

Universität Regensburg
Institut für Germanistik
Lehrstuhl für Deutsche Sprachwissenschaft

**Arbeitstipp
Empirisches Arbeiten
im Fach Deutsche Sprachwissenschaft**

1 Ausgangsüberlegungen

Was genau soll untersucht werden?

Der Untersuchungsgegenstand muss exakt festgelegt werden. Hilfreich ist meist die Formulierung einer Forschungsfrage/ Hypothese.

ACHTUNG: Wichtig ist, das Thema der Untersuchung im Vorfeld möglichst exakt einzugrenzen (Was untersuche ich? In welche kleineren und größeren wissenschaftlichen Rahmenbedingungen und Abhängigkeiten ist meine Untersuchung eingebettet?) oder es spätestens in Anpassung an die erhobenen Daten zu konkretisieren (**Mögliche**

Problemfälle: Meine Daten sind zum Thema meiner Fragestellung nur bedingt aussagekräftig [nur zu einem Teilbereich meiner Fragestellung] => Fragestellung konkretisieren; meine Daten stehen in weiteren Abhängigkeiten über meine Fragestellung hinaus => Abhängigkeiten abklären und zu erfassen versuchen; meine Daten sind aufgrund ihrer Erhebung nicht für diese Fragestellung als Untersuchungsgegenstände geeignet => Fragestellung möglichst im Rahmen des größeren Themas abändern).

Zu welchem Zweck sollen diese Daten erfasst werden?

Die empirische Untersuchung muss so angelegt sein, dass die Daten nachher mit Bezug auf die Fragestellung valide interpretierbar sind.

Die **konkrete Durchführung** einer empirischen Untersuchung gliedert sich in **vier** aufeinander folgende **Schritte**¹:

- a) Ausrichtung der empirischen Daten an der **Hypothese:** Wie und wo kann ich meine Untersuchung empirisch fassen?
- b) **Erhebung der empirischen Daten.**
- c) **Verarbeitung** und Aufbereitung des erhobenen Materials: Z.B. die Übertragung Quantitativer Daten in Tabellen und Graphiken.
- d) **Analyse** der Daten, Ableitung von Schlussfolgerungen, die aus dem Datenmaterial gezogen werden können, und Test der postulierten Hypothese(n). Erfassen von eventuellen Problemfällen.

¹ Vgl. Schlobinski: Empirische Sprachwissenschaft, 20-21.

2 Einige Methoden empirischer Forschung

2.1 Die nicht-teilnehmende Beobachtung

V.a. in der Konversationsanalyse angewandt eignet sich dieses Verfahren besonders zur Analyse gesprochener Sprache, da die zu befragenden Personen frei agieren können.

2.2 Die Befragung

2.2.1 Arten der Befragung

a) Unterscheidung nach **Situationen**, in der sich die **befragte Person** befindet:

- 1) **Mündliche Befragung:** Einer Person werden im persönlichen Gespräch mündlich Fragen gestellt [= Interview].
- 2) **Schriftliche Befragung:** Eine Person antwortet auf schriftlich gestellte Fragen schriftlich [= Fragebogen].

b) Unterscheidung nach **Antwortmöglichkeiten:**

- 1) **Offene Konzepte (explorative Interviews):** Durch die Möglichkeit der Befragten, frei zu antworten und eigene Gedanken und Formulierungen einzubringen, kann ein Überblick über verschiedene Teilaspekte eines Themas gewonnen werden. Deshalb haben diese meist vorbereitenden Charakter.

Beispiel für eine offene Frage: „*Wie gefällt Ihnen die vorliegende Grammatik?*“

- 2) **Geschlossene Konzepte:** Durch entweder mündlich (Interview) oder schriftlich (Fragebogen) vorher vorgegebene Fragenkataloge werden die Befragten in ihren Antwortmöglichkeiten beschränkt. Geschlossene Konzepte können also nur Meinungen der Befragten zu vorgegeben Aspekten widerspiegeln.

Beispiel für eine geschlossene Frage:

„Was bedeutet Ihrer Meinung nach das Verb ‚verfilmen‘?“

etw. bis zu Ende filmen ()

etw. falsch filmen ()

etw. zu einem Film verarbeiten ()

etw. mit einem Film versehen ()“

2.2.2 Wahl der Betroffenen und Größe der Stichprobe

Da es grundsätzlich unmöglich ist, alle Betroffenen, über die man eine Aussage treffen will (= **Grundgesamtheit**), tatsächlich zu befragen, muss ein Teil der Betroffenen (= **Stichprobe**), ausgewählt werden, um eine möglichst generelle Aussage zu einem Untersuchungsgegenstand treffen zu können. Da empirische Analysen im Rahmen studentischer Seminar- und Abschlussarbeiten in der Regel immer exemplarischen Charakter haben, ist eine möglichst gezielte Auswahl der Betroffenen wichtiger bzw. den Aufbau des Arguments stützender als die Untersuchung möglichst vieler Personen.

- 1) **Zufallsstichprobe:** Zufällige Aufnahme der Mitglieder einer Grundgesamtheit ohne Berücksichtigung weiterer Aspekte (explorativ-vorbereitend, nicht argumentativ-stützend!)
- 2) **Quotenverfahren:** Auswahl der Mitglieder einer Grundgesamtheit nach bestimmten, für die angestrebte Untersuchung relevanten Merkmalen (z.B. Alter, Schulabschluss, sozialer Stand, Studienfach, Geschlecht, Stadt-/ Landbevölkerung; **ACHTUNG:** Relevanz der Auswahlkriterien für die Untersuchung!).

2.2.3 Der Fragebogen

a) Phasen der Fragebogenentwicklung²:

- 1) **Fokussierung des Themas**, Aufstellung von Hypothesen, mündliche oder schriftliche Befragung, Auswahlkriterien zur Begrenzung der Gruppe der befragten Personen.
- 2) **Formulierung von Fragen zu den Hypothesen:** Verständlichkeit und Eindeutigkeit, Festlegung auf einen Fragetyp (geschlossen vs. offen; evtl. Mischung beider Fragetypen).
- 3) **Anordnung der Fragen** in eine Reihenfolge.
- 4) **Überprüfung des Fragebogens** durch eine Testperson zur Feststellung von Ungenauigkeiten der Fragenformulierung.

b) Aufbau des Fragebogens:

- 1) Sog. **Sekundärdaten** (Sozialdaten der befragten Person, z.B. Geschlecht, Alter, Schulabschluss, Beruf etc.) ganz an den Anfang stellen.
- 2) Fragen, die das **Interesse** des/der Befragten wecken, an den Anfang, eher schwierig zu beantwortende Fragen ans Ende stellen.
- 3) Der Fragebogen darf **keine Ordnung hinsichtlich des Ergebnisses** erkennen lassen, das der **Fragesteller selbst annimmt**.
- 4) Alle **Informationen**, die für die **Interpretation des Ergebnisses maßgeblich** sind, dürfen **nicht aus den Inhalten** der Fragen **erschlossen** werden können.
- 5) **Vermeidung** des sog. **Halo-Effektes** durch überlegte Abfolge der Fragen: Die Beantwortung einer Frage darf nicht abhängig von der Beantwortung einer Vorgängerfrage sein.

2.3 Arbeiten mit Textkorpora

2.3.1 Definitionsvorschlag Korpus:

[Achtung: das Korpus, die Korpora]

„Sammlung einer möglichst hohen, notwendigerweise aber immer begrenzten Anzahl möglichst zusammenhängender sprachlicher Äußerungen (gesprochen oder/und geschrieben) aus möglichst **natürlichen Kommunikationssituationen**. Aufbereitet sind solche Korpora wichtige Hilfsmittel z.B. in der Lexikografie [...]. Korpora dienen zur

² Vgl. Atteslander: Methoden der empirischen Sozialforschung, 175.

Ermittlung und **Beschreibung sprachlicher Regularitäten** bzw. zur **Überprüfung von Hypothesen und Theorien** und sind Grundlage der Korpusanalyse.“³

2.3.2 Digitalisierte Textkorpora im Internet

Je nach Fragestellung eines linguistischen Themas kann auf das Material zurückgegriffen werden, das einem breiten Benutzerkreis von verschiedenen Institutionen für wissenschaftliche Untersuchungen in Form eines digitalisierten Korpus zur Verfügung gestellt worden ist. Dabei muss überprüft werden, ob die Rahmensituationen und die Darstellung der fremden Daten hinsichtlich der eigenen Forschungsfrage/ Hypothese verwendet werden können.

2.3.2.1 DeReKo: Das DEUTSCHE REFERENZKORPUS

- 1) **Entwickler und Betreuer:** Institut für Deutsche Sprache (IDS);
- 2) **Umfang:** 3,2 Milliarden Textwörter insgesamt, 2,2 Milliarden Textwörter online zugänglich;
- 3) **Quellen:** Deutschsprachige geschriebene Texte aus der Gegenwart und der jüngeren Vergangenheit (Neuhochdeutsch!);
- 4) **Aufbau:** Frei wählbare Teilkorpora, u.a. Frankfurter Rundschau, Goethe-Korpus, Grammatik-Korpus, GRIMM-Korpus, Korpus Magazin Lufthansa Bordbuch, Wendekorpus, komplette Einzelübersicht (<http://www.ids-mannheim.de/kl/projekte/korpora/archiv.html>);
- 5) **Anmeldung:** <https://cosmas2.ids-mannheim.de/cosmas2-web/>, kostenlos, grundsätzliche Voraussetzung für Online-Recherchen;
- 6) **Service:** Z.B. Überführung von Recherchedaten in Word-Dokumente;
- 7) **Recherche-Programm:** Cosmas II.

2.3.2.2 AGD: Das ARCHIV FÜR GESPROCHENES DEUTSCH

- 1) **Entwickler und Betreuer:** Institut für Deutsche Sprache (IDS);
- 2) **Umfang:** 28 Einzelkorpora aus Einzelspracherhebungen und Forschungsprojekten, in Erweiterung begriffen (jüngstes Projekt = „Deutsch heute“, http://www.ids-mannheim.de/prag/AusVar/Deutsch_heute/);
- 3) **Quellen:** Texte von gesprochenem Deutsch in Tonaufnahmen und Transkripten binnen- und auslanddeutscher Varietäten und Dialekte (Onlinezugang), Materialien zum gesprochenen Standarddeutsch, Umgangssprachen und aus verbalen Interaktionen in verschiedenen sozialen und situativen Kontexten (Archiv);
- 4) **Aufbau:** Frei wählbare Teilkorpora in der Datenbank für Gesprochenes Deutsch (DGD), u.a. Deutsche Mundarten (DDR), Deutsche Umgangssprachen (Pfeffer-Korpus), Emigrantendeutsch in Israel;
- 5) **Anmeldung:** <http://dsav-wiss.ids-mannheim.de/>, kostenlos, Möglichkeit zur Online-Recherche in einem Teilbestand der Tonaufnahmen;

³ Glück: Metzler-Lexikon Sprache, 357-358. Hervorhebungen durch den Autor.

- 6) **Service:** Z.B. Überführung von Recherchedaten in Word-Dokumente, Lieferservice von Kopien von Tonaufnahmen, Transkripten und Begleitmaterialien aus dem Bestand (kostenpflichtig).

2.3.2.3 Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS)

- 1) **Entwickler und Betreuer:** Berlin-Brandenburgische Akademie der Wissenschaften;
- 2) **Umfang:** Versch. Einzelkorpora: DWDS-Kernkorpus [122.816.010 Textwörter aus Schöner Literatur (ca. 26%), Journalistischer Prosa (ca. 27%), Fachprosa (ca. 22%), Gebrauchstexten (ca. 20%), Transkribierten Texte gesprochener Sprache (ca. 5%)], Juilland-D-Korpus (500.000 Textwörter), Corpus Gesprochene Sprache (ca. 2.500.000 Textwörter aus Reden, Rundfunkansprachen, Auszügen aus Parlamentsreden), DDR-Corpus (9.000.000 Textwörter aus in der DDR zwischen 1949 und 1990 erschienen Texten), Corpus Jüdischer Periodika (26.247.390 Textwörter), Korpus Anonymisierung von Eigennamen (Bearbeitungen eigentlich urheberrechtlich geschützter Texte), Zeitungskorpora z.B. das ZEIT-Korpus (ca. 106.000.000 Textwörter in den Zeitausgaben ab 1996), Berliner Wendekorpus (Befragung von 39 Ostberlinern und 38 Westberlinern in narrativen Interviews zu den Ereignissen des 9. November und anderen „Wende-Ereignissen“, nähere Informationen unter <http://www.dwds.de/textbasis/>);
- 3) Quellen: Je nach Teilkorpus verschieden;
- 4) Aufbau: Zugriff auf DWDS-Kernkorpus, Auswertung z.T. nach Teilkorpora;
- 5) Anmeldung: <http://www.dwds.de/>, kostenlos, verbesserte Möglichkeit der Online-Recherche;
- 6) Service: Überführung von Ergebnisdaten in andere Formate möglich.

2.3.2.3 Weitere Korpora

Weitere Korpora finden sich im Internet, beispielsweise das „Wortschatz-Lexikon“ der Universität Leipzig: <http://wortschatz.uni-leipzig.de/>.

2.3.3 Textkorpora auf CD-Rom

Textausgaben, die auf CD-Rom gespeichert sind, werden von Universitätsbibliotheken für Recherchezwecke kostenlos zur Verfügung gestellt.

Der CD-ROM-Server der Universität Regensburg stellt u.a. folgende Korpora zur Verfügung:

a) Einzelwerke der Weltliteratur:

- Bertolt Brechts Werke: Jubiläumsausgabe auf CD-ROM
- Deutsche Literatur von Luther bis Tucholsky
- Goethes Werke auf CD-ROM.

b) Zeitungen und Zeitschriften:

- FAZ auf CD-ROM *
- Mittelbayerische Zeitung *
- Neue Zürcher Zeitung *
- Der Spiegel *
- Süddeutsche Zeitung *

- Die Zeit *

c) Nachschlagewerke und Wörterbücher:

- Der Brockhaus in Text und Bild
- Brockhaus-Enzyklopädie Online
- Atlas Deutsche Literatur
- Duden Deutsches Universal-Wörterbuch A - Z *
- Langenscheidts Taschenwörterbuch Französisch : französisch-deutsch *
- Oxford English Dictionary.

[Die mit * gekennzeichneten Korpora sind nur an Einzelplätzen im Info-Zentrum der Zentralbibliothek zugänglich.]

2.3.4 Materialgewinnung für eigene Untersuchungen

Auch eine eigene Zusammenstellung geeigneten Sprachmaterials bezeichnet man als Korpus. Das Erstellen eigener Textkorpora ist sehr zeitaufwändig und erfordert eine eigene Beschäftigung damit, kann aber bezüglich folgender Kriterien durchgeführt werden:

- a) Repräsentative Textmenge
- b) Zeitliche und räumliche Ausdehnung der Texte
- c) Textsorten
- d) Sprachschichten und Sprachvarietäten innerhalb der Texte.

3 Literaturhinweise

Atteslander, Peter (⁹2009): Methoden der empirischen Sozialforschung. Berlin.

Bergmann, Rolf/Peter Pauly/Stricker, Stefanie (⁴2005): Einführung in die deutsche Sprachwissenschaft. Heidelberg, 217-225.

Brons-Albert, Ruth/Koster, Cor J. (2002): Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch. Tübingen.

Glück, Helmut (Hrsg.) (³2005): Metzler-Lexikon Sprache. Stuttgart, Weimar: Verlag J.B. Metzler. [Einträge: Datenerhebung; Empirie; Empirische Sprachwissenschaft; Interview; Korpus; Korpusanalyse].

Schlobinski, Peter (1996): Empirische Sprachwissenschaft. Opladen.

Empfehlenswert sind außerdem:

Lemnitzer, Lothar/Zinsmeister, Heike (2010): Korpuslinguistik. Eine Einführung. Tübingen.

Scherer, Carmen (2006): Korpuslinguistik. Heidelberg.

Handreichung zur Benutzung des DWDS-Korpus

Vorgehensschritte:

- Anmeldung (Erweiterte Funktionen: Zusammenstellung eigenes Korpus)
- Bloße Stichworteingabe
- Komplexe Suchanfrage => Operatoren:

&&	=	UND	
	=	ODER	
&& !	=	NICHT	
@	=	nur diese Wortform	
#n	=	Maximalabstand	(alle Wörter zählen!)
*	=	Platzhalter	(0 bis unendlich)
„“	=	Komplexe Abfrage	
- Bei Setzung mehr als eines Operators: Klammerung
- Beispielsuchanfrage: Phraseologismus *Haus und Hof* => „@Haus #3 Hof“
- Tipp: Eingabeprobleme resultieren oft aus Setzung der Leerzeichen

Anzeige/ Ergebnisse:

- Panel-Anzeige:

1=	DWDS-Wörterbuch
2=	Etymologie
3=	Openthesaurus
4+6=	DWDS-Kern-/Zeitungskorpus
5=	Kollokationsstatistik [+ Relationenfilter]

 Hinzufügung möglich: z.B. Wortverlaufsstatistik, Korpusfrequenzen)
- Problem: Bei komplexeren Suchanfragen v.a. mit zusätzlicher morphosyntaktischer Spezifikation sind die meisten Zusatzfunktionen nicht mehr gegeben!

Arbeitsmöglichkeiten mit dem DWDS-Korpus:

- Gutes heuristisches Mittel semantisch/ soziolinguistisch/ morpho-syntaktisch!
- Schlechtes exaktes morpho-syntaktisches Auswertungsmittel!

Handreichung zur Benutzung von COSMAS II

Welche Arbeitsschritte sind bei der Arbeit mit COSMAS II zu berücksichtigen?

- Anmeldung + Login: <https://cosmas2.ids-mannheim.de/cosmas2-web/> -> „Anmeldung/ Login“, „Registrieren/ Recherche“
- Archivauswahl: Erste Auswahl der Suchtexte, vordefiniert
- Korpusauswahl: Zweite Auswahl der Suchtexte, vordefiniert oder benutzerorientiert -> „benutzerdefiniert/ Korpusverwaltung“
- Formulierung der Suchanfrage: Frage bestimmt wesentlich Ergebnis! -> „Suchen“ (Online-Hilfe „Suchanfragesyntax“ als Read-me-Datei)
- Ergebnisstrukturierung: Zweite wesentliche Eingrenzung der Suchergebnisse/ Überblick -> „Öffnen“
- Ergebnisauswertung: Verschiedene Anzeigemöglichkeiten und Schachtelung je nach Kontextmasse: KWIC-Anzeige (= key-word-in-context) -> „Ergebnisse/ KWIC“
- Ergebnisexport: Einzelauswahl und Export als Word-Datei per E-Mail -> „Export“

Welche Suchanfragen können formuliert werden?

- Einfache Suchanfrage: Ohne Einschränkung über Annotation
- Morphologisch-syntaktische Suchanfrage: Annotation über Flexionsformen und Stellung im Satz
- Kookkurrenzsuchanfrage: Annotation nach Umgebung der Suchwörter geordnet (semantisch)

Wie können Suchanfragen formuliert werden?

- * = beliebig viele Zeichen im gleichen Wort vorher, z.B. Präfigierungen
- & = beliebig viele Zeichen im gleichen Wort vorher und nachher, auch verändert, z.B. Flexionsformen
- Wortabstand: /+w(Zahl1):(Zahl2) = (Zahl1) Wörter vorher und (Zahl2) Wörter nachher berücksichtigen, z.B. Kombination zweier Wörter zu einer Suchanfrage
- Satzabstand: /+s(Zahl1):(Zahl2) = (Zahl1) Sätze vorher und (Zahl2) Sätze nachher berücksichtigen, z.B. Wiederauftreten eines Wortes im Satzabstand

Welche Ergebnisse erhalte ich aus den Suchanfragen?

- Ergebnisanzeige: Verschiedene Treffer in Anzahl von Texten aus vielen bestimmten Korpora
- Kontextueller Umfang: Satz mit Suchwort „+“, bzw. „Volltext“
- Ergebnisexport: Word-Datei mit Zusatzinformationen zur Suchanfrage, eventuell auch zum Korpus
- Ergebnispräsentation: Je nach Dateiauswahl beim „Export“

Welche Grenzen und Möglichkeiten bieten Untersuchungen mit COSMAS II?

- Suchanfrage determiniert Ergebnis => eher Theorie stützendes (gezielte Suchanfrage) als heuristisches Mittel (ungezielte Suchanfrage)
- Fundierung theoretischer Feststellung durch realistische Sprachdaten!