neg[p_1 \vee p_2] \rightarrow \neg p_2]$
 - $[[p_1 \rightarrow p_2] \rightarrow [p_2 \rightarrow p_1]]$
- 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.*

17. Werkzeuge der Modellierung: Aussagen- und Prädikatenlogik