

Ruth Albert / Cor J. Koster

Empirie in Linguistik und Sprachlehrforschung

Ein methodologisches Arbeitsbuch

gnV Gunter Narr Verlag Tübingen

Albert, Ruth:

Empirie in Linguistik und Sprachlehrforschung : ein methodologisches Arbeitsbuch /

Ruth Albert ; Cor J. Koster. – Tübingen : Narr, 2002

(Narr Studienbücher)

ISBN 3-8233-4985-6

© 2002 · Gunter Narr Verlag Tübingen

Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Gedruckt auf chlorfrei gebleichtem und säurefreiem Werkdruckpapier.

Internet: <http://www.narr.de>

E-Mail: info@narr.de

Druck: Gulde, Tübingen

Verarbeitung: Nädele, Nehren

Printed in Germany

ISSN 0941-8105

ISBN 3-8233-4985-6

Inhaltsverzeichnis

0	Zur Einführung	1
	Aufgabe	10
1	Gütekriterien für empirische Untersuchungen	11
	1.1 Verlässlichkeit (Reliabilität)	12
	1.2 Gültigkeit (Validität)	13
	1.3 Geltungsbereich	14
	Aufgaben	16
2	Die Beobachtung	17
	Aufgaben	23
3	Die Befragung	24
	3.1 Die Wahl der Stichprobe	27
	3.1.1 Die Größe der Stichprobe	29
	3.2 Befragungsarten	30
	3.2.1 Offene Konzepte, explorative Interviews	30
	3.2.2 Geschlossene Konzepte, festgelegte Fragefolgen	32
	3.2.3 Offene und geschlossene Fragen	34
	3.2.4 Direkte und indirekte Fragen	35
	3.3 Aufbau eines Fragebogens	36
	3.4 Umgang mit der Gefahr von Artefakten	37
	3.5 Kodierung der Antworten, Auswertungsmöglichkeiten	39
	Aufgaben	43
4	Das Experiment	46
	4.1 Experimentelle Forschung	46
	4.2 Der Entwurf des Forschungsvorhabens	48
	4.2.1 Vorbereitungen für ein Experiment	51
	4.2.2 Labor- vs. Feldexperiment	52
	4.2.3 Kontrollgruppen	53
	4.3 Mehrfaktorielle Fragestellungen	56
	4.4 Ergebnisdarstellung	59
	Aufgaben	67
5	Arbeiten mit Textkorpora	68
	Aufgabe	72
6	Beschreibende Statistik	74
	6.1 Messskalen	74
	6.2 Streuung, Mittelwert, Standardabweichung	77
	6.3 Häufigkeiten, relative Häufigkeit	80
	6.4 Die Darstellung der Daten	81
	Aufgaben	85
7	Korrelationen	87
	7.1 Kalkulation des Korrelationskoeffizienten	90
	7.2 Wann ist eine Korrelation hoch genug?	93
	7.3 Korrelation bedeutet keine Kausalität	95
	7.4 Wie man Korrelationen präsentiert	97
	Aufgaben	98

8	Das Testen des Tests	100
8.1	Geschlossene Tests	100
8.1.1	Einheitenanalyse (item analysis)	101
8.2	Offene Tests	104
8.3	Gemischte Tests	105
8.4	Validität	105
8.5	Verlässlichkeit (Reliabilität)	106
	Aufgaben	110
9	Wann sind Unterschiede bedeutsam?	111
9.1	Der t-Test	112
9.1.1	t-Test für korrelierende Stichproben	112
9.1.2	t-Test für unabhängige Gruppen	114
9.1.3	Annahmen, die der t-Test voraussetzt	117
9.1.4	Wie man die Ergebnisse eines t-Tests präsentiert	118
9.2	Der Chi-Quadrat-Test	118
9.2.1	Wie man Chi-Quadrat-Ergebnisse präsentiert	127
9.3	ANOVA: Varianzanalyse (analysis of variance)	128
9.3.1	ANOVA: Beispiel 1	128
9.3.2	ANOVA: Beispiel 2	131
9.4	Wann man welchen Test benutzt	134
9.5	Signifikanz und Aussagekraft	135
9.5.1	Interpretation des Signifikanzniveaus	135
9.5.2	Erklärte Varianz	137
9.5.2.1	Korrelation und r-Quadrat	137
9.5.2.2	t-Test und Omega zum Quadrat	138
9.5.2.3	Chi-Quadrat und Cramér's V	139
9.5.2.4	ANOVA und Eta zum Quadrat	140
9.6	Weiterführende Literatur	141
	Aufgaben	142
10	Statistik- und Tabellenkalkulationsprogramme	144
10.1	Freeware-/Shareware-Statistikprogramme	144
10.2	Das Tabellenkalkulationsprogramm Excel	145
10.3	Berechnung von Korrelationen	147
10.4	t-Test für korrelierende Stichproben	148
10.5	t-Test für zwei unabhängige Stichproben	151
10.6	Chi-square	152
10.7	ANOVA	153
11	Lösungen der Aufgaben	156
	Anhänge	177

Vorwort

Wir richten uns an Studierende der Geisteswissenschaften, die ihre Examens- oder Doktorarbeit schreiben und haben uns deshalb auf die Beschreibung der Verfahren beschränkt, die Geisteswissenschaftler tatsächlich benutzen. Dabei zeigen wir neben den Rechenverfahren "von Hand" die Benutzung von Computerprogrammen, vor allem die Statistik-Anwendungen im verbreiteten Tabellenkalkulationsprogramm Excel.

Dieses Buch kann keine Wunder wirken. Es ersetzt nicht, mit dem/der Betreuer(in) der Arbeit zu besprechen, was genau untersucht werden soll, welche Methoden dabei zur Datenerhebung eingesetzt werden können und wie bei der Analyse der erhobenen Daten vorzugehen ist. Wir vermitteln einen ersten Eindruck, erklären häufig benutzte Verfahren und warnen vor häufig vorkommenden Fehlern.

Wenn man ein Buch für Personen mit recht verschiedenem Hintergrundwissen schreibt, dann tut man gut daran, bei der Darstellung der einzelnen Methoden Beispiele zu wählen, die man verstehen kann, ohne dass vorher komplizierte linguistische Hypothesen oder lerntheoretische Annahmen erklärt werden müssen. Deshalb haben wir hier mit möglichst einfachen und teils auch erfundenen Beispielen gearbeitet, die man mit Alltagswissen verstehen kann und in Kauf genommen, dass diese Beispiele gelegentlich banal erscheinen mögen.

Die einzelnen Kapitel enden jeweils mit Übungsaufgaben, die es Ihnen ermöglichen sollen zu überprüfen, ob Sie das Gelesene schon selbstständig auf eine neue, konkrete Aufgabenstellung anwenden können. Erfahrungsgemäß ist man recht sicher, dass man die Aufgabe genau so gelöst hätte, wie es im Lösungsteil angegeben ist, wenn man unmittelbar nach dem Lesen der Aufgabe die Lösung nachliest. Wenn man die Aufgabe aber tatsächlich zu lösen versucht, ergeben sich doch Fragen. Wir raten Ihnen deshalb, falls die Aufgabe für Ihre geplante Arbeit relevant sein könnte, sie wirklich erst ohne die Lösungshinweise in Kapitel 11 zu bearbeiten.

Wir wollten einen möglichst leicht lesbaren Text schreiben, sodass wir uns entschieden haben, in Bezug auf die politische Korrektheit bei den Personenbezeichnungen wie folgt vorzugehen: Wenn eine tatsächlich geschlechtsneutrale Form (wie z.B. "Lehrperson") nicht existiert, sind die Forscher und Forscherinnen im Text immer weiblich, die Versuchsteilnehmer und Versuchsteilnehmerinnen (das ist der politisch korrekte Ausdruck für "Versuchspersonen") oder Schüler und Schülerinnen immer männlich bezeichnet. In beiden Fällen sind selbstverständlich beide Geschlechter gemeint.

Wir danken Tamara Faschingbauer, Sabine Jentges, Gebhard Aders und den Studierenden unserer Seminare für wertvolle Hinweise, die zur Verbesserung des Manuskripts beigetragen haben.

0 Zur Einführung

Manche Sprachlehrpersonen fragen sich Dinge wie beispielsweise:

Fragestellungen

- Wie lehrt man am besten Vokabeln? Sollte man Vokabeln den Schülern mit Hilfe von Texten beibringen, indem man eine Übersetzung der unbekannten Wörter eines Textes gibt, oder sollte man die Schüler die Bedeutung der Wörter aus dem Text heraus selbst erraten lassen?
- Wie sollte man Grammatik lehren? Ist es besser, sie zuerst zu erklären und dann einige Übungen zu machen oder sollte man mit einer Situation oder einem Text beginnen und die darin enthaltenen wichtigen grammatischen Erscheinungen mit den Schülern gemeinsam erarbeiten?
- Würden die weniger begabten Schüler mehr lernen, wenn ich die Gruppe aufspalte in eine mit den besseren und eine mit den schlechteren Schülern?
- Macht es einen Unterschied, ob in Gruppen mit Schülern unterschiedlicher Sprachlerneignung oder ob in homogenen Gruppen unterrichtet wird?
- Sind Schüler mit zwei arbeitenden Elternteilen tatsächlich im Schnitt besser als diejenigen, bei denen nicht beide Eltern arbeiten, oder ist das nur ein Eindruck, den ich aufgrund meiner eigenen Schülergruppe habe?
- Sind Mädchen wirklich sprachbegabter als Jungen? Und falls ja, warum? Oder fällt ihnen nur die Aussprache von Fremdsprachen leichter?

Oder sie interessieren sich für ganz andere Fragestellungen, wie beispielsweise:

- Wie unterscheidet sich die gesprochene von der geschriebenen Sprache in Bezug auf den Gebrauch komplexer Satzkonstruktionen?
- Worin unterscheidet sich die Sprache der Boulevardpresse von der seriöser Zeitungen?
- Ist es tatsächlich so, dass in politischen Talkshows in England mehr “tag questions” (wie *isn't it?*) verwendet werden als in anderen Sendungen? Und wenn das stimmt, woran liegt es?

Einige dieser Fragen können sehr leicht beantwortet werden, weil auf diesem Gebiet schon viel Forschung betrieben wurde. Trotzdem kann man der Meinung sein, dass man selbst mehr darüber herausfinden möchte, oder vielleicht glaubt man auch, dass die vorhandene Forschung zu einer bestimmten Frage für die eigene Situation nicht besonders relevant ist, oder man hat einfach Spaß an der Überprüfung von Hypothesen. Für diese neugierige Sorte der besonders enthusiastischen Sprachlehrpersonen – und angehenden Sprachlehrpersonen – sind dieses und die folgenden Kapitel geschrieben.

Auch Linguistinnen, die keine praktische Anwendung ihrer Forschung im Klassenzimmer planen, arbeiten oft empirisch. Empirisches Arbeiten liegt schon dann vor, wenn man seine Urteile des “kompetenten Sprechers” der jeweiligen Sprache nicht mehr bei sich selbst durch Introspektion gewinnt, sondern versucht, eine Anzahl “kompetenter Sprecher” systematisch nach ihren Grammatikalitätsurteilen zu befragen. Selbst eine so einfache Befragung kann man geschickt und auch ausgesprochen ungeschickt angehen.

Obwohl dies häufig nicht geglaubt wird, kann Forschung von Sprachlehrpersonen auch in tatsächlichen Klassenzimmersituationen betrieben werden. Alles, was man braucht um Forschung zu betreiben, ist Neugier, eine Menge Geduld, etwas Wissen über Statistik und eine gehörige Portion gesunden Menschenverstand. Zusätzlich muss man über die möglichen Fallen Bescheid wissen, die einen bei einem Forschungsprojekt erwarten.

Arten von Forschung

Arten von Forschung

Forschung wird gewöhnlich in qualitative und quantitative Forschung aufgeteilt. Quantitative Forschung wird dann wieder aufgeteilt in experimentelle und nicht-experimentelle Forschung. Nur als Anmerkung: “Qualitative Forschung” ist natürlich nur eine Bezeichnung und bedeutet nicht notwendigerweise, dass diese Forschung von besonders hoher Qualität ist, es gibt in der qualitativen und in der quantitativen Forschung gute und schlechte Arbeiten.

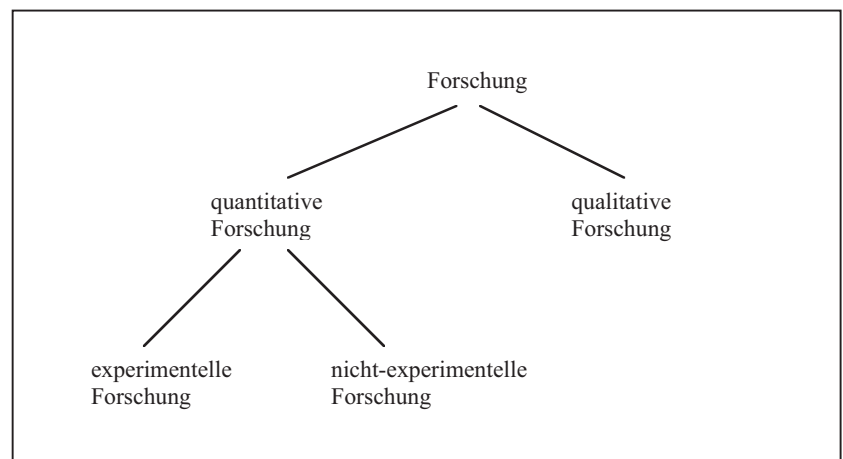


Abb. 1: Arten von Forschung

Quantitative Forschung beschäftigt sich mit Zahlen, die zusammengerechnet oder in komplizierte Formeln eingesetzt werden; bei dieser Art von Forschung beschäftigt man sich mit Konzepten wie dem Mittelwert und der Standardabweichung. Bei der qualitativen Forschung zählt man vielleicht auch (wie beispielsweise die Anzahl der weiblichen Charaktere in “Hamlet” im Vergleich zum “Sommernachtstraum”), aber man interessiert sich hauptsächlich für Meinungen, Gefühle und Intuitionen von anderen Menschen (oder sich selbst). Qualitative Forschung ist sehr häufig introspektiv. Wir lassen qualitative Forschung völlig außer Acht – einfach deshalb, weil qualitative Forschung sich nicht mit numerischen Daten beschäftigt.

Nicht-experimentelle quantitative Forschung zählt oft einfach nur; Meinungsumfragen sind ein typisches Beispiel dafür (“Wie viele Menschen beantworten eine bestimmte Frage mit *ja*, wie viele mit *nein*?”). Experimentelle Forschung ist dadurch charakterisiert, dass etwas (beispielsweise die Menge Alkohol, die man zu sich nimmt, oder die Art und Weise, wie man lernt) manipuliert wird, d.h. vom Forscher kontrolliert wird. Tabelle 1 enthält typische Beispiele für die verschiedenen Arten von experimenteller Forschung.

Tabelle 1: Beispiele für die drei Arten von Forschung

Experimentelle Forschung	nicht-experimentelle Forschung	qualitative Forschung
Der Effekt von Alkohol auf die Fahrtüchtigkeit	Änderungen im Familieneinkommen während der letzten 10 Jahre	Die Rolle der Frau bei Schiller
Ergebnisse eines frühen Lernanfangs im Englischen	Die Entwicklung der Wortstellung bei Kindern	Die Endung “-mark” in Ortsnamen
Effekte von Hausaufgaben im Sprachunterricht	Häufigkeit der einzelnen Vergangenheitstempora im Deutschen	Arten der Pronominalisierung im Frühneuhochdeutschen
Der Effekt von regressiver Assimilation auf die Worterkennung	Sprachlicher Hintergrund in multi-ethnischen Grundschulen	Kritik am Begriff “Regressive Assimilation”

Wann arbeitet man eigentlich in der Linguistik und Sprachlehrforschung sinnvollerweise empirisch? Nicht für jedes linguistische Problem und auch nicht für jedes Problem der Beschreibung einer bestimmten Sprache ist empirische Forschung nötig. Nötig ist sie aber dann, wenn man nicht durch die Beobachtung seines eigenen Sprachvermögens zu einer Lösung des anstehenden Problems gelangen kann. Das ist z.B. dann der Fall, wenn vermutet werden muss, dass es Unterschiede gibt zwischen der sprachlichen

Gegenstand

Norm und dem tatsächlichen Sprachgebrauch; in diesem Fall muss man den tatsächlichen Sprachgebrauch empirisch untersuchen, um die Unterschiede zur Norm zu erfahren. Auch kann es sein, dass man sich nicht sicher ist, ob jeder Sprecher der zu beschreibenden Sprache denselben Sprachgebrauch hat wie man selbst. Ebenso kann es sein, dass das zu untersuchende Phänomen äußerst komplex ist. Das ist z.B. in der Wirklichkeit des Fremdsprachenunterrichts immer so. Aber auch in der "reinen Linguistik" gibt es sehr komplexe Phänomene zu untersuchen. Dazu gehören vor allen Dingen die Beschreibung des Ablaufs von Gesprächsformen, z.B. Kommunikation vor Gericht, Kommunikation in der Schule, Verkaufsgespräche, Verhandlungsgespräche u.a.m. Für diese Untersuchungen gibt es eine spezielle linguistische Methode, die Gesprächsanalyse oder Konversationsanalyse. Sie beruht auf speziellen Verfahren der Datenerhebung und der Datenaufbereitung, darunter besonders die Transkription der Gespräche nach festgelegten Notationen, die nicht nur den Wortlaut der Äußerungen, sondern auch Betonungen, Gleichzeitig-Sprechen, Pausen, Tonhöhenverlauf und vieles andere mit berücksichtigen. In die Gesprächsanalyse oder Konversationsanalyse gibt es zwei Einführungen, einerseits Helmut Henne/Helmut Rehbock, 1982, *Einführung in die Gesprächsanalyse*, erschienen bei Walter de Gruyter – Sammlung Götschen und andererseits Klaus Brinker/Sven Sager, 1989, *Linguistische Gesprächsanalyse: Eine Einführung*, erschienen bei Schmidt, Berlin. Da es für diese Art empirischer Forschung bereits zwei gute Einführungen gibt, soll dieses Thema hier nicht weiter aufgegriffen werden. Zudem gelten für die Gesprächsanalyse einige Sonderbedingungen. Aufgrund des enormen Aufwands bei der Datenerhebung und -aufbereitung ist es bei konversationsanalytischen Arbeiten schwierig, mit einer angemessenen Anzahl von untersuchten Personen bzw. untersuchten Gesprächen zu arbeiten, so dass hier teilweise nicht auf den Standards bestanden wird, die wir im Folgenden für andere Arten von empirischer Forschung anführen. Wir beschäftigen uns also mit der nicht konversationsanalytisch ausgerichteten empirischen Forschung im Bereich Linguistik/Sprachlehrforschung.

Empirische Forschung im Bereich Linguistik außerhalb der Konversationsanalyse folgt den üblichen Regeln der empirischen Sozialforschung, d.h., die dort geltenden Gütekriterien und die dort geltenden Methoden werden von Linguisten genauso benutzt wie von Soziologen auch. Das Ziel dieser einführenden Darstellung ist, einen Überblick über diese Methoden zu verschaffen und die Leserinnen in die Lage zu versetzen, selbst eine empirische Untersuchung in diesem Bereich zu planen und durchzuführen. Dazu werden die folgenden Themen behandelt: Zunächst einmal die *Gütekriterien* für empirische Sozialforschung allgemein. Danach die Form der *Befragung*, wobei auf Interviews und Fragebogen-Befragungen eingegangen wird. Der Aufbau eines *Fragebogens* ist bei einer empirischen Untersuchung dieser Art der Teil, der ganz besonders sorgfältig geplant sein

muss. Wir erleben es immer wieder, dass Studierende mit einem gut gemeinten, aber falsch angelegten Fragebogen enorme Arbeit in eine wissenschaftlich nicht verwertbare Untersuchung stecken. Es folgen die *Beobachtung* und das *Experiment* als Forschungsmethode. Im anschließenden Teil geht es um die *statistische Auswertung* der Ergebnisse, die man bei seinen durch Befragung, Beobachtung oder Experiment erhobenen Daten gewonnen hat. Dieser Teil ist oft für Geisteswissenschaftler etwas angsterregend. Schließlich hat man das Studium einer Philologie gewählt, weil man schon immer nicht rechnen konnte.¹ Es gibt aber keinen Anlass vor Statistik Angst zu haben. Die eigentliche Rechenarbeit übernehmen heutzutage Computer-Programme, und der in dieser Einführung behandelte Teil der Statistik ist nur der, den man als Geisteswissenschaftler tatsächlich braucht.

Bevor man eine empirische Untersuchung beginnt, sollte man sich über die folgenden Fragen völlig klar sein:

Vorfragen

1. *Was genau soll untersucht werden?* Damit ist gemeint, der Untersuchungsgegenstand muss genau beschrieben werden. Man muss *exakt* festlegen, welchen Ausschnitt aus der Wirklichkeit man untersuchen möchte. Gerade in der Sprachlehrforschung, bei der es um recht verschiedene Dinge gehen kann, die alle auf den Unterrichtsprozess ihren Einfluss haben, muss man seine Untersuchungsfragestellung sehr klar formuliert haben, was keineswegs einfach ist. Ein Beispiel dazu: Nehmen wir an, es soll untersucht werden, ob man das Hörverständnis von Fremdsprachenlernern besser mit der Methode X oder mit der Methode Y fördern kann. Dann wird eine ganz entscheidende Frage für den Wert der Untersuchung sein, ob es gelingt, den Faktor "Hörverständnis" hinreichend von anderen Faktoren zu isolieren.
2. *Wie soll das zu Untersuchende erfasst werden?* Damit ist die Wahl der Methode gemeint. Abhängig von der Art der Fragestellung muss die richtige Methode zur Datenerhebung gewählt werden. Grundsätzlich können Daten über Beobachtung, Befragung oder durch ein Experiment gewonnen werden. Untersucht man z.B. Fehlerkorrekturen im Fremdsprachenunterricht, so kann man die Frage, welche Arten von Fehlerkorrekturen überhaupt vorkommen, kaum anders als durch Beobachtung klären, keine Lehrperson wird ihr Verhalten so genau beschreiben können, dass man tatsächlich alle Arten von Fehlerkorrekturen erfahren würde, die diese Lehrperson anwendet. Die Zufriedenheit der Schüler mit bestimmten Arten der Korrektur kann man über eine Befragung erheben und eine Erfolgsmessung verschiedener Korrekturverfahren gelingt am exaktesten in einem Experiment, in dem man alle störenden Einflussfaktoren ausschließen kann.

¹ Oder nicht? :-)

3. *Zu welchem Zweck sollen diese Daten erfasst werden?* Die Frage bezieht sich darauf, dass man vor der Datenerhebung bereits bedenken muss, welche Schlussfolgerungen man aus den Resultaten hinterher tatsächlich ziehen kann. Viele laienhaft angelegte empirische Untersuchungen sammeln Daten, die hinterher nicht interpretierbar sind. Soll z.B. eine bestehende Hypothese oder Theorie überprüft werden, so wird man sich fragen, welche Vorhersagen diese Hypothese für bestimmte Situationen macht. Dann kann man diese Situationen versuchen zu beobachten oder sie künstlich zu schaffen, um zu überprüfen, ob sich die Wirklichkeit entsprechend den Vorhersagen dieser Theorie verhält. Man beobachtet z.B. bei Befragungen oft, dass Studierende Fragen stellen, ohne sich vorher genügend genau überlegt zu haben, was die möglichen Antworten für ihre Untersuchung überhaupt bedeuten können. Den daraus resultierenden Problemen kann man entgehen, indem man, nachdem der Fragebogen konzipiert ist, systematisch untersucht, ob man die zu erwartenden Antworten überhaupt für seine Argumentation gebrauchen kann. Es ist nicht sinnvoll eine Befragung so anzulegen, dass man Fragen stellt mit dem Ziel "Ich sehe mal, was dabei herauskommt". Wenn man z.B. für die im Folgenden als Beispiel gebrauchte Untersuchung eine Operationalisierung von "katholisch" nach der Lohnsteuerkarte benutzt ("katholisch ist derjenige, in dessen Lohnsteuerkarte als Religion "katholisch" eingetragen ist"), dann ist eine Frage wie "Gehen Sie regelmäßig in die Kirche?" oder "Was halten Sie vom Papst?" nicht relevant für die Untersuchung. Derartige Fragen hätten nur einen Sinn, wenn anzunehmen wäre, es spiele eine Rolle beim Perfektgebrauch, ob jemand das "Katholisch-Sein" mehr oder weniger intensiv praktiziert.

Hypothesenprüfung

Was heißt "Hypothese"?

Empirische Forschung untersucht normalerweise, ob eine bestimmte Hypothese der Überprüfung in der Realsituation standhält. In diesem Zusammenhang muss zunächst einmal ausgeführt werden, was eine Hypothese für die empirische Sprachforschung überhaupt bedeutet. Eine Hypothese ist ein Satz, der empirisch falsifizierbar ist. Sinn der empirischen Untersuchung ist, dass durch sie eine Hypothese bestätigt oder verworfen werden kann. Die wichtigsten Kriterien für eine solche falsifizierbare Hypothese sind:²

1. Eine Hypothese ist eine Aussage, nicht eine Frage. Beispielsweise könnte man die Hypothese aufstellen: "Katholische Deutschsprachige gebrauchen Perfekt bei der Bezeichnung von Vergangenem, andere Deutschsprachige nicht". Diese Hypothese könnte man vereinfachen

² Die folgende Darstellung lehnt sich an die Ausführungen in Atteslander 1995 an.

zu: “Wenn jemand katholisch ist, dann gebraucht er das Perfekt bei der Bezeichnung von Vergangenem, andernfalls nicht”.

2. Die Aussage enthält mindestens zwei semantisch gehaltvolle Begriffe. Im Beispiel sind das “katholisch” und “Perfekt”.
3. Die Begriffe sind durch den logischen Operator “wenn-dann” verbunden. Also in unserem Fall, **wenn** jemand katholisch ist, **dann** gebraucht er das Perfekt.
4. Die Aussage ist nicht tautologisch, d.h. ein Begriff deckt den anderen semantisch nicht vollständig ab. Dies ist ein Problem der Definition der beiden Begriffe, auf die die Untersuchung sich bezieht. In unserem Fall würden wir “katholisch” wie oben erwähnt als eingetragene Zugehörigkeit zur katholischen Kirche (auf der Lohnsteuerkarte) definieren. Bei “Perfekt” würden wir uns vermutlich an die Morphologie halten und sowohl für starke als auch für schwache Verben genau festlegen, woran wir Perfektformen erkennen. In diesem Fall war es nicht schwer Tautologien auszuschließen. Das kann bei anderen sprachwissenschaftlichen Fragestellungen ganz anders sein. Angenommen, wir hätten die Hypothese zu überprüfen, “Jeder deutsche Satz enthält ein Subjekt”, dann würden wir vor der Frage stehen, wie “Subjekt” überhaupt ohne den Rückgriff auf den Begriff “Satz” zu definieren sein kann und umgekehrt.
5. Die empirischen Geltungsbedingungen sind implizit oder explizit im Einzelnen aufgezählt. In unserer Hypothese gilt implizit eine Universal-Aussage, d.h. ein einziger Katholischer, der das Präteritum statt des Perfekts gebrauchen würde, würde unsere Hypothese schon widerlegen. Betrachten wir nun den realistischeren Fall, wir hätten die Hypothese aufgestellt “Katholische Deutschsprachige gebrauchen häufiger Perfekt als Nicht-Katholische”. Dann würde stattdessen eine statistische Behauptung gelten, nämlich die, dass wenn man eine hinreichend große Gruppe von katholischen und von nicht-katholischen Sprechern des Deutschen untersucht, dass man dann bei den katholischen Sprechern prozentual mehr Perfekt, bezogen auf die Gesamtverteilung der Tempora, findet als bei den nicht-katholischen. Natürlich ist hier die Beweisführung viel komplizierter und erfordert einen viel größeren Aufwand beim Einsatz empirischer Methoden.
6. Die Begriffe sind auf Wirklichkeits-Phänomene hin operationalisierbar. Das haben wir bereits getan, indem wir “katholisch” operationalisiert haben mit “Zugehörigkeit zur katholischen Kirche (Eintrag auf der Lohnsteuerkarte)”. Auch “Perfekt” müssten wir nun operationalisieren durch eine feste Definition von “Perfekt”, die immer wieder gleich angewandt wird, also zu einer Liste der infrage kommenden Formen führt. Wir können nicht willkürlich bei den Nicht-Katholischen einen anderen Begriff von “Perfekt” zu Grunde legen als bei den Katholischen. Die Definition ist übrigens nicht so einfach, denn man

muss valenzreduzierte Formen wie *Die Straße ist gesperrt* und *Das Pferd hat die Fesseln bandagiert* ausnehmen.

7. Die Aussage ist falsifizierbar. Die ursprüngliche Hypothese wäre bereits mit dem Nachweis falsifiziert, dass eine einzige nicht-katholische Person einmal ein Perfekt benutzt hat. Bei der schwächeren Hypothese "Katholiken gebrauchen häufiger Perfekt" wäre die Hypothese falsifiziert, wenn sich bei einer hinreichend großen Anzahl von katholischen und nicht-katholischen Sprechern kein im statistischen Sinne signifikanter Unterschied (d.h. kein Unterschied, der groß genug ist, um den Zufall mit ausreichender Sicherheit als Ursache auszuschließen) in der Häufigkeit des Gebrauchs des Perfekts aufzeigen lässt. Das heißt noch nicht, dass das Gegenteil bewiesen wäre (also dass Katholiken das Perfekt genauso gebrauchen wie andere Sprecher des Deutschen), es heißt nur, dass unsere Daten keinerlei Unterstützung für die Annahme liefern, Katholiken gebrauchten mehr Perfekt als Nicht-Katholiken.

Die einzelnen Methoden

Gegenstand unserer Untersuchungen ist im allerweitesten Sinn menschliches Verhalten, nämlich Sprachverhalten, Kommunikationsverhalten oder Lernverhalten. Um über dieses Verhalten mehr zu erfahren, gibt es drei verschiedene Herangehensweisen. Zunächst einmal kann ich das zu untersuchende Verhalten da untersuchen, wo es natürlicherweise stattfindet. Das ist in der Sprachlehrforschung das verbreitetste der angewandten Verfahren. Man beobachtet das Unterrichtsgeschehen einfach im "normalen" Unterricht. Wichtig bei diesem Verfahren der *Beobachtung* ist es, dass die beobachtenden Forscher die zu beobachtende Wirklichkeit so wenig wie möglich verändern. Sie können sich vorstellen, dass eine Gruppe von Schülern – und ganz bestimmt auch eine Lehrperson – sich anders verhält, wenn eine Gruppe von Video-filmenden Forscherinnen in ihrem Unterricht auftritt, als in der normalen Unterrichtssituation.

Die zweite Möglichkeit ist die *Befragung*. Hier stört die Forscherin nicht den Prozess, über den sie etwas aussagen will, sondern sie bittet die am Prozess Beteiligten um Auskünfte. In diesem Fall wird die Wirklichkeit gefiltert durch bewusste Prozesse derjenigen, die befragt werden. Das ist besonders dann problematisch und erfordert besonders ausgefeilte Techniken, wenn Normen im Spiel sind. Jeder Befragte wird natürlich das Bestreben haben, sich so darzustellen als verhalte er sich den Normen entsprechend, vor allem, wenn er selbst diese Normen akzeptiert (was bei sprachlichen Normen häufiger der Fall ist als bei anderen Verhaltensnormen). In einer von Ruth Albert durchgeführten Untersuchung (Brons-Albert 1977) wurden Sprecher des Deutschen gefragt, welches Tempus sie zur Bezeichnung von zukünftigen Ereignissen verwenden. Fast alle Befragten erklärten, sie gebrauchten ständig das Futur. In später erhobenen Daten ihrer aktuellen Sprachproduktion gebrauchten sie allerdings nur in etwa 5% der Fälle das Futur. Wir können also nicht immer davon ausgehen,

dass die Antworten, die wir in Befragungen bekommen, tatsächlich die Wirklichkeit widerspiegeln.

Die dritte Möglichkeit der Verhaltensbeobachtung ist das *Experiment*. Der Vorteil von Experimenten ist, dass wir dort versuchen alles auszuschließen oder konstant zu halten, was für die Untersuchungsfragestellung keine Rolle spielt. Ein gut geplantes Experiment ist ausschließlich auf die Untersuchungsfragestellung hin geplant und erhebt keine anderen Daten als die, um die es bei der entsprechenden Hypothesenüberprüfung geht. Ein Beispiel dafür wären die bekannten Experimente zu den von Sprechern benutzten Regeln der Pluralbildung. In diesen Experimenten bildeten die Sprecher Pluralformen von verschiedenen aufgebauten Kunstwörtern, z.B. zu einem Wort wie *Wug* oder *Emmel*. Das erreicht man mit einem ganz einfachen Verfahren. Man zeigt eine Zeichnung oder Stoffpuppe eines seltsamen Tieres und erklärt: "Das ist ein Wug." Dann zeigt man eine Zeichnung von zwei Exemplaren davon oder zwei Stoffpuppen und erklärt: "Und hier ist noch ein Wug. Das sind jetzt zwei..." Die Versuchspersonen ergänzen ihren Plural von "Wug". Hier ist völlig ausgeschlossen, dass die Versuchspersonen die entsprechenden Wörter schon einmal in einem Plural gehört haben, sie müssen also die Pluralregeln anwenden, die sie mental gespeichert haben. Störende Einflüsse kann es aber trotzdem geben, z.B. durch die Reihenfolge, in der die Kunstwörter präsentiert (die Versuchsteilnehmer bleiben gern bei einer vorher gewählten Endung). Man muss also nach Techniken suchen, derartige Effekte möglichst auszuschalten, hier wäre eine ganz einfache Technik, dass man die Kunstwörter jedem Versuchsteilnehmer in einer anderen Reihenfolge zeigt.

Aufgabe

Nehmen wir an, Sie sollten Forschung auf einer Reihe von Gebieten, die unten in etwa beschrieben werden, betreiben. Welche Art der Forschung wäre am geeignetsten (experimentell, nicht-experimentell, qualitativ)? Beschreiben Sie, falls die experimentelle Forschung vorzuziehen ist, den Entwurf des Forschungsvorhabens, die verschiedenen Variablen und die Art und Weise, wie diese operationalisiert werden sollen. In anderen Worten: Wie würden Sie das Forschungsvorhaben angehen?

- a. Der Einfluss eines Partybesuchs in der Nacht vor einer Prüfung auf diese Prüfung
- b. Die Beziehung zwischen Sprachlaborübungen und Aussprache
- c. Die Bedeutung von Farben in den Werken von Chaucer (ein englischer Dichter des 14. Jahrhunderts)
- d. Die Fremdsprachenkenntnisse niederländischer Geschäftsleute
- e. Der Nutzen von Korrekturen der Grammatikfehler von Fremdsprachenlernern
- f. Verwendung von Höflichkeitsformen bei internationalen Verhandlungen
- g. Arten der Selbstkorrektur beim L1- und L2-Spracherwerb
- h. Ein Vergleich des stillen und des lauten Lesens bei der Entwicklung der Lesefähigkeit.

1 Gütekriterien für empirische Untersuchungen

Bei jeder empirischen Untersuchung entstehen einige nahe liegende grundsätzliche Fragen, die wir anhand von drei Beispielen erläutern wollen.

Einführung

Angenommen, wir wollen die Übersetzungsfertigkeit von Studenten messen und wir haben dazu einen Text von 200 Wörtern ausgesucht, der in die Fremdsprache übersetzt werden sollte. Manche Studenten haben eine gute Übersetzung geschrieben, manche eine mittelmäßige oder schlechte. Wissen wir auf Grund dieser Übersetzungen dann, ob diese Studenten gut oder schlecht übersetzen können? Oder könnte es sein, dass wir ganz andere Resultate bekommen hätten, wenn wir einen anderen Text ausgesucht hätten, einen Text über ein anderes Thema oder einen viel längeren Text?

Und angenommen, wir hätten diesen Text von einer von drei Gruppen Erstsemester-Studenten der Anglistik übersetzen lassen, können wir dann anhand der Resultate dieser Übersetzungen etwas über das zu erwartende Übersetzungsfertigkeitsniveau der anderen beiden Gruppen aussagen? Oder über die Fähigkeiten von Erstsemestern allgemein im Übersetzen?

Oder nehmen wir an, wir lassen einen Dozenten Essays beurteilen, die Studenten in einem Kurs "Schriftlicher Ausdruck" geschrieben haben. Wie sicher können wir sein, dass die Kriterien, die der Dozent benutzt, gut und konsistent sind? Würde ein anderer Dozent dieselben Noten vergeben? Und können wir, wenn ein Student eine gute Note für diesen Essay bekommen hat, annehmen, dass er sich gut schriftlich ausdrücken kann?

Um ein ausführliches Beispiel zu geben: Wir wollen herausfinden, wie Wörter in unserem mentalen Lexikon zusammenhängen, ob Wörter, die in bestimmten Kontexten häufig zusammen vorkommen, auch im Gehirn so organisiert sind, dass sie einander aktivieren. Wir betrachten Wörter aus demselben Script wie *Arzt*, *Krankenschwester*, *Krankenhaus* oder wie *Schlüssel* und *abschließen*, bei Mehrsprachigen auch in ihrer zweiten Sprache (*key*, *lock*), oder wir betrachten Wörter aus demselben Wortfeld wie *hell* und *dunkel* oder *Vogel* und *Spatz*. Um das zu untersuchen, haben wir ein Experiment entwickelt, in dem Wortpaare auf einem Computerbildschirm präsentiert werden, wobei das zweite "Wort" manchmal ein existierendes Wort und manchmal eine Buchstabenfolge ohne Bedeutung ist. Unter den Wörtern gibt es dann solche, die eine Beziehung zum anderen Wort des Wortpaares haben, und solche, die keine von den o.a. Beziehungen aufweisen. Das erste Wort wird kurz auf dem Computerbildschirm gezeigt und direkt danach das zweite. Die Versuchsteilnehmer (Studenten) müssen auf einen roten Knopf drücken, wenn das zweite kein Wort ist, und auf einen grünen, wenn es ein Wort ist. Dabei interessiert uns nur, wie sie auf tatsächliche Wörter reagieren. Wir messen die Reaktionszeit von der Präsentation des zweiten Wortes bis zum Drücken des Knopfes. Wir erwarten, dass

die Reaktionszeit kürzer ist, wenn das präsentierte Wortpaar *key – lock* ist, als wenn das präsentierte Wortpaar *fee – lock* ist.

Dabei kommen Fragen auf wie:

- Wie viele Wortpaare brauchen wir eigentlich, um ein einigermaßen zuverlässiges Ergebnis zu bekommen? Reichen 5? Oder 10? Oder brauchen wir erheblich mehr?
- Und wenn wir Unterschiede in der Reaktionszeit finden, was können wir auf Grund davon genau über unsere Fragestellung aussagen?
- Und, angenommen wir haben tatsächlich gefunden, dass die Reaktionszeit kürzer war, wenn es eine – wie auch immer geartete – Beziehung zwischen den beiden Wörtern des Wortpaares gab, können wir dann für die Schlüsse, die wir daraus ziehen, mit der nötigen Sicherheit sagen, dass sie nicht nur für unsere kleine Gruppe von Versuchsteilnehmern gelten, sondern für alle Deutschsprachigen?

Die oben gestellten Fragen beziehen sich unter anderem auf die Reliabilität, die Validität und den Geltungsbereich der jeweiligen Untersuchung.

1.1 Verlässlichkeit (Reliabilität)

Verlässlichkeit

Die Begriffe “Verlässlichkeit”, “Zuverlässigkeit” und “Reliabilität” werden synonym gebraucht; sie bezeichnen dasselbe, und zwar ob das Messverfahren exakt ist, ob es das, was gemessen werden soll, exakt erfasst. Als verlässlich gilt eine Erhebung und das bei dieser Erhebung benutzte Instrument dann, wenn bei einer Wiederholung des Verfahrens unter gleichen Bedingungen die gleichen Ergebnisse erzielt werden. Das ist bei einfachen Messverfahren kein Problem: die Länge meines Tisches sollte dieselbe sein, wenn ich sie zweimal nacheinander mit demselben oder einem anderen Zollstock messe, aber bei unserer Art von Untersuchungen macht die Ermittlung der Reliabilität schon einigen Aufwand nötig.

Zur Ermittlung der Verlässlichkeit gibt es im Prinzip vier Verfahren:

- **Testwiederholung:** Wenn es möglich ist, kann man – nach einer gewissen Zeit, damit sich die Versuchsteilnehmer nicht mehr an das erinnern, was sie beim letzten Mal produziert haben – einfach denselben Test noch einmal machen. Dabei sieht man, ob das Ergebnis der zweiten Durchführung des Tests mit der ersten übereinstimmt. Dieses Verfahren kann man bei Grammatikalitätsurteilen durchaus anwenden, bei Sprachbeherrschungstests geht es meist nicht, weil die Lerner in der Zwischenzeit – oder durch den ersten Test selbst – Lernfortschritte gemacht haben können.
- **Paralleltest:** Man untersucht dieselben Versuchsteilnehmer ohne nennenswerten zeitlichen Abstand mit einer zweiten Version des Tests,