



Rainer Dietrich / Johannes Gerwien

Psycholinguistik

Eine Einführung

3. Auflage



LEHRBUCH
J.B. METZLER



J.B. METZLER

Rainer Dietrich / Johannes Gerwien

Psycholinguistik

Eine Einführung

3., aktualisierte und erweiterte Auflage

Mit Abbildungen und Grafiken

J. B. Metzler Verlag

Die Autoren

Rainer Dietrich ist Professor em. für germanistische Linguistik an der HU Berlin und der Universität Heidelberg.

Johannes Gerwien, Dr., ist Akademischer Rat an der Universität Heidelberg.



Gedruckt auf chlorfrei gebleichtem, säurefreiem und alterungsbeständigem Papier

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet
über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-476-02644-6

ISBN 978-3-476-05494-4 (eBook)

Dieses 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.

J.B. Metzler, Stuttgart

© Springer-Verlag GmbH Deutschland, 2017

Einbandgestaltung: Finken & Bumiller, Stuttgart (Foto: photocase.com / Dirk Hinz)

Satz: primustype Robert Hurler GmbH, Notzingen

Druck und Bindung: Ten Brink, Meppel, Niederlande

J.B. Metzler ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer-Verlag GmbH Deutschland

www.metzlerverlag.de

info@metzlerverlag.de

Inhaltsverzeichnis

Vorwort zur ersten Auflage	IX
Vorwort zur dritten Auflage.....	X
1 Psycholinguistische und angrenzende Fragen.....	1
1.1 Sprache und andere Kommunikationssysteme	1
1.2 Gegenstand und Ziele der Psycholinguistik	8
1.3 Wissenschaftliche Erkenntnis in der Psycholinguistik	12
1.4 Methoden	14
1.4.1 Beschreiben, Erklären, Vorhersagen.....	14
1.4.2 Messen, Messmethoden und Güte der Messung	16
1.4.3 Elizitationsmethoden	19
2 Sprachliches Wissen	23
2.1 Das mentale Lexikon	25
2.1.1 Die lexikalische Einheit	26
2.1.2 Die Struktur der lexikalischen Einheit.....	30
2.1.3 Beschaffenheit der lexikalischen Information.....	33
2.1.4 Komplexe Wörter	39
2.2 Strukturelles Wissen	42
2.2.1 Strukturelles Wissen auf der semantischen Ebene	46
2.3 Repräsentation sprachlichen Wissens.....	47
2.4 Die Lokalisationsfrage	48
3 Spracherwerb.....	57
3.1 Die Herausforderung	59
3.2 Die Voraussetzungen.....	60
3.3 Methoden – historisch	62
3.3.1 Longitudinalstudien und Methoden der Verhaltensbeobachtung	63
3.3.2 Experimentelle Methoden der Spracherwerbsforschung	68
3.4 Die Sprachentwicklung.....	71
3.4.1 Die lautliche Entwicklung.....	73
3.4.2 Erwerb des Wortschatzes	78
3.4.3 Der Erwerb von Morphologie und Syntax.....	83
3.5 Erklärungsansätze	87
3.5.1 Das Nativismus-Modell	89
3.5.2 Das Sprachgebrauchsmodell.....	93
3.5.3 Der kognitivistische Ansatz	94
3.6 Der bilinguale Erstspracherwerb	96
3.7 Zweitspracherwerb	101
3.8 Gebärdenspracherwerb	108
4 Sprechen.....	113
4.1 Die Kommunikationsplanung	116
4.2 Die Konzeptualisierung.....	120

4.3	Die Formulierung	125
4.3.1	Überblick der ablaufenden Prozesse	125
4.3.2	Evidenz für die Teilsysteme der Sprachproduktion	128
4.3.3	Die Dynamik des Zusammenspiels der einzelnen Prozessebenen	135
4.3.4	Die Erzeugung der Satzstruktur	148
4.3.5	Die Erzeugung der Lautkette	152
4.3.6	Das Zusammenspiel von syntaktischer und phonologischer Kodierung	156
4.4	Die Artikulation	157
4.5	Selbstkontrolle und Selbstkorrektur	158
5	Sprachverstehen	163
5.1	Die Schritte des auditiven Sprachverstehens im Überblick	164
5.2	Lexikalische Erkennung	165
5.2.1	Lauterkennung	165
5.2.2	Die Worterkennung	173
5.3	Die Analyse der Satzstruktur	183
5.3.1	Die Herausforderung an das Parsing-System	183
5.3.2	Das syntaktische ›Wissen‹ des Parsers	187
5.3.3	Vom grammatischen Prinzip zur Analysemaxime	192
5.3.4	Nicht-syntaktisches Wissen	195
5.3.5	Strategien der Satzanalyse	199
5.3.6	Modelle	200
5.4	Verstehen der Satzbedeutung	204
5.5	Lesen	208
5.5.1	Lesen als Kommunikationsform	208
5.5.2	Lesen als Prozess der Sprachverarbeitung	209
5.6	Sprachverstehen – ein separates Teilsystem?	219
6	Sprachstörungen	223
6.1	Die erworbenen Sprach- und Sprechstörungen	224
6.1.1	Die Aphasien	226
6.1.2	Die Dysarthrien	227
6.1.3	Die Sprechapraxie	227
6.2	Psycholinguistische Betrachtungen und einige Relativierungen	228
6.2.1	Psycholinguistische Betrachtungen	228
6.2.2	Relativierungen und Revisionen	229
6.3	Die Sprachentwicklungsstörungen	231
6.3.1	SES bei Autismus	232
6.3.2	SES bei Fragilem X-Syndrom (Fra-X)	233
6.3.3	SES bei Down-Syndrom	234
6.3.4	SES bei Williams-Beuren-Syndrom (WBS)	234
6.3.5	Spezifische Sprachentwicklungsstörung (sSES)	235
6.3.6	Nicht-klinische Sprachentwicklungsstörungen	236

7	Literaturverzeichnis	239
7.1	Lehrbücher	239
7.2	Einführende Darstellungen	239
7.3	Handbücher	239
7.4	Weitere im Text zitierte Titel	240
8	Sachregister	261

Vorwort zur ersten Auflage

Die *Psycholinguistik* soll zwei Zwecken dienen. Der Band präsentiert das Bild, das man sich heute von den kognitiven Zuständen und Vorgängen des sprachlichen Wissens, des Spracherwerbs, von Sprachproduktion und -verstehen sowie von Sprachstörungen macht. Zugleich ist der Stoff so ausgewählt, geordnet und dargestellt, dass er in einem zweistündigen Einführungskurs in einem Semester behandelt werden kann. Das Buch bietet die empirisch begründete Skizze der Sprachfähigkeit des Menschen und – damit verbunden – viele weiterführende Fragen, die aus Befunden folgen und die sicherlich durch weitere Messungen und Deutungen zu beantworten sind, wenn unsere Ideen unsere Laborarbeit in eine glückliche Richtung lenken.

Dass der Text schließlich zustande gekommen ist, habe ich vielen zu verdanken. Am meisten den Studierenden, die mit Neugier und Eifer an meinen Psycholinguistikkursen der letzten zehn Jahre an der Humboldt-Universität zu Berlin teilgenommen und – wie Thorsten Fellberg – mancherlei Unklarheiten in meinen Ausführungen angemerkt haben. Die Mitarbeiter, Kollegen und Freunde in und an der Berliner Psycholinguistik mit Guido Kiecker im Mittelpunkt haben nicht nur die dazu gehörenden Räume zu einem professionellen, lebendigen und produktiven Labor gemacht, sondern auch mehr als sie bemerkt haben, zu meinem Blick auf die Inhalte des Faches beigetragen. Dass es Ingrid Häfner geschafft hat, mehrere hundert Seiten handschriftlichen Manuskripts quasi nebenbei in Word-Dateien zu überführen, sollte Anlass geben, ihr die Solopartie in einer Neueinspielung von Lee Roy Andersons Typewriter anzubieten.

Inhaltlich besonders förderlich waren die ausgedehnten Arbeitsaufenthalte von Herbert Schriefers und Jörg Jescheniak am Berliner Labor und die gründlichen Kommentare von Jürgen Weissenborn zu einem Entwurf von Teilen des Buches. Als geradezu unentbehrlich hat sich erwiesen, dass ich bei Bedarf jederzeit am Max-Planck-Institut für Psycholinguistik arbeiten und mit Wolfgang Klein über alles reden konnte. Während eines solchen Aufenthalts dort hat Tilman von Meltzer Literaturverzeichnis und Referenzen bearbeitet und später auch das Register erstellt; er und Thomas Wilhelmi haben Korrektur gelesen und auch ihre Anmerkungen zum Inhalt waren mir sehr hilfreich. Die Stimme, die in Kapitel 5.1 zu sehen ist, ist die von Katja Kühn.

Das Buch wäre vielleicht gar nicht oder jedenfalls noch nicht fertig geworden, hätte ich nicht die ungestörte Zeit am Stück gehabt, die durch ein Forschungssemester der Deutschen Forschungsgemeinschaft gewährleistet ist.

Gewidmet ist das Buch Lu aka Lise. Sie hat mir zu der Erfahrung verholfen, dass es eine Steigerung von Sprechen und Hören gibt, das Aussprechen und das Zuhören.

Vorwort zur dritten Auflage

Nach fast zehn Jahren liegt nun die dritte Auflage der »Psycholinguistik« vor. Im Rückblick auf die zweite Auflage fällt zweierlei spontan ins Auge; das Buch ist nun das Werk von zwei Autoren und das Format sowie die graphische Gestaltung sind gänzlich neu.

Ersteres liegt in der Konsequenz der beträchtlichen Zunahme an neuen Befunden und der damit einhergehenden Diversifizierung der Forschungslage in allen Teilgebieten der Psycholinguistik. Aus ihr ergibt sich die Notwendigkeit zur Spezialisierung und infolgedessen wiederum zur Teamarbeit.

Letzteres, die graphische Gestaltung, ist der sichtbare Ausdruck einer erweiterten Zielsetzung. Vor zehn Jahren war die Darstellung des Faches noch von der Absicht geprägt, den großen Fragen des Gebiets folgend den Stand der Forschung aufzuzeigen und anhand von exemplarischen Befunden zu illustrieren und zu problematisieren. Seither hat sich zweierlei verändert. Der Anteil an kanonisierten Ergebnissen im Forschungsfeld von Spracherwerb, sprachlichem Wissen, Sprachverarbeitung und Sprachstörungen ist enorm gewachsen. Zum zweiten sind Befunde und Ziele der Psycholinguistik zunehmend in den Blick benachbarter Disziplinen gerückt. Dem wachsenden Interesse von Sprachwissenschaft, Kognitionswissenschaft und Neuropsychologie wird mit der 3. Auflage nun ein Aufriß des Faches angeboten, der sich noch besser als Lehrbuch für Zwecke einer Einführung in die Psycholinguistik in diesen angrenzenden Fächern eignen sollte.

Für diese dritte Auflage sind alle Kapitel inhaltlich aktualisiert und entsprechend die Hinweise auf die relevante Literatur ergänzt worden. Ein besonderes Augenmerk lag dabei auf den bahnbrechenden Befunden der neurolinguistischen Forschung.

Wie schon in der Vergangenheit haben auch zu dieser Überarbeitung viele mit Hinweisen, Fragen und Anregungen beigetragen. Besonderen Dank dafür sagen wir Despina Giesel, deren gründlicher Lektüre es zu danken ist, dass mehr als ein halbes Hundert von Tippfehlern bereinigt werden konnte. In besonderem Maße förderlich war uns die laufende Begleitung von Ute Hechtfisher, J.B. Metzler Verlag, bei unserem Bemühen, das »Taschenbuch« der 1. und 2. Auflage in ein respektables Lehrwerk zu überführen. Schließlich haben wir auch dieses Mal den Studierenden in unseren psycholinguistischen Grundkursen, Proseminaren, Seminaren und Kolloquien an der Universität Heidelberg für inspirierende, oft verblüffend grundsätzliche Fragen und Kommentare zu danken.

Rainer Dietrich und Johannes Gerwien
Heidelberg im September 2016

1 Psycholinguistische und angrenzende Fragen

- 1.1 Sprache und andere Kommunikationssysteme
- 1.2 Gegenstand und Ziele der Psycholinguistik
- 1.3 Wissenschaftliche Erkenntnis in der Psycholinguistik
- 1.4 Methoden

1.1 | Sprache und andere Kommunikationssysteme

Von den vielerlei lautlichen Äußerungen, zu denen die Ente in der Lage ist, werden einige von fast allen Arten, einige hingegen nur von wenigen beherrscht. Von den zwanzig Arten, die Lorenz (1965) vergleicht, verfügen offenbar alle über die Äußerung vom Typ EPV, das einsilbige Pfeifen des Verlassenseins. Bei nur zwei Arten, beide echte Gänse im Unterschied zu den Pfeifgänsen, beobachtet man den mehrsilbigen Kükenstimmführungs-laut der Anserinen. Die beiden Gansarten, die gerade diese Äußerungsart lernen können, unterscheiden sich auch biologisch von den übrigen Entenarten. Aber es gibt auch biologische Gemeinsamkeiten; alle sind eben Enten und alle beherrschen das einsilbige Pfeifen des Verlassenseins.

Das Verhalten biologischer Gattungen zu vergleichen, zielt darauf ab, ihre stammesgeschichtliche Verwandtschaft zu untermauern und ihre biologische Ausstattung funktional zu erklären. Ersteres ist heute auch anhand vergleichender Analysen des Erbgutes mit molekularbiologischen Methoden möglich, und die stammesgeschichtliche Stellung des Menschen relativ zu seinen Gattungsnachbarn und Vorfahren ist biologisch so weit bekannt, dass die entsprechenden Erkenntnisse zum Schulwissen gehören. Was der Biologieunterricht aber nicht lehrt, ist, wie ähnlich das menschliche Verhalten dem von Angehörigen benachbarter Gattungen ist. Zwar herrscht weitgehend Konsens darüber, dass Moralempfinden, Gewissen, Abschätzung der Folgen eigenen und fremden Handelns bei keiner anderen Spezies als beim Menschen nachgewiesen sind. Ob aber der Mensch das einzige Wesen ist, das begrifflich denkt und mittels Äußerungen einer gegliederten Wortsprache kommuniziert, wie es (Medicus 1985) postuliert, lässt sich gerade mit der Methode der vergleichenden Verhaltensanalyse in Frage stellen.

Aussagekräftige Kategorien für einen Vergleich könnte schon die Liste von dreizehn Merkmalen kommunikativen Verhaltens liefern, die Hockett (1960) verwendet, um die menschliche Sprache mit dem Tanz der Bienen und mit den Äußerungen von Gibbons zu vergleichen. Danach weisen die sprachliche Kommunikation des Menschen und die Rufe der Gibbons gemeinsame Merkmale auf; verschieden sind sie hingegen in der zweifachen Gegliedertheit des Kommunikationssystems, der lautlichen und der nicht-lautlichen, die der menschlichen Sprache eigen ist, nicht aber den Gibbon-Rufen. Gemeinsam ist beiden andererseits die Arbitrarität der Äußerungs-

mittel, die sie aber wiederum nicht mit dem Tanz der Bienen teilen, dem aber durchaus Bedeutungshaltigkeit zugesprochen wird, was die Bienen damit mit der Affen- und Menschenkommunikation teilen.

Was ist mit einer solchen Klassifizierung gewonnen? Die Fakten lassen sich in geordneter Weise beschreiben und das ist nicht wenig. Aber welche Fakten? Im sprachwissenschaftlichen Grundkurs lernen wir die Unterscheidung zwischen Sprache und Rede, d. h. dem System der Ausdrucksmittel (*langue*) und der kommunikativen Verwendung der jeweils geeigneten Mittel in einer Situation (*parole*). Ursprünglich zur Unterscheidung linguistischer Gegenstände gedacht, bietet diese begriffliche Unterscheidung auch einen Ansatzpunkt zu weiterer Klarheit in der ethologischen Forschung. Unterscheidet man nämlich zwischen Kommunikation und Sprachsystem generell, nicht nur beim Menschen, so stellt sich die **ethologische Ausgangslage** so dar:

Genetische Verankerung? Viele Arten haben die Fähigkeit, zweckmäßig und eindeutig zu kommunizieren und jede verwendet dazu ein je artspezifisches System. Die Kommunikationsverfahren sind bestimmt von biologischen, sozialen und kognitiven Voraussetzungen der jeweiligen Art und ihrer Lebenswelt. Und die artspezifischen Sprachen sind es auch – aber anders (vgl. Hauser 1996). Die Kommunikationsverfahren sind angepasst einerseits an die peripheren Organe der Produktion und Rezeption, andererseits an die psychischen Repräsentationen der jeweiligen Organismen. Es ist trivial, dass die menschliche Kommunikation primär über den lautlichen Kanal und im Frequenzbereich zwischen 100 und 20.000 Hertz betrieben wird, da wir nicht unter Wasser leben und unsere Sprech- und Hörorgane nicht für die Verarbeitung von Ultraschall geeignet sind. Nicht trivial dagegen ist die Frage im Blick auf die sprachlichen Systeme. Hauser/Chomsky/Fitch (2002) nehmen die Sicht eines Besuchers vom Mars ein und wundern sich:

»If a Martian graced our planet it would be stuck by one remarkable similarity among Earth's living creatures and a key difference. [...] It would note that all living things are designed on the basis of [...] systems that read a universal language encoded in DNA-base pairs. [...] In contrast, it would notice the absence of a universal code of communication« (S. 1569).

Damit destilliert sich ein **engerer Begriff von Sprachfähigkeit** heraus, als der, den Hockett zum Gegenstand seiner Klassifikation gewählt hat, nämlich der des ›*code of communication*‹, linguistisch gesprochen, das Konzept der ›Grammatik‹. Warum haben die Lebewesen bei aller Verschiedenheit der Kommunikationsverfahren nicht immerhin eine universale Grammatik entwickelt? Stellen wir die Warum-Frage einstweilen hintan bis die Was-Frage formuliert und geklärt ist. Inwiefern gibt es keine universale ›Grammatik‹ im Unterschied zur universalen Codierung der Genome? Es sind grundsätzlich zwei Befunde denkbar:

- Jede Artengrammatik (Bientanz, Gesang der Stare, die Lautsprache der Bonobo-Männchen etc.) ist strukturell vollkommen verschieden von jeder anderen. Sie teilen keine, aber auch gar keine strukturellen Eigenschaften.

- Die alternative Erkenntnis könnte sein, dass es ein endliches Repertoire grammatischer Relationen gibt, das aber nicht alle Artengrammatiken alle diese Relationen aufweisen. Bei Relationen denken wir an Nachbarschaftsrelationen (Laut 1 – Laut 2 – Laut 1 – Laut 3; Wort 1 – Wort 2 – Wort 3 – Wort 2 ...) und hierarchische Relationen wie in Abb. 1.1, etc.

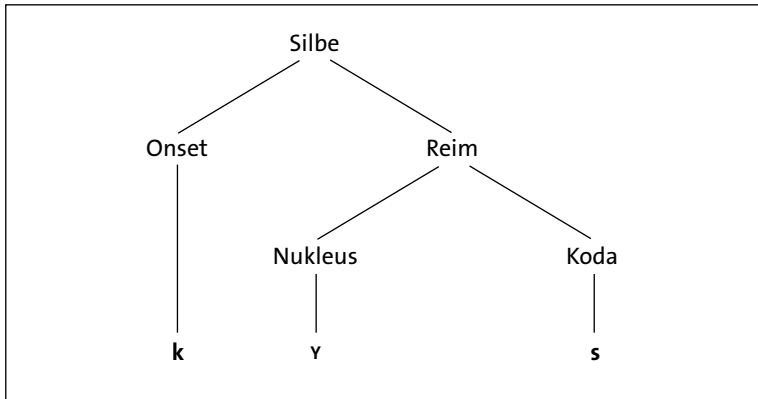


Abb. 1.1: Hierarchische Gliederung der Silbe /kvs/ im Wort ›Küster‹

Ein Spezialfall dieses Befundes wäre, dass die Grammatiken intrinsisch hierarchisch geordnet sind; jede höhere weist alle Strukturmerkmale aller hierarchisch niedrigeren auf und zusätzlich ein artspezifisches Merkmal. Dass die Artengrammatiken nicht komplett verschieden sind, ist unmittelbar einsichtig, denn viele, die wir aus dem täglichen Umgang mit Tieren kennen, weisen in ihren Ausdrücken die Nachbarschaftsrelationen auf: Stare singen ›Sätze‹, in denen sich ›Wörter‹ wiederholen (vgl. z. B. Gentner/Fenn/Margoliash/Nusbaum 2006, S. 1204) und Bonobo-Äußerungen sind akustisch der Lautform einer grammatischen Äußerung von Menschen nicht unähnlich.

Die Annahme einer vollständigen gegenseitigen Disjunktion (Verschiedenheit) aller Artengrammatiken ist also leicht zu widerlegen, ebenso diejenige einer universalen Grammatik aller Wirbeltiere. Was nun aber letztlich der Fall ist, ist eine empirische Frage. Es müssten die Struktureigenschaften aller Grammatiken ermittelt und diese dann auf ihre Teilmengeninklusionen hin untersucht werden, eine Herkulesarbeit. Machbar wäre, theoriegeleitet eine Annahme zu formulieren und diese dann empirisch zu testen. Eine solche Annahme ist, dass die Grammatik der Menschensprache viele Merkmale mit den Grammatiken anderer Arten teilt, sich aber in genau einem Strukturmerkmal von den Kommunikationssystemen aller Tiere unterscheidet. Mit jedem Falsifikationsversuch, dem sie standhält, gewinnt sie an Wahrscheinlichkeit. Hauser et al. (2002) gehen diesen Weg.

Rekursivität: In der natürlichen Sprache des Menschen gibt es die Relation der Rekursivität, d. h., sie weist Regeln der Form $A \rightarrow A-B$ auf. Die Konstituente A kann eine Kette dominieren, die wiederum ein A als Glied enthält. Wenn diese Annahme zutrifft und zugleich keine Tier-Grammatik gefunden wird, die ein Strukturmerkmal aufweist, das in der Menschengrammatik nicht auch belegt ist, so würde das heißen, dass die menschl-

Artspezifik

che Sprache einzigartig ist – unter den Lebewesen der Erde. Dies ist die einstweilige Folgerung von Chomsky/Lasnick (1993) aus einer Reihe von evolutionstheoretischen Fakten und Überlegungen. Sie lassen wenig Zweifel daran, dass diese wie jene nichts als Spekulationen sind – aber entscheidbare – in the long run. Das ist eine starke Behauptung und sie ist erwartungsgemäß nicht unumstritten – in beide Richtungen. Gentner et al. (2006) behaupten, dass auch die Grammatik des Starengesanges Rekursivität aufweist und Jackendoff/Pinker (2005) problematisieren Grundbegriffe von Hauser et al. (2002), darunter den der »Grammatik« als ein System von Prinzipien, angewandt auf elementare lexikalische Einheiten.

Einzigkeit der menschlichen Kommunikation: Anthropologisch wird die Einzigartigkeit der menschlichen Kommunikation nicht erst an der Fähigkeit zur Sprache verankert, sondern an der artspezifischen Fähigkeit, absichtsgeleitet zu handeln und die Absicht von Interaktionspartnern zu erkennen. Zwar führen z. B. Affen wie auch Kleinkinder Zeigegesten aus, um die Aufmerksamkeit eines anderen auf Objekte in der Umgebung zu lenken; die Intention des Menschen dabei, so Tomasello/Carpenter/Call/Behne/Moll (2005) und Tomasello/Carpenter (2005) ist aber eine fundamental andere. Sie ist eingebunden in ein gegenseitiges Verstehen von Intentionen, eine Fähigkeit über die nach dieser Theorie ausschließlich der Mensch verfügt. Wir gehen hier nicht weiter ins Detail. Noam Chomsky soll die Stichhaltigkeit der Feststellung, menschliche Sprache sei Kommunikationssystemen der Tiere ähnlich, mit derjenigen gleichgesetzt haben, das Verhalten des Menschen, der in die Luft springt und die Arme auf und ab schwingt, sei dem Flug des Vogels ähnlich; also könne der Mensch im Prinzip fliegen wie der Vogel. Manch einem wird es fraglich sein, ob derlei Behauptungen überhaupt wissenschaftlich diskutiert und entschieden werden können. Die Fragen jedenfalls tauchen immer wieder auf, typischerweise auch in Einführungsseminaren. Ob man darauf entschieden oder dilatorisch antwortet – eine wissenschaftliche Antwort setzt in jedem Fall voraus, dass man über die Sprachfähigkeit des Menschen Bescheid weiß. Und bei der Annäherung an dieses Wissen verhilft gerade die ethologische Perspektive zu der nützlichen Distanziertheit, die die Wahrnehmung und die Aufnahme der Befunde prägen sollte. Der Mensch ist kein Delphin, keine Biene und kein Schimpanse. Dennoch ergibt sich beim Durcharbeiten dieses Textes mehr Plastizität, wenn man sein Kommunikationsverhalten mit der Kommunikation im Tierreich hie und da vergleicht. Natürlich fällt der Kontrast krass ins Auge, wenn dem Pfeifen der Graugans ein Stück aus einem routinierten und inhaltlich zudem fesselnden Gespräch gegenübergestellt wird.

Beispiel Cockpit-Voice-Recorder-Mitschnitt

Ein packendes Beispiel eines solchen Gesprächs ist etwa der Dialog zwischen zwei Flugzeugführern in einer kritisch werdenden Lage beim Landeanflug auf einen Flughafen. Quelle dieser authentischen Aufzeichnung ist eine Transkription des Cockpit-Voice-Recorder-Mitschnitts (CVR) durch die Nationale Transport-Sicherheitsbehörde der U. S. A. Zur Situation: Im September 1993 befindet sich ein Airbus 320–211 von

Frankfurt am Main kommend auf dem Anflug auf den Flughafen von Warschau. Der Flug hat die Nummer Lufthansa 2904. An Bord sind Pilot und Copilot, vier weitere Mitglieder der Besatzung, darunter der Purser, der u. a. die Kommunikation zwischen Kabine und Cockpit wahrnimmt, sowie vierundsechzig Passagiere. Der Anflug findet unter ungünstigen Windbedingungen statt, vom Boden werden der Maschine Böen von starkem Seitenwind über der Landebahn gemeldet. Dadurch setzt die Maschine seitlich gekippt auf der Bahn auf; es berühren nicht alle Räder des Fahrwerks gleichzeitig den Boden, was zu Bremsproblemen und schließlich zu einer gefährlichen Annäherung an die hintere Landebahnbegrenzung führt.

Hier sind die letzten vier Minuten der Kommunikation aus der Sicht des Cockpit:

Sprecher bzw. Hörer sind:

PF: Der Pilot, der fliegt.
 PNF: Der Pilot, der nicht fliegt (Copilot)
 PUR: Purser
 TWR: Tower (Fluglotse im Tower)
 APP: Approach (Fluglotse für den Luftraum über dem Flughafen)
 ACO: Stimme der automatischen Höhenansage im Cockpit.

<i>Zeit</i>	<i>Sprecher</i>	<i>Äußerung</i>
15.29:38	PNF	Noch hundert. < Klack > Bist leicht überm glideslope.
15.29:42	PF	Stimmt.
15.29:46	PNF	Ungefähr eine Daumenbreite.
15.29:52	PUR	Kabine ist klar.
15.30:06	PNF	Gear down. <= Fahrwerk raus >
15.30:07	PF	Gear down. < Bestätigung > < Geräusch von Hebel und Fahrwerk > < Gong; Höhenwarnung >
15.30:16	PF	Jetzt hat er ihn. < den glideslope >
15.30:17	PNF	Genau.
15.30:23	PF	Dreitausendachthundert. < Fuß. Die Höhe unterhalb der in der gegebenen Situation kein Durchstarten mehr möglich ist. > [...]
15.30:45	PNF	Lufthansa two-nine-zero-four is established.
15.30:48	APP	Lufthansa twoninezerofour continue ILS <= Instrumentenlandesystem > approach runway oneone as number one, call tower one-two-one point six, thank you, see you later.
15.30:56	PNF	Roger, one-six, bye bye.
15.31:00	PNF	Tower, good afternoon, Lufthansa two-nine-zero-four ILS runway one-one.
15.31:07	TWR	Lufthansa two-nine-zero-four, continued ILS approach, call me Outer Marker, wind one-six-zero degrees two-five kilometers per hour and before landing, eh it was report windshear on the final runway one-one
15.31:18	PF	Windshear <= Seitenwind >

	Zeit	Sprecher	Äußerung
Beispiel für Funktionen der Sprach- verwendung	15.31:21	PNF	Roger, that's understood, I call you Outer Marker.
	15.31:25	PF	< Click > Flaps three. (Landeklappen in Stellung drei)
	15.31:26	PNF	Flaps three. (Bestätigung) [...]
	15.32:14	PF	Dreht ja noch ordentlich.
	15.32:16	PNF	Muss ja drehen.
	15.32:17	PF	Muss ja, das kann ja gar nicht anders sein.
	15.32:23	PNF	Gut [...]
	15.32:57	PNF	Field in sight. (Landebahn in Sicht)
	15.32:58	PF	Visual. (Umschalten von Instrumentenflug auf Sichtflug)
	15.33:00	PF	Immer noch von achten.
	15.33:02	PF	Rate of descent is out of limits. (Sinkgeschwindigkeit außerhalb der Grenzwerte)
	15.33:03	PNF	Ja, elf hundert Fuß. [...]
	15.33:15	ACO	Four hundred.
	15.33:20	ACO	Three hundred.
	15.33:26	ACO	Two hundred.
		TWR	< Funkverkehr in polnischer Sprache >
	15.33:29	PNF	Von rechts kommt jetzt.
	15.33:31	PF	Jetzt kommt die windshear.
	15.33:33	ACO	One hundred.
	15.33:36	PNF	Dreht, dreht (turning, turning)
	15.33:37	ACO	Fifty.
	15.33:39	ACO	Thirty.
	15.33:40	ACO	Retard, retard. < Clack >
	15.33:49	PF	Brems mal mit.
	15.33:52	PF	Full braking.
	15.33:56	PNF	Reverse auf?
	15.33:57	PF	Ja's voll. < Clack >
	15.34:01	PNF	Hundert.
	15.34:02	PF	Weiter bremsen.
	15.34:05	PF	Scheiße.
	15.34:06	PF	Was machen wir jetzt?
	15.34:08	PNF	Tja, du kannst nix mehr machen.
	15.34:10	PF	Ich möchte nicht da gegen knallen.
	15.34:11	PNF	Dreh'n weg.
	15.34:12	PF	Was?
		PNF	Dreh ihn weg.
	15.34:16	PF	Scheiße!
	15.34:17		< Krachen >

Man braucht den Inhalt dieser Kommunikation nicht in allen Details zu verstehen, um spontan die Behauptung zu wagen, dass sie von keinem anderen Lebewesen als vom Menschen und auch von keinem Automaten vollzogen werden könnte.

Was ist in dieser Kommunikation und in dem Verhalten, in das sie eingebettet ist, alles im Gange? Offenbar sind Aktivitäten mehrerer Verhal-

tensbereiche der Beteiligten zu beobachten. Was die Transkription am augenfälligsten macht, ist, dass alle Beteiligten Äußerungen produzieren. Sie verfügen mithin über die dazu erforderlichen Sprechorgane und die Fähigkeit, mit ihnen lautsprachliche Äußerungen zu artikulieren. Die Äußerungen sind sprachlich wohlgeformt, ein Hinweis auf die Beherrschung der verwendeten Sprache, d. h. Kenntnis der Wörter und der Gesetzmäßigkeiten ihrer Zusammenfügung zu Sätzen und Satzfolgen. Pilot, Copilot und der Sprecher in der Anflugkontrolle verwenden abwechselnd Äußerungen in zwei Sprachen, was zum einen die entsprechenden Kenntnisse und Fertigkeiten einer Person in mehreren Sprachen belegt, zum anderen annehmen lässt, dass die Sprecher die Veranlassungen kennen, mal die eine, mal die andere Sprache zu verwenden, und diese Veranlassungen kalkulieren und befolgen können. Mehrsprachigkeit in diesem Sinne kommt übrigens auch im Reich der Vögel vor (vgl. Williams/Zeigler/Marler 2004).

Sprachproduktion: Welche Einzelaktivitäten des Weiteren die gesamte Produktion einer Äußerung ihrerseits enthält, wird in Kapitel 4 genauer behandelt. Vorläufig genügt es festzuhalten, dass dem Artikulieren der Äußerung Denkvorgänge, also kognitive Prozesse, vorangehen. Der Sprecher plant offenbar bei jeder Äußerung, an wen sie gerichtet ist, was mitgeteilt werden soll, welchem Ziel die Äußerung dienen soll, z. B. einer Bitte, einer Frage, einer Aufforderung etc. Das alles lässt sich unter das sprachliche Verhalten unterordnen.

Sprachverstehen: Dazu zählen auch die Aktivitäten des Verstehens, die dem Dokument nur indirekt zu entnehmen sind. Um 15.29:42 sagt der Pilot ›Stimmt‹, um 15.30:17 der Copilot ›Genau‹. Äußerungen wie diese zeigen, dass seitens des Sprechers auf eine vorangegangene Äußerung des Kommunikationspartners reagiert wird, diese also wahrgenommen, verstanden und ihr Inhalt auf seine Zustimmungsfähigkeit durch den Adressaten beurteilt worden ist. Auch dem Verstehen und den dabei stattfindenden Einzelaktivitäten wird ein eigenes Kapitel (Kap. 5) gewidmet sein.

Eingebunden ist das sprachliche Verhalten in diesem Beispiel in verschiedene nicht-sprachliche Aktivitäten von Sprecher bzw. Hörer. Da ist zunächst einmal das gesamte motorische Verhalten, das Bewegen von Körper und Körperteilen, ferner die laufende auditive, visuelle, taktile und olfaktorische Wahrnehmung.

Direkte Hinweise auf die visuelle Wahrnehmung finden sich in den Äußerungen 15.29:40 *Bist leicht überm Glideslope* und 15.32:57 *Field in sight*. Die erste zeigt an, dass der Copilot einen Zustand eines Anzeigeinstruments gesehen hat, die zweite, dass er die Landebahn direkt sieht. Visuelle Sprachverarbeitung, also Lesen, findet in der obigen Kommunikation nicht statt. Sie ist aber, wie man sich denken kann, auch Bestandteil der menschlichen Sprachfähigkeit und wird in Kapitel 5.6 behandelt.

Dem sprachlichen Verhalten nur mittelbar zu entnehmen sind Vorgänge des nicht-sprachlichen Denkens, des Sozialverhaltens und des Empfindens. Nicht-sprachliches Denken findet sich z. B. mittelbar belegt in der Äußerung 15.29:58, mit der der Pilot zu verstehen gibt, dass die Entscheidung, zeitweise oberhalb des Radarleitstrahls zu fliegen, der Sache nach begründet ist. Der Gedankengang des Piloten ist auch ohne flugtechnisches Wissen ein Stück weit zu verstehen. Jedenfalls deutet die Äuße-

Beteiligte Systeme

rung darauf hin, dass angesichts der Windverhältnisse, der Sinkrate, der Entfernung von der Landebahn und der Fluggeschwindigkeit eine vollständige Annäherung an den Leitstrahl mit dem Landemanöver als Ganzem nicht zu vereinbaren ist. Ausdruck von Sozialverhalten ist eine Äußerung wie ›Danke‹ (PF, 15.29:52) und emotionales Verhalten schließlich belegt mehr oder weniger indirekt die Äußerung 15.34:10 und unmissverständlich die von 15.34:16.

Sprachliche Kommunikation: Nach Maßgabe und Vermögen seiner artspezifischen Ausstattung fügt sich das sprachliche Kommunizieren des Menschen in sein Gesamtverhalten ein:

- Vorgänge der peripheren Sprachorgane
- Wahrnehmung
- Gedächtnisvorgänge
- Denken im weiten Sinne, d. h. von den autonomen, dezentralen Informationsverarbeitungen bis zu den bewussten Problemlöseprozessen
- Soziale Orientierung
- Gefühlsleben

1.2 | Gegenstand und Ziele der Psycholinguistik

Wenn man sich daran macht, die Abläufe des sprachlichen Verhaltens und seine Voraussetzungen genauer zu untersuchen, betrachtet man es aus einer der genannten Perspektiven. Die psycholinguistische Perspektive ist die, sprachliches Verhalten als einen Bestandteil des menschlichen Denkens, der Kognition, zu betrachten. Der griffigen Hilfsvorstellung folgend, dass wissenschaftliches Arbeiten systematisches Suchen nach Antworten auf offene Fragen ist, ergibt sich:

Zum Begriff

Psycholinguistik ist das systematische Suchen nach einer Antwort auf die Frage: Welches Wissen und welche kognitiven Verarbeitungssysteme machen die Sprachfähigkeit des Menschen aus?

Komponenten des sprachlichen Wissens: Die Frage ist vielleicht zu umfassend formuliert, so dass ihr keine fassbare Vorstellung vom Gegenstand psycholinguistischer Arbeit zu entnehmen ist. Und tatsächlich wird sie in dieser Allgemeinheit auch nicht in einem Schritt angegangen, jedenfalls heutzutage nicht mehr. Ein Ergebnis der psycholinguistischen Forschung der letzten rund hundert Jahre ist, dass die Sprachfähigkeit des Menschen auf dem Zusammenwirken mehrerer Wissensbestände und Verarbeitungssysteme beruht. Dasselbe gilt für die Vorgänge des Erwerbs der Sprachfähigkeit und ebenso für das seinerseits wieder umfassende Feld der Störungen.

Betrachten wir noch einmal die Kommunikationsvorgänge in dem obigen Cockpit-Beispiel. Was die Beteiligten da sprachlich tun, lässt auf verschiedene Bestände an Wissen und verschiedene Teilfertigkeiten schließen. Grob zu unterscheiden sind in erster Annäherung die Kenntnis von

Wörtern einerseits und Kenntnisse davon, wie welche Wörter kombiniert werden können und müssen, damit eine Äußerung herauskommt, die dem Angesprochenen gerade die Information zu verstehen gibt, die der Sprecher aktuell ausdrücken will. Die Kenntnis von Wörtern, ihren Bedeutungen, ihren syntaktischen Bedingungen, ihrem inneren Aufbau (kannst, komm-t, Daumen-breit-e) und ihrer Ausdrucksseite, d. h. Lautung und Schreibung, bildet zusammengefasst **das lexikalische Wissen**. Es ist statisch, verändert sich beim Erwachsenen eher langsam, jedenfalls verglichen mit dem Wissen über Aktienkurse, Benzinpreise oder die Stimmung der nächsten Mitmenschen; kurzum, es ist ein relativ stabiles Langzeitwissen.

Das Wissen der grammatischen Kombinatorik ist zwar auch langfristig – der Mensch hat es auch dann, wenn er nicht gerade redet oder mit ihm geredet wird – es ist aber dem Inhalt nach ein prozedurales Wissen um das Zusammenpassen sprachlicher Einheiten unter strukturellen Bedingungen. Wer die Kombinationsregeln des Deutschen beherrscht, verfügt z. B. über die Kenntnis, dass der Artikel vor einem Nomen, mit dem er zusammengehört, sich hinsichtlich seiner Genusform nach dem Nomen richtet und nicht umgekehrt. Dieser Kenntnis gemäß verwendet der Copilot in der Äußerung 15.29:46 die Form *eine* und nicht *ein* oder *einen*, weil *eine* die feminine Form des Akkusativs ist, was dem Genus von *Daumenbreite* entspricht. Dieses sog. grammatische Wissen ist ebenfalls äußerst reichhaltig und kompliziert, wessen man sich als Muttersprachensprecher kaum bewusst ist. Eine Ahnung davon entsteht mitunter, wenn man eine fremde Sprache zu lernen beginnt, sagen wir als Deutschsprachiger das Italienische, und wenn man dann, im vermeintlichen Besitz des qualvoll Erworbenen, in einem Hotel in der Toscana an der Rezeption ein ruhiges Zimmer für zwei Personen, drei Nächte und mit nur einem Frühstück bestellen will, nach Wörtern und Konstruktionen, Flexionsformen und der richtigen Aussprache sucht. Dass der Mensch sich den Inhalt seines grammatischen Wissens nicht bewusstmachen kann und dass er auch nicht auf anderem Wege unmittelbar beobachtet werden kann, stellt die Wissenschaftler vor besondere methodische Herausforderungen. Das grammatische Wissen ist nur auf dem Wege systematischer Beobachtung des sprachlichen Verhaltens und der darin belegten Äußerungen möglich. In einer einzelnen Äußerung wird aber jeweils nur von einem winzigen Ausschnitt des gesamten grammatischen – und übrigens auch lexikalischen – Wissens Gebrauch gemacht.

Prozeduren der Sprachverwendung: Ebenso wie das sprachliche Wissen selbst sind dem Menschen auch die meisten Prozeduren der jeweiligen Sprachverwendung nicht bewusst. Diese machen in ihrer Gesamtheit neben dem sprachlichen Wissen den anderen Hauptanteil der Sprachfähigkeit aus. Eine entsprechende Unterscheidung ist in der Linguistik schon zu Anfang des 20. Jahrhunderts durch den Indogermanisten und Sprachtheoretiker Ferdinand de Saussure getroffen worden. In dem nach seinen Vorlesungen geschriebenen Buch *Grundlagen der allgemeinen Sprachwissenschaft* wird die Sprachfähigkeit als ganze als *faculté de langage* bezeichnet, der Wissensbestand als *la langue* und die Rede als *la parole*.

Spracherwerb

Vom sprachlichen Wissen und der Fähigkeit, es zu verwenden, führt ein direkter Weg zu der Frage, wie beides erworben wird. Ein Hauptgebiet der Psycholinguistik untersucht die Frage, wie die Sprachfähigkeit des Menschen sich entwickelt

Beispiel Beobachtung des Spracherwerbsverlaufs

Clara und Wilhelm Stern, ein Forscherehepaar in Breslau Anfang des 20. Jahrhunderts, haben ihre drei Kinder beim Erwerb der Sprache genau beobachtet, besonders genau Hilde, geb. am 7. April 1900. Über Hildes Sprechen und Verstehen etwa 12 Monate nach der Geburt berichten die Sterns die folgenden Beobachtungen:

a) Verstehen:

Gesagt wird: *Wer bin ich denn?*

Hilde: *mamma*

Gesagt wird: *Gib einen Kuss!*

Hilde: Macht eine schmatzende Kussbewegung in die Luft.

b) Sprechen:

Hilde: *puppe*.

Kommentar der Eltern: *puppe* wurde anfangs zu fast allen Spielobjekten gesagt, zu richtigen Puppen ebenso wie zum Stoffkaninchen und zur Stoffkatze, dagegen nicht zu ihrem Glöckchen (Hervorhebung von uns, RD und JG).

Hilde: *babà* < Betonung auf der letzten Silbe > .

Kommentar: *babà* sagte Hilde öfter, wenn sie etwas in den Mund nahm, was ihr nicht schmeckte.

(vgl. Stern/Stern 1920, S. 18–20)

Vergleichen wir diese kommunikativen Aktivitäten von Hilde mit solchen zweieinhalb Jahre später (vgl. Stern/Stern 1920, S. 72)

Situation: Die Mutter kommt zu Hilde ins Zimmer, die mit einer Puppe beschäftigt ist; es entsteht ein Gespräch:

Mutter: *Ist das auch dein Puppenkind?*

Hilde: *Nein, die kleine ist bloß für den Günther* < der kleine Bruder > .
komm, hinsetzen Püppelchen. ich wer dich halten; ja?

M.: *Was hast du gespielt mit Püppchen?*

H.: *Ich hab schlafen elegen; sie is schon ausgeschlafen.*

M.: *War sie artig?*

H.: *Ja sie war artig – und dann hat sie geschrieen.*

Die Sprachentwicklung ist offensichtlich. In zweieinhalb Jahren hat Hilde fast alle Sprachkenntnisse erworben, über die unsere Piloten verfügen, jedenfalls, wenn man vom Fachwortschatz und von den Englischkenntnissen absieht. Sie kann Äußerungen anderer verstehen und angemessen darauf reagieren. Und das alles, obwohl sie im Erwerb sprachlichen Wissens weder besonders ausgebildet, noch in der Verwendung systematisch angeleitet worden ist. Und sie hat diesen Erwerbsvorgang in einem Alter

bewältigt, in dem ihr der Erwerb des Schachspielens oder des Lösen von Gleichungen nicht so flott von der Hand gegangen wäre. Eine Sonderstellung hat die Sprachfähigkeit außerdem damit, dass sie, sofern nicht Krankheit oder Störungen sie verhindern, sozusagen naturwüchsig bei jedem Kind entsteht.

Die Fähigkeiten der produktiven und rezeptiven Sprachverwendung umfassen in sich jeweils Gruppen von funktional zusammenhängenden Prozeduren, die scheinbar relativ autonom, d. h. unabhängig voneinander und unabhängig von zentraler Steuerung ›arbeiten‹. Werfen wir dazu noch einen Blick auf die kurzen Redewechsel in der Cockpit-Kommunikation, 15.29:38 bis 15.29:52. *Bist leicht über'm Glideslope*. Was geschieht hier alles? Der Copilot nimmt eine Anzeige auf einem Gerät wahr und folgert mit seinem Wissen, dass Ist- und Sollzustand des Flugzeugs nicht übereinstimmen. Er stellt sich daraufhin die kommunikative Aufgabe, den Piloten auf diese Abweichung aufmerksam zu machen. Beobachten lässt sich nur das Ergebnis der Äußerungsproduktion, die Äußerung selbst, technisch gesprochen, eine Folge von akustischen Erscheinungen, Schallschwingungen in der Luft der Flugkanzel. Zwischen dem Zeitpunkt der kommunikativen Aufgabe und dem Ende des Sprechens haben eine Reihe von kognitiven Prozessen stattgefunden; darunter die Umwandlung eines mentalen Bildes, hier das von der Abweichung zwischen Ist- und Soll-Lage des Flugzeugs, in einen Typ von Information, der sprachlich zum Ausdruck gebracht werden kann, ferner die zeitliche und räumliche Bezugsetzung dieses Gedankens auf die aktuelle Sprechsituation, das Finden der Wörter im lexikalischen Wissen, das Zusammenbauen des Satzes, die Erstellung einer Vorstellung von einer lautlichen Realisierung, die Aktualisierung der motorischen Artikulationssteuerung etc. Das alles hat schätzungsweise nicht länger als zwei Sekunden gedauert. Mindestens ebenso bemerkenswert ist, dass die Reaktion des Piloten ohne nennenswerte Verzögerung darauf erfolgt. Immerhin muss auch in seinem geistigen Programm des Sprachverstehens eine Kette von Prozessen stattgefunden haben. Dass derartige Vorgänge zeitlich so dicht hintereinander möglich sind, deutet darauf hin, dass der Verstehensvorgang zeitlich überlappend mit dem Sprechvorgang abläuft und umgekehrt die Sprechplanung des Antwortenden wahrscheinlich beginnt, bevor der Verstehensvorgang vollständig abgeschlossen ist.

Natürlich kann die Entwicklung der Sprachfähigkeit auch gestört sein. Ebenso kann schon erworbene Sprachfähigkeit wieder verlorengehen.

Sprachproduktion eines Aphasiepatienten

Beispiel

Stark/Stark (1991, S. 280) dokumentieren eine Nacherzählung von »Rotkäppchen und der Wolf« durch einen 46-jährigen Patienten, der infolge eines Schlaganfalls an einer sog. Aphasie leidet.

Ich zähle Rotkäppchen. ... <... = längere oder kürzere Pause> Ah Mutter . ah . backt Kuchen . und . andere Sachen .. Und . schickt eine Mädchen weg. .. Und ein Wolf ... kommt ... und sieht Rotkäppchen. .. Und ah .. ah .

*ah . Blumen ah / < Abbruch > Und ah und .. p
flückt . Blumen. . Und ah . ah Wolf geht. Und ah ah und ah . ah ah . eine
Mutter. ... Und pflückt Blumen. viele Blumen. . Und geht und nichts mehr.
.. Rotkäppchen .. und ah .. hören . hörn und sehn . und ah/ Weiß ich nicht!
Ah Kopf und und .. Mutter .. ah ah gibt sie mehr Milch.*

Krankheit und Störung der Sprachfähigkeit (s. Kap. 6) zu untersuchen, ist zum einen natürlich direkt motiviert durch die gewonnenen Einblicke und die damit verbundenen Heilungsmöglichkeiten. Darüber hinaus liefert die Beobachtung, welche Komponenten der Sprachfähigkeit separat geschädigt werden können, auch Anhaltspunkte für das Verständnis ihres Aufbaus und ihrer Funktionsweise beim Gesunden.

Ein eigenes Kapitel über die **Geschichte der Psycholinguistik** gibt es in diesem Buch nicht. Auf Arbeiten, die die Erkenntnis ihrer Zeit nachhaltig erweitert haben, wird innerhalb der Darstellung der Teilgebiete zurückgeblickt. Eine umfassende Darstellung der Geschichte der Psycholinguistik bietet Levelt (2012).

1.3 | Wissenschaftliche Erkenntnis in der Psycholinguistik

Die Sprachfähigkeit als »Black Box«-Phänomen: Begründet durch die Tatsache, dass Vorhandensein und Beschaffenheit geistiger Phänomene, die mit der Sprachfähigkeit des Menschen in Verbindung stehen, unserer Beobachtung eben nicht in dem Sinne unmittelbar zugänglich sind, wie der Aufbau des und die Abläufe in unserem Sonnensystem oder die Tätigkeit des Herzmuskels, ergibt sich für den wissenschaftlichen Erkenntnisgewinn in der Psycholinguistik eine methodische Besonderheit, die oft mit der Metapher der »Black Box« beschrieben wird. Gemeint ist hiermit, dass die fehlende Möglichkeit der direkten Beobachtung eine Situation darstellt, in der man eine »Kiste« vor sich hat, von der man weiß, dass sie einen spezifischen, normalerweise relativ gut beschreibbaren Input in einen relativ gut beschreibbaren Output verwandelt, man aber nicht weiß, wie genau dieser Umwandlungsprozess von staten geht. Wie andere Disziplinen der Kognitionswissenschaft versucht auch die Psycholinguistik die in der »Black Box« ablaufenden Prozesse zu rekonstruieren.

Methodenentwicklung: Bei der Planung einer neuen Untersuchung ist äußerste Flexibilität und Kreativität im Ersinnen von aussagekräftigen Verfahren, die möglichst weitreichende und allgemeine Behauptungen rechtfertigen, nötig. Beide Aspekte sollen in den folgenden Kapiteln deutlich werden, indem Methodisches jeweils im Zusammenhang mit spezifischen Befunden behandelt wird. Die Vorgehensweise bei der psycholinguistischen Methodenentwicklung und -anwendung richtet sich nach dem Inhalt der Fragestellung, dem Umfang des darüber schon vorliegenden Wissens und der technischen und ethischen Machbarkeit des Verfahrens. Als Faustregel kann gelten, dass zur Untersuchung qualitativer Eigenschaften

der Sprachfähigkeit wie in allen Naturwissenschaften Methoden der vergleichenden Datenanalyse geeignet sind.

In erster Annäherung lassen sich **drei Verfahrensweisen** feststellen, in deren Rahmen Untersuchungsmethoden entwickelt und angewendet werden. Die erste betrifft die Verhaltensbeobachtung in natürlicher oder kontrollierter Umgebung, die zweite die Beobachtung der Struktur und Tätigkeit des Zentralnervensystems und die dritte die Simulation sprachlichen Verhaltens.

Die Methode der Verhaltensbeobachtung umfasst eine Spannweite, die von der Beobachtung typischen sprachlichen Verhaltens in gewohnter Umgebung (die Tagebuchaufzeichnungen der Sterns, siehe oben) über die Zusammenstellung und Auswertung spezifischer sprachlicher Phänomene im Rahmen sogenannter Korpora (Korpora zu Versprechern) bis hin zur Messung spezifischer Reaktionen eines Versuchsteilnehmers im Experiment reicht. Alle drei Spielarten der Verhaltensbeobachtung sind für die psycholinguistische Forschung unverzichtbar, wobei zu beachten ist, dass die Verhaltensbeobachtung im Experiment wohl am häufigsten zum Einsatz kommt.

Neurowissenschaftliche Methoden: Verfahren wie die funktionale Magnetresonanztomographie (fMRT), die Positronen-Emissions-Tomographie (PET), Magnetoenzephalographie (MEG) oder die Messung ereignisrelatierte Potentiale (ERP) werden zur Messung der Hirnaktivität während der Bearbeitung sprachlicher oder sprachrelatierter kognitiver Prozesse eingesetzt. Hier ist erst ein Anfang gemacht. Die Aussicht auf Fortschritte wird u. a. davon bestimmt, wie genau und mühelos neuronale Aktivitäten gemessen werden können. Anfang der 1970er Jahre schrieb Harry Whitaker, ein Pionier der Neurolinguistik: »Neurophysiology is faced with a peculiar surfeit of data: it is possible to record all the activity of a single neuron or the activity of collections of neurons [...]. The problem is, what do these spike trains mean? They are undoubtedly the coded form of information, but yet there is little to indicate what the units of communication are.« (Whitaker 1971, S. IX-X).

Simulation sprachlichen Verhaltens: Die Existenz leistungsstarker Computer seit etwa den 1970er Jahren hat es der Psycholinguistik ermöglicht, sich zunehmend auch der mathematischen Simulation sprachlichen Verhaltens zuzuwenden. Das Ziel solcher Methoden ist es bisher gewesen, zunächst einen ganz spezifischen Bereich der Sprachverarbeitung sozusagen ›außerhalb‹ des menschlichen Gehirns nachzustellen. Dabei wird der interessierende Ausschnitt der kognitiven Gesamtstruktur mit Hilfe sogenannter Netzwerke abgebildet, die ihrerseits aus Knoten und Verbindungen zwischen Knoten bestehen. Der zentrale Gedanke bei solchen Simulationen ist, dass der reale Zeitbedarf des menschlichen kognitiven Systems, für die Reaktion auf einen Reiz benötigt wird, aus der Existenz eines Knotens sowie dessen Fähigkeit Aktivationsenergie an einen anderen Knoten abzugeben, berechnet werden kann. Die Güte der Simulation wird dann für gewöhnlich durch einen Vergleich mit Daten aus Reaktionszeitexperimenten bestimmt. Zu den einflussreichsten Simulationen in der Psycholinguistik gehören die Konzeption von Dell (1986) zur lautlichen Enkodierung während der Sprachproduktion, das WEAVER-Modell (mehrere Ver-

Methoden

sionen) zur Wortselektion und zur lautlichen Enkodierung (vgl. u. a. Roelofs 1992; 1996; Levelt/Roelofs/Meyer 1999), das Modell von Chang/Dell/Bock/Griffin (2000) zur Generierung der syntaktischen Struktur eines Satzes sowie verschiedene **constraint-based-Modelle** zur syntaktischen Verarbeitung von Sätzen während des Sprachverstehens (vgl. u. a. McRae/Spivey-Knowlton/Tanenhaus 1998).

Skepsis als Grundhaltung: Die mehrfach erwähnte Tatsache, dass die kognitiven Zustände und Vorgänge der Sprachfähigkeit nicht direkt zu beobachten sind, begründet eine gewisse Skepsis. Und diese Skepsis ist als Grundhaltung in der heutigen Psycholinguistik auch generell anerkannt und gepflegt; so kommt der Methode, dem Verfahren, mit dem eine Beobachtung gewonnen wird, eine wesentliche Bedeutung in der Beurteilung der Behauptungen zu, die sich darauf stützen. Ein psycholinguistischer Befund ist nur in dem Maße aussagekräftig, wie die verwendete Methode ihn aussagekräftig macht. Die Notwendigkeit gewissenhafter methodischer Reflexion zeitigt in der psycholinguistischen Arbeit zwei komplementäre Wirkungen, je nachdem, ob sie sich auf die Bewertung gewonnener Ergebnisse richtet oder auf die Planung neuer Untersuchungen. Indem eine psycholinguistische Erkenntnis Voraussagen und damit Allgemeingültigkeit immer nur in dem Rahmen zulässt, der mit den Beobachtungsmodalitäten gesteckt ist, wirkt das Methodenbewusstsein in der Beurteilung eines Befundes häufig restriktiv. Eine Behauptung wird zunächst einmal nur für die Verhältnisse für gültig gehalten, unter denen sie gewonnen worden ist, nämlich eben genau den im Experiment gegebenen.

1.4 | Methoden

Mit dem Inhalt dieses Abschnitts soll die Lektüre der folgenden Kapitel erleichtert werden, indem grundlegende Begriffe vorab eingeführt werden. Eine Einführung in die Methoden des Faches ist nicht beabsichtigt. Stattdessen werden methodologische Details im Zusammenhang mit den illustrierenden Experimentbeschreibungen en passant behandelt. Zur systematischen Vertiefung eignen sich die Handbücher von Gazzaniga (2014) und Gullberg/Indefrey (2006).

1.4.1 | Beschreiben, Erklären, Vorhersagen

Als Startpunkt für die Entwicklung und Anwendung typischer psycholinguistischer Testverfahren, wie der Reaktionszeitmessung oder der Messung von Blickbewegungen während der Beschreibung visuell dargebotener Stimuli, dient wie auch in allen anderen empirisch arbeitenden Wissenschaften eine spezifische Theorie. Eine Theorie kann als System von Aussagen aufgefasst werden, mit deren Hilfe bestimmte Phänomene beschrieben und erklärt sowie ihr Auftreten oder ihre spezifische Ausprägung vorhergesagt werden können.

Eine Theorie etwa über den Gesamtprozess der Sprachproduktion muss es ermöglichen, alle beteiligten Teilprozesse zu beschreiben, d. h.

alle einzelnen Schritte von der Vorbereitung einer Idee bis hin zur Bewegung der Artikulationsorgane, mit der die Produktion von Lauten bewerkstelligt wird, um die Idee an Mitmenschen zu übermitteln. Erst wenn das Feld sozusagen vollständig »kartographiert« ist, können Erklärungen über mögliche Zusammenhänge gemacht werden, zum Beispiel, dass ein bestimmtes in der Theorie beschriebenes Phänomen einem anderen zeitlich vorausgeht. In Hinblick auf eine Theorie über die menschliche Sprachproduktion könnte eine Erklärung sein, dass der Abruf der Bedeutung eines Wortes dem Abruf der Wortform vorausgeht. Ist eine Theorie dann so weit entwickelt, dass Zusammenhänge erklärt werden können, können auf Grundlage der Erklärungen bestimmte Vorhersagen gemacht werden. Wenn es, um beim Beispiel zu bleiben, stimmt, dass die Wortbedeutung im Sprachproduktionsprozess zeitlich vor der Wortform aktiviert wird, dann sollte es gelingen, unter Verwendung einer geeigneten Methode, die Aktivierung der Wortbedeutung zu einem früheren Zeitpunkt zu stören als die Wortform, beides relativ zu einem Referenzzeitpunkt (vgl. Schriefers/Meyer/Levelt 1990).

Doch wie kommt man überhaupt zu einer Theorie? Der erste Schritt besteht für gewöhnlich in der Exploration. Die Tagebuchaufzeichnungen der Sterns (siehe oben) sind ein gutes Beispiel für diese Herangehensweise.

Die Exploration führt über die Klassifikation der Einzelfälle zu Kategorien, die aus irgendwelchen Erwägungen als erkenntnisrelevant gelten können. In der Grammatikforschung sind dies z. B. die Vorkommensbedingungen von Wörtern, die zu Wortklassen (Nomen, Adjektiv, Adverb, Artikel) führen, in der Phonologie die Verteilung von Lauten, die bei sonst gleichem lautlichen Rest des Wortes eine semantische Unterscheidung anzeigen (*Haus – Maus*). Sind alle Fälle kategorisiert, wird nach Zusammenhängen zwischen den Kategorien gesucht, z. B. derart, dass das Auftreten von Substantiven von Eigenschaften der Verben (transitiv, intransitiv) abhängt und nicht umgekehrt. Die Exploration kann im Übrigen auch unter experimentellen oder quasiexperimentellen Bedingungen stattfinden.

Der Hypothesentest ist ein Weg, der voraussetzt, dass der betrachtete Gegenstand schon in vielerlei Hinsicht gut bekannt ist, nur eben in einer Hinsicht noch unklar. Aber auch bezüglich dieser Hinsicht müssen gut begründete Antwortalternativen vorliegen. Zum Beispiel könnte es durch verschiedene Evidenz als relativ gesichert gelten, dass die Bedeutung eines Wortes zeitlich vor der Wortform aktiviert wird, aber gleichzeitig könnte Unklarheit darüber herrschen, wie genau die Aktivierung von staten geht: Wenn mehrere Kandidaten auf der Bedeutungsebene in Wettbewerb stehen (sage ich *Couch* oder *Sofa*?), »wartet« das Sprachproduktionssystem dann zunächst bis ein Kandidat ausgewählt ist bevor die Wortform aktiviert wird oder wird für beide Wortbedeutungen von Anfang an zumindest teilweise die Forminformation aktiviert? Dies ist ein typisches Szenario dafür, dass eine Theorie bereits einen gewissen Entwicklungsstand erreicht hat, über ein Detail, welches spezifische Zusammenhänge betrifft, aber noch mehrere Antwortalternativen im Spiel sind. In solchen Fällen richtet sich die psycholinguistische Methodik auf den Hypothesentest, wobei die zu testenden Hypothesen aus den theoretischen Alternativ-erklärungen abgeleitet werden.

Zur Vertiefung

Schritte des empirischen wissenschaftlichen Arbeitens

Die hier skizzenartig beschriebene Herangehensweise in der psycholinguistischen Forschung führt zu einem ganz bestimmten Arbeitsstil, der mindestens die sechs folgenden Arbeitsschritte umfasst:

- die Identifikation eines Phänomens in einem spezifischen Bereich und seine Abgrenzung gegenüber anderen Phänomenen,
- die Formulierung einer theoretischen Hypothese in Bezug auf das Phänomen vor dem Hintergrund eines theoretischen Modells (bzw. einer nicht fachfremden Theorie),
- die Operationalisierung, d. h. die Übertragung/Abbildung theoretischer Begriffe auf beobachtbare (messbare) Variablen,
- die Wahl einer geeigneten Beobachtungsmethode,
- die Erhebung von Daten und deren Analyse,
- die Interpretation der Befunde in Hinblick auf die theoretische Hypothese sowie das Inbeziehungsetzen der Ergebnisse mit der aktuellen Forschungslage.

1.4.2 | Messen, Messmethoden und Güte der Messung

Experimentieren

Ein wesentlicher Punkt, der sich aus der bereits mehrfach genannten Tatsache ergibt, dass kognitive Prozesse nicht direkt beobachtbar sind, ist, dass objektives Messen in der Psycholinguistik den Schritt der sogenannten **Operationalisierung** voraussetzt.

Variablen zur Beschreibung von Merkmalen: Vorab sollte allerdings der für jede empirisch arbeitende Wissenschaft essentielle Begriff »Variable« eingeführt werden. Eine **Variable** ist ein wissenschaftliches Hilfskonstrukt, mit dessen Hilfe Eigenschaften bzw. Merkmale von Phänomenen erfasst werden können. Zum Beispiel könnte die Variable »Körpergröße« verwendet werden, um ein bestimmtes Merkmal von Menschen zu erfassen. Dasselbe Ziel könnte mit den Variablen »Haarfarbe« oder »Augenfarbe« erreicht werden. Eine Variable dient dabei dazu, genau ein Merkmal zu erfassen, welches in unterschiedlichen Ausprägungen auftreten kann. Eine wichtige Eigenschaft einer Variable ist daher, dass sie immer mindestens zwei **Ausprägungen** hat. Die Ausprägungen der Variable »Körpergröße« könnten entweder lediglich in »groß« bzw. »klein« bestehen oder in »180 cm«, »18 cm«, usw. Variablen können nach Bedarf kombiniert werden, wodurch unter Umständen neue Variablen entstehen. Eine wichtige Unterscheidung ist, auf was sich eine Variable im Rahmen einer empirischen Untersuchung bezieht. Eine Gruppe von Variablen, die sogenannten **unabhängigen Variablen**, beschreiben Merkmale eines Reizes oder einer Situation, unter denen eine Beobachtung, zum Beispiel die Messung einer bestimmten Reaktion, stattfindet. **Abhängige Variablen** auf der anderen Seite erfassen die Ausprägung eines Merkmals, mit dem das interessierende Verhalten beobachtet bzw. gemessen wird. Abhängige Variablen hängen von (den Ausprägungen der) unabhängigen Variablen ab, nämlich insofern, dass die in einem Experiment auf kontrollierte Weise manipulierten Ausprägungen der un-

abhängigen Variablen die Ausprägung der abhängigen Variable bestimmen. Beide Typen von Variablen spielen sowohl auf der theoretischen als auch auf der empirischen Ebene eine Rolle, wobei der Zusammenhang zwischen beiden Ebenen durch die sogenannte Operationalisierung hergestellt wird.

Operationalisierung: Unter diesem Begriff versteht man das Beobachtbar-Machen theoretischer Variablen. Eine theoretische psycholinguistische Hypothese könnte zum Beispiel lauten, dass das Erkennen seltener Wörter mehr Verarbeitungsaufwand benötigt als das Erkennen häufiger Wörter. Um diese Hypothese empirisch zu überprüfen, müssen im Zuge der Operationalisierung alle theoretischen Begriffe bzw. Variablen so transformiert werden, dass sie objektiv beobachtbar werden. Eine Möglichkeit, die Seltenheit oder Häufigkeit von Wörtern zu operationalisieren, besteht in der Bestimmung der relativen Wortfrequenz. Man könnte hierfür das Vorkommen von Wörtern in einem Korpus zählen, in dem Tausende von Texten gespeichert sind, und in Bezug zur Gesamtgröße des Korpus setzen. Dadurch würde pro Wort ein Wert zur Verfügung stehen, wie zum Beispiel 100 Mal pro 10 Millionen Wörter, der es einem erlauben würde, objektiv einzuschätzen, ob ein Wort ›häufig‹ oder ›selten‹ ist. Viele der heutigen Online-Wörterbücher geben bereits die relative Wortfrequenz zu jedem Wort an.

In Bezug auf den zweiten theoretischen Begriff in der hier verfolgten Beispielhypothese, den **Verarbeitungsaufwand**, müsste ein Psycholinguist auf ähnliche Weise vorgehen. Wie kann der Verarbeitungsaufwand objektiv angegeben werden? Die entscheidende Erkenntnis hierzu stammt bereits aus dem 19. Jahrhundert und wurde von Franciscus Cornelis Donders (1868) in einer psychologischen Fachzeitschrift publiziert: Kognitive Verarbeitungsprozesse benötigen Zeit und komplexere Verarbeitungsprozesse benötigen mehr Zeit als weniger komplexe. In unserem Beispiel könnte der Verarbeitungsaufwand also durch die Messung der Zeit operationalisiert werden, die zwischen dem Beginn und dem Abschluss der Reizverarbeitung vergeht. Die operationalisierte Version der oben vorgestellten Hypothese würde demnach lauten: Die Verarbeitung hochfrequenter Wörter benötigt weniger Verarbeitungszeit als die Verarbeitung niederfrequenter Wörter.

Offline- vs. Online-Messung: Eine wichtige Unterscheidung bei den psycholinguistischen Messmethoden betrifft, was genau gemessen wird. Bei der sogenannten Offline-Messung wird das Resultat der ablaufenden kognitiven Prozesse gemessen, wohingegen bei der Online-Messung der zeitliche Verlauf der kognitiven Prozesse selbst im Vordergrund steht. Hierzu zählen insbesondere auch alle Messungen mit neurowissenschaftlichen Verfahren. Die Art der Messung bzw. die Entscheidung für ein spezifisches Messverfahren sollte immer in Hinblick auf verschiedene Kriterien bewertet werden, welche es zusammengefasst gestatten, eine Aussage über die Güte des Messverfahrens zu machen.

Beispiele

Beispiel für Offline-Messung

- Wie oft wurde in einem Experiment eine bestimmte syntaktische Struktur verwendet? – z. B. strukturelles Priming (vgl. Bock 1986; Pickering/Ferreira 2008; Gerwien/Flecken 2015).
- Wie häufig wurde eine bestimmte Region auf verschiedene Stimuli insgesamt angeschaut? – Messung der Blickhäufigkeit mit Eye-Tracking (vgl. van der Meulen/Meyer/Levelt 2001; Stutterheim/Andermann/Carroll/Flecken/Schmiedtová 2012).
- Wie häufig wird ein Satz, ein Bild oder ein Objekt richtig oder auf bestimmte Weise falsch erinnert? – z. B. Einfluss eines Prime-Satzes auf das Erinnern der syntaktischen Struktur eines Target-Satzes (vgl. Griffin/Weinstein-Tull 2003).
- Als wie akzeptabel wird eine bestimmte grammatische Konstruktion von Versuchspersonen bewertet? (vgl. Mandell 1999).

Beispiele für Online-Messung

- Wieviel Zeit vergeht durchschnittlich zwischen dem Beginn der Reizpräsentation und der Reaktion per Knopfdruck? – z. B. lexikalische Entscheidung (vgl. Schreuder/Baayen 1997; Bertram/Baayen/Schreuder 2000).
- Wieviel Zeit vergeht durchschnittlich zwischen dem Beginn der Reizpräsentation (z. B. Zeigen eines Bildes) und einer sprachlichen Reaktion? (vgl. Schriefers et al. 1990: Objektbenennungsaufgabe, Griffin/Bock 2000: Bild-/Szenenbeschreibungsaufgabe).
- Wieviel Zeit vergeht zwischen dem Beginn der visuellen Reizaufnahme in einer bestimmten Region und der Verbalisierung der dort befindlichen Information? (vgl. Griffin/Bock 2000 – Messung der Eye-Voice-Span mit Eyetracking).
- Wann relativ zum akustischen Beginn (Onset) eines auditiv wahrgenommenen sprachlichen Reizes identifizieren Versuchspersonen ein korrespondierendes visuelles Objekt? (vgl. Altmann/Kamide 1999: Messung der Augenbewegung im Visual-World-Paradigma).
- Wie lange hält sich der Blick beim Lesen eines Satzes auf einem bestimmten Wort auf? (vgl. Staub 2010).
- Wie schnell erfolgt der Tastendruck, der beim Wort-für-Wort-Lesen das Verschwinden eines aktuellen und das Erscheinen eines neuen Wortes bewirkt? ›*self-paced reading*‹ (vgl. Traxler/Pickering/McElree 2002).

Validität, Objektivität, Reliabilität, Trennschärfe und Einfachheit: Ein Verfahren ist umso mehr **valid**, je genauer mit ihm das beobachtet/gemessen wird, was gemessen werden soll. Inwieweit das für ein gegebenes Verfahren relativ zu einer Beobachtungsabsicht der Fall ist, ist nicht immer so augenfällig wie beim Metermaß und der Längenmessung. Neben der theoretischen Prüfung werden deshalb vielfach empirische Prüfungen der Validität unternommen. Ein Standardverfahren ist das Vergleichen des zu prüfenden Verfahrens mit einem unabhängigen, validen Testverfahren. **Objektivität** und **Reliabilität** geben an, wie wahrscheinlich es ist, dass ein

Verfahren bei wiederholter Anwendung auf denselben Wirklichkeitsausschnitt (Reliabilität) bzw. durch andere Beobachter (Objektivität) zu denselben Ergebnissen kommt. Ein Verfahren ist umso besser in seiner **Trennschärfe** relativ zu einer Beobachtungsabsicht, je genauer es für einen beobachteten Wirklichkeitsausschnitt die Unterscheidungen trifft, die für die Untersuchung relevant sind. Ein plumpes Beispiel: Für die alltägliche Unterscheidung der Größe von Schuhen ist eine Messung nach Millimetern zu differenzierend, eine nach Dezimetern zu grob. **Einfachheit** ist nach vielen Hinsichten zu beurteilen; die gängigen Kriterien sind relativer Aufwand an Personal, Zeit, Geräten und Material pro Anwendung.

1.4.3 | Elizitationsmethoden

In den bisherigen Abschnitten wurde beschrieben, dass typische psycholinguistische Methoden auf der Beobachtung/Messung spezifischen sprachlichen oder sprachrelatierten Verhaltens unter kontrollierten Bedingungen beruhen. Doch wie bringt man Versuchspersonen dazu, genau das Verhalten an den Tag zu legen, das gemessen werden soll? Dies ist die Frage nach den Elizitationsmethoden. Von ihnen gibt es in der Psycholinguistik viele verschiedene. Aus Platzgründen können in diesem Abschnitt allerdings nur eine Handvoll vorgestellt werden. Insgesamt ist zu beachten, dass die gewählte Elizitationsmethode von der im Zentrum des Interesses stehenden theoretischen Forschungsfrage, der Operationalisierung und dem gewählten Messverfahren abhängig ist.

Benennen und Beschreiben

Für die Untersuchung der Prozesse, die im Zusammenhang mit dem Aufbau oder der Arbeitsweise des Sprachproduktionssystems stehen, werden Versuchsteilnehmer häufig gebeten, auf Bildern dargestellte Objekte zu benennen oder auf Bildern bzw. Videoclips dargestellte Szenen zu beschreiben. Hierbei spricht man von der Objektbenennung bzw. der Szenenbeschreibung. In einigen Studien erfolgt das Benennen oder Beschreiben auch in schriftlicher Form.

Lesen

Um Prozesse, die im Zusammenhang mit dem Sprachverstehen stehen, oder um die Art der Repräsentation spezifischen Wissens zu untersuchen, werden in vielen Experimenten Wörter, Sätze oder Texte präsentiert, die von Versuchspersonen gelesen werden sollen. Die Messung der Reaktion kann bei dieser Art der Elizitation ganz unterschiedlich ausfallen.

Hören von Lauten, Wörtern oder Sätzen

Das Sprachverstehen kann natürlich auch durch die Verarbeitung von auditivem Input untersucht werden, was in manchen Probandengruppen auch die einzig einsetzbare Methode darstellt (Kleinkinder, Sprecher einer Sprache ohne Schriftform). Dabei bedient man sich in der Psycholinguistik je nach Fragestellung unterschiedlich großer Einheiten. Während Studien, die sich mit der Verarbeitung der kleinsten bedeutungsunterscheiden-

Beispiele für Elizitationsmethoden

den Einheiten konzentrieren, primär mit der Fähigkeit zur Unterscheidung verschiedener Laute – teilweise von jeglichem Kontext isoliert – konzentrieren, fokussieren Studien, die sich mit dem Aufbau komplexer Bedeutungsstrukturen beschäftigen, entsprechend auf Sätze oder sogar kurze Texte. Oft wird das Hören bzw. Verstehen von Sprache mit dem Blickverhalten der Probanden in Verbindung gesetzt (Visual-World-Paradigma, vgl. Altmann/Kamide 1999).

Entscheiden

Die Verarbeitung von Sprache wird in vielen Experimenten auch durch Entscheidungsaufgaben elizitiert. Bei der sogenannten lexikalischen Entscheidung beispielsweise sollen Versuchspersonen so schnell wie möglich (meist durch das Drücken einer Taste) entscheiden, ob ein dargebotener Reiz ein Wort ihrer Sprache darstellt (Monat vs. Motan). Um diese Entscheidung treffen zu können, müssen Versuchspersonen auf ihr mentales Lexikon zugreifen. Dasselbe gilt für Aufgaben, bei denen Versuchspersonen so schnell wie möglich eine semantische Entscheidung treffen sollen, beispielsweise, ob ein präsentierte Wort ein belebtes Objekt bezeichnet (*Kamel* vs. *Karamell*).

Erinnern

Auch Gedächtnisaufgaben können benutzt werden, um die Verarbeitung von Sprache zu elizitieren. Dabei kann das Erinnern tatsächlich als Messung im Vordergrund stehen oder es dient gewissermaßen als Methode, um vom eigentlichen Ziel der Untersuchung abzulenken.

Priming

Das Priming-Paradigma ist ein sehr generelles Verfahrensmuster, mit Hilfe dessen, die Vorhersagbarkeit spezifischer Reaktionen auf einen Stimulus untersucht werden kann. Es beruht auf dem Funktionszusammenhang, dass die Durchführung einer kognitiven Prozedur erleichtert ist, wenn sie in einem bestimmten zeitlichen Abstand davor schon einmal oder mehrfach aktiviert war. Beispiel: Es wird eine einfache Handlung – auf Bildern dargeboten – beschrieben, z. B. *Eine Katze fängt eine Maus*. Wird vor der Bildpräsentation ein passivischer Satz zu lesen gegeben, der mit dem Inhalt des Stimulusbildes nicht verbunden ist (*Die Hecke wird von dem Gärtner gegossen*), so ist die Wahrscheinlichkeit, dass auch das Bild mit einem Passivsatz beschrieben wird, erhöht (vgl. das Experiment von Bock 1986).

Normierung und Auswahl der Probanden: Um zu gewährleisten, dass anhand der Messungen generalisierbare Aussagen getroffen werden können, werden in Experimenten häufig Messungen bei einer Versuchsperson (verschiedene Stimuli) bzw. bei einem Stimulus (verschiedene Versuchspersonen) wiederholt. Hierfür ist es natürlich notwendig, dass die zu messenden Reaktionen miteinander vergleichbar sind und deshalb über Einzelbeobachtungen hinweg generalisiert werden kann. Aus diesem Grund werden experimentelle Stimuli oft vor der eigentlichen Messung genormt. Soll zum Beispiel gemessen werden, wie schnell Sprecher die Bezeich-

nung eines Objekts aus einer bestimmten Kategorie aktivieren und produzieren können, muss vorher sichergestellt werden, dass möglichst alle Probanden dieselbe Bezeichnung für ein Objekt produzieren. Ähnliches gilt teilweise auch für die Auswahl der Kohorte (Probandengruppe). Um etwa Aussagen über die Fähigkeit von Sprachlernern, bestimmte sprachliche Strukturen auf bestimmte Weise zu verarbeiten, können natürlich nur dann aussagekräftig sein, wenn die Untersuchten einen vergleichbaren Spracherwerbsstand haben.

2 Sprachliches Wissen

2.1 Das mentale Lexikon

2.2 Strukturelles Wissen

2.3 Repräsentation sprachlichen Wissens

2.4 Die Lokalisationsfrage

Immer, wenn Menschen sprechen, schreiben, lesen oder hören, finden kognitive Prozesse statt, die nach Inhalt und zeitlichem Ablauf vom Inhalt und dem Aufbau der Sprachfähigkeit ermöglicht und beschränkt sind. Daher bildet die Kenntnis dieses sprachlichen Wissens eine Voraussetzung für die Erklärung des sprachlichen Verhaltens. Nachdrücklich wird noch einmal daran erinnert, dass mit dem Terminus ›Wissen‹ hier und im Weiteren auch das Wissen bezeichnet ist, das man sich normalerweise nicht begrifflich bewusstmacht und auch nicht bewusstmachen kann, also z. B. das uns unbewusste syntaktische Programm, dessen Anwendung, ebenso unbewusst, dazu führt, dass wir korrekte Sätze sprechen und schreiben. Eine luzide linguistische Analyse der Begriffe Sprachvermögen, Sprachsystem und Kommunikation bietet Klein (2005). Was im Folgenden **sprachliches Wissen** genannt ist, entspricht dort **Sprachsystem**, der Bezeichnung **Sprachverhalten** hier dort der Terminus **Kommunikation**.

Was ist sprachliches Wissen? Was besagt es nun, dass Inhalt und Aufbau des sprachlichen Wissens die Aktivität des Sprachverhaltens ermöglicht? Es besagt ganz lapidar, dass mittels dieses Wissensbestandes etwas möglich ist, was ohne ihn nicht möglich wäre, nämlich der sprachliche Ausdruck von theoretisch jedem beliebigen Gedanken, der uns im Leben durch den Kopf gehen mag. Ebenso ermöglicht das sprachliche Wissen das Verstehen von sprachlich ausgedrückten Gedanken der Menschen um uns herum. Ohne sprachliches Wissen hätte der Purser in der Lufthansamaschine (s. Beispiel 15.29:52 am Anfang von Kap. 1) seine Absicht nicht realisieren können, die Piloten zu informieren, dass in der Passagierkabine des Flugzeugs die Voraussetzungen für die Landung hergestellt sind, und die Piloten hätten diese Mitteilung ohne sprachliches Wissen nicht aufnehmen können.

Das sprachliche Wissen stellt dem Menschen die Mittel für die sprachliche Kommunikation über nicht-sprachliche psychische Dinge bereit, und die Organisation dieses Wissensbestandes prägt alle Modalitäten der sprachlichen Verständigung.

Zum Begriff

Wie für das Sprachsystem gilt für das sprachliche Wissen, dass es zwei Hauptteile umfasst: das Wissen sprachlicher Einheiten und Kombinationswissen. Letzteres kann man sich als Ensemble von Prinzipien denken, die grundsätzliche Einschränkungen für die Zusammensetzung von Ein-

Überblick

heiten festlegen. Die Kenntnis der sprachlichen Einheiten nennt man **das mentale Lexikon**, die Kenntnis der Kombinationsmöglichkeiten **die mentale Grammatik**.

Inhalt des mentalen Lexikons: Eine scheinbar so einfache Äußerung wie *Kabine ist klar* zu sprechen oder zu verstehen, beruht auf der Kenntnis von drei lexikalischen Einheiten und der raschen Kontrolle ihrer kombinatorischen Eigenschaften. Welche Eigenschaften sind das im Einzelnen?

Da sind zum einen die **gedanklichen Inhalte**, im Beispiel also die Bedeutungen *Kabine*, *sein* und *klar*, sowie, dass das Klar-Sein der Kabine in einer bestimmten Zeitspanne der Fall ist, nämlich zu der aktuellen Sprechzeit.

Zum zweiten wird eine **Äußerungsabsicht** zu verstehen gegeben, nämlich, dass es sich um eine Mitteilung handelt, nicht um eine Frage. Alle diese Eigenschaften sind begrifflicher Natur und werden als **konzeptuelle Bestandteile** bezeichnet, unterscheidbar in lexikalische Konzepte (*Kabine*, *klar* etc.) und indexikalische Konzepte (zeitliche und räumliche) sowie Intensionskonzepte.

Zum dritten sind da die **syntaktischen Eigenschaften** jedes Wortes, zusammen als **Lemma** bezeichnet: Wortklasse (hier: Nomen, Kopula, Adjektiv), Genus (hier: feminin), Anforderungen an die syntaktische Umgebung, *klar sein*: Subjekt notwendig) etc.

Zum vierten zeigt die Äußerung Kenntnis **flexionsmorphologischer Kategorien**, hier Person, Kasus, Numerus, Tempus, Modus, Determiniertheit, Komparation.

Schließlich belegt das Beispiel die Kenntnis der **lautlichen** und die Kenntnis der **schriftlichen** Ausdrucksseite der Lautsegmente, ihrer Gliederung zu Silben und – im Textbeispiel nicht ausgedrückt – der Platzierung des Akzents im Wort. Sie werden zusammengefasst als **Lexem** bezeichnet.

Inhalt der mentalen Grammatik: Zur mentalen Grammatik im weiteren Sinne zählen alle **universalen** und einzelsprachlichen **Prinzipien**, die festlegen, wie aus elementaren Einheiten komplexere zusammengesetzt sind, aus diesen wieder komplexere, aus diesen die Äußerung und aus Äußerungen letztlich die Rede, der Text, ein Gesprächsbeitrag. Die mentale Grammatik umfasst nach allgemeiner Annahme Prinzipien für die Kombination:

- von Konzepten zu komplexeren semantischen Gebilden, Bedeutungsstrukturen genannt,
- von syntaktischen Einheiten zu komplexen syntaktischen Gebilden wie Phrasen und Sätzen,
- von einfachen lexikalischen Einheiten zu komplexeren, also Wortformen und neuen Wörtern, und
- von lautlichen Einheiten zu komplexeren wie phonologischen Phrasen und der intonatorischen Einheit der Äußerung.

Linguistischer vs. psychologischer Zugang: Restriktionen hinsichtlich der Strukturbildung bestehen natürlich; wäre eine verletzt, würde die resultierende Äußerung nicht wohlgeformt sein, z. B. *Kabine sind klar*, *Kabine*

bist klar, Kabine will klar. Daraus folgt zwangsläufig, dass sie auch im sprachlichen Wissen enthalten sein müssen; es würden sonst eben abweichende Sätze produziert bzw. Abweichungen nicht erkannt werden. Der entscheidende Punkt ist aber, dass aus der linguistischen Erkenntnis der strukturellen Beziehungen im Satz nicht eindeutig hervorgeht, wie das sprachliche Wissen aufgebaut ist und wie mit ihm diese Beziehungen erzeugt bzw. verstanden werden. Kurz gesagt, aus Strukturkenntnissen kann nicht die mentale Grammatik des Menschen abgeleitet werden, ebenso wenig aus den linguistischen Beschreibungen der Lexik einer Sprache der Aufbau des mentalen Lexikons.

Das Spezifikum der psycholinguistischen Fragestellung lässt sich umso gründlicher verstehen, je eindringlicher man sich die Verschiedenheit und die Nachbarschaftsbeziehung der linguistischen und der psychologischen Perspektive bewusstmacht. Der linguistische Zugang erfolgt über die Analyse der Struktur sprachlicher Äußerungen. Der psychologische Zugang geht über die Analyse von Verarbeitungsvorgängen, den Zusammenhang zwischen sprachlicher Aufgabe und dem beobachteten Ergebnis, der Zeit, die dafür gebraucht wird, der Störbarkeit der Verarbeitung und der Logik der Fehlproduktionen. In den letzten Jahren werden zunehmend auch Ergebnisse von Messungen neurophysiologischer Vorgänge ausgenutzt. Da aus beiden Perspektiven dieselbe Sache gesucht wird, sollten die jeweiligen Behauptungen nicht zu entgegengesetzten Schlussfolgerungen führen.

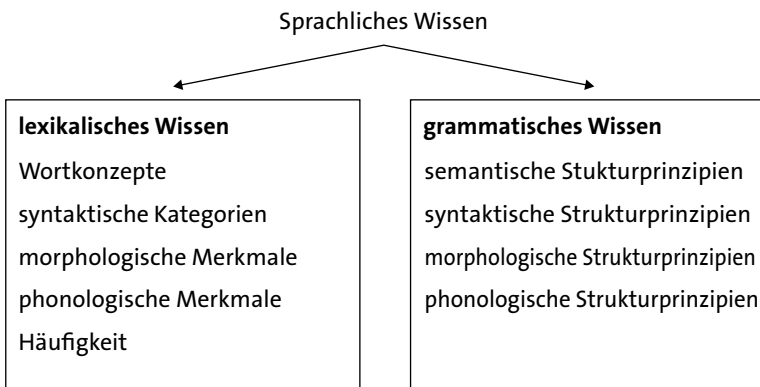


Abb. 2.1: Aus linguistischen Analysen ableitbare Bestandteile des sprachlichen Wissens.

2.1 | Das mentale Lexikon

Allen Versuchen, den Inhalt des mentalen Lexikons zu beschreiben, sind einige elementare Fragen gemeinsam. Die generellste ist wohl die, was eine lexikalische Einheit sei. Anders gefragt, gibt es überhaupt begründeten Anlass zu der Annahme, dass im Langzeitgedächtnis bestimmte Lautfolgen dauerhaft mit bestimmten Bedeutungen verknüpft sind? Im positiven Fall schließt sich natürlich die Frage nach dem Inhalt einer solchen Wissensseinheit an und an diese die Frage, wie die im Wortwissen enthaltenen Informationen gegliedert sind. Sofern sich das Lexikon als komple-

xer Informationsspeicher erweisen sollte, folgt natürlich die Frage, wie die einzelnen Informationen (lautliche, syntaktische etc.) ihrerseits aus psycholinguistischer Perspektive beschaffen sind.

Zum Begriff

In der Psycholinguistik versteht man unter **mentales Lexikon** einen sprachlichen Wissensbestand im Langzeitgedächtnis.

2.1.1 | Die lexikalische Einheit

Die kognitive Realität der lexikalischen Einheit

Evidenz

Untersuchungen darüber, ob das Langzeitgedächtnis tatsächlich sprachliche Einheiten enthält, liegt bei allen methodischen Unterschieden der einfache Gedanke zu Grunde, dass ein Informationsbündel, sofern es im geistigen Bestand eine Einheit bildet, auch als Einheit verarbeitet werden sollte, also anders als die darin enthaltenen Komponenten einzeln.

Wortüberlegenheitseffekt: Ein Prozess, der eine gewisse Kontrolle der zu verarbeitenden Einheit erlaubt, ist die Sprachrezeption; besonders das Lesen ist mit einfachen Mitteln leicht zu kontrollieren und zu beobachten. Der erste Schritt des Lesens ist die visuelle Wahrnehmung der Buchstaben, das heißt die kurzfristige Aufnahme ihrer visuellen Bilder in einen begrenzten Wahrnehmungsspeicher. Dieser Vorgang braucht Zeit und daraus folgt, dass die Aufnahme von einem einzelnen Buchstaben schneller abgeschlossen sein sollte als die Aufnahme von einer Folge von Buchstaben. Nun wurde aber beobachtet – übrigens schon im 19. Jahrhundert durch Cattell (1885) – dass unter Umständen die Verarbeitung eines einzelnen Buchstabens scheinbar länger dauern kann, als die Verarbeitung von mehreren, nämlich dann, wenn letztere zusammen ein Wort ausdrücken. Allerdings zeigt sich dieser Effekt nur bei einigen Wörtern, eher kürzeren, und nicht bei beliebigen, was den Einwand nahelegt, es handle sich gar nicht um einen allgemeinen Wortüberlegenheitseffekt, sondern nur um eine Eigenschaft spezieller Wörter.

Der Verdacht, dass die Beobachtung doch etwas mit der Worthaftigkeit zu tun hat, bedarf also der Überprüfung. Dazu kann man zunächst einmal die Lesegeschwindigkeit für Buchstabenfolgen, die Wörter bilden, mit der von Nichtwörtern vergleichen. Wieder zeigt sich, dass Ketten aus Wörtern schneller verarbeitet werden als gleich lange Nichtwortketten. Nun kann diese Wirkung natürlich auch anders als unter Bezug auf die Worthaftigkeit erklärt werden, z. B. mit der sprachspezifischen Wahrscheinlichkeit, mit der Buchstaben aufeinanderfolgen – in der jeweiligen Sprache natürlich. Sie könnte in wortbildenden Buchstabenketten höher sein, was beim Lesen den Übergang von einem Buchstaben zum nächsten beschleunigt. Der zu Grunde liegende Effekt wäre dann wieder ein Häufigkeitseffekt auf Buchstabenebene und nicht eine Wirkung der Worthaftigkeit. Zu einer vorläufigen Entscheidung dieser Alternativen führte ein Experiment von Reicher (1969).

Experimenteller Nachweis des Wortüberlegenheitseffekts

Experiment

Reicher (1969) machte sich eine Besonderheit der Arbeitsweise im Wahrnehmungsspeicher zu eigen, die rund zehn Jahre zuvor entdeckt worden ist, die der rückwirkenden Maskierung.

Der Effekt der rückwirkenden Maskierung ist auf die Art und Weise zurückzuführen wie die visuelle Wahrnehmung des Menschen zum Beispiel beim Lesen funktioniert. Hierbei wird ein Außenreiz für sehr kurze Zeit in einen Wahrnehmungsspeicher aufgenommen und unter Umständen weitergeleitet oder – unter anderen Umständen – einfach gelöscht, jedenfalls aber durch den folgenden Wahrnehmungsinhalt überschrieben.

Der Effekt tritt u. a. ein, wenn kurz hintereinander zunächst eine Buchstabenfolge, dann dieselbe Buchstabenfolge aber mit einem buchstaben großen Quadrat an einer Position gesehen werden. Ist dann die Entscheidung zu treffen, welcher Buchstabe an der maskierten Position der erstgesehenen Kette stand, kann diese Entscheidung aus der Erinnerung nicht mehr getroffen werden, wenn die Lesezeit entsprechend kurz war (100 ms (= Millisekunden)) oder weniger, wie man heute weiß. Sie kann aber noch getroffen werden, wenn die Löschung durch rückwirkende Maskierung erst nach der Weiterleitung zur lexikalischen Erkennung stattfindet. Um zu zeigen, dass die bessere Verarbeitung einer worthaften gegenüber einer nicht-worthaften Buchstabenfolge nicht auf einen reinen Häufigkeitseffekt, sondern eben auf einen Wortüberlegenheitseffekt zurückgeht, verwendete Reicher (1969) folgende Prozedur:

In schneller Folge und sehr kurz werden eine Buchstabenfolge, MAUS (= Wortbedingung) oder AMUS (Non-Wort-Bedingung) und sodann eine entsprechende maskierte Buchstabenfolge, z. B. MAU□ bzw. AMU□ gezeigt. Dann folgt die Aufgabe: Es wird der Buchstabe S oder L gezeigt und die Entscheidung verlangt, ob es der letzte Buchstabe der zuerst gesehenen Folge, also MAUS bzw. AMUS war.

Das Ergebnis ist, dass die Entscheidung unter der Wort-Bedingung signifikant häufiger korrekt ist, als unter der Non-Wort-Bedingung, obwohl beide Buchstaben die MAU□-Maske zu einem Wort ergänzen (Maus, Maul) und auf »u« im Deutschen gleich häufig folgen.

Dieses Ergebnis ist also nicht mit Übergangshäufigkeiten von Buchstaben zu erklären, sondern dadurch, dass ein Wort als Buchstabenfolge entschieden schneller vom Wahrnehmungsspeicher verarbeitet wird, als eine nicht-worthaftige Kette von Buchstaben. Zwar sind in beiden Fällen die rechten Buchstaben durch die Maske im Wahrnehmungsspeicher gelöscht worden; das Wissen über den rechten Rand des Wortes ist aber korrekter als das über den rechten Rand des Nichtwortes, weil ersteres schneller verarbeitet worden ist. Die Erklärung des Wortüberlegenheitseffekts wird nun darin gesehen, dass bei Wahrnehmung eines Wortes zwangsläufig und äußerst schnell lexikalische Verarbeitung unter Einbezug des lexikalischen Wissens, z. B. der Bedeutungsvorstellung stattfindet.

An der Aufklärung des Wortüberlegenheitseffekts wird weiter gearbeitet, u. a. an seiner Präsenz bei lautlicher Wahrnehmung (vgl. Krueger 1992;

allgemein Hildebrandt/Caplan/Sokol/Torreano 1995). Im Zuge dieser Bemühungen ist eine andere Erklärung in den Blick gerückt. Sie geht von der Frage aus, ob die Transparenz der Schrift einer Sprache bei dem Effekt eine Rolle spielt. So ist z. B. die Schrift im Englischen der lautlichen Form vielfach weniger nah als das im Italienischen der Fall ist; ein immer gerne zitiertes Beispiel ist das englische Wort für *lachen*, *laugh* [la:f]. Ripamonti/Traficante/Crippa/Luzzatti (2014) sind dieser Vermutung nachgegangen.

Experiment Wortüberlegenheitseffekt Revisited

In einer Studie von Ripamonti et al. (2014) wurden Probanden Wörter, Non-Wörter und sog. Pseudo-Wörter jeweils sehr kurz visuell präsentiert. Ein Pseudo-Wort ist eine Buchstabenfolge, die nach den Regeln in der jeweiligen Sprache ein Wort darstellen könnte, aber nicht im mentalen Lexikon enthalten ist. Ein Non-Wort ist eines, das ebenfalls nicht im mentalen Lexikon vorhanden ist und dessen Buchstabenfolge zusätzlich kein Wort der jeweiligen Sprache sein kann. Im Deutschen ist z. B. h-a-u-s ein Wort, h-a-u-m ein Pseudowort und h-s-u-a ein Non-Wort.

In einem Durchgang wurden die Stimuli mit dem linken Auge, im zweiten Durchgang mit dem rechten Auge gelesen. Die Aufgabe war dieselbe, wie in dem Experiment von Reicher (1969). Es zeigte sich, dass Non-Wörter schlechter erinnert werden als Wörter und Pseudowörter, dass aber kein signifikanter Unterschied in der Verarbeitung von Wort und Pseudowort bestand – im Italienischen, wie gesagt.

Die Autoren schließen daraus, dass an dem Lesevorgang auch schon prälexikalische Prozesse entscheidend beteiligt sind. Ebenfalls für diese Erklärung spricht die Beobachtung, dass die Reaktionen keinen Unterschied zwischen rechtem und linkem Auge aufweisen, was für die Annahme spricht, dass die relevanten Prozesse links- wie rechtshemisphärisch ablaufen können, mithin kein Zugriff auf das mentale Lexikon erfolgt.

Stroop-Effekt: Für die Annahme, dass die Informationen, die das lexikalische Wissen ausmachen, im mentalen Lexikon zu einer Einheit integriert sind, wird häufig auch eine andere Beobachtung angeführt. Sie heißt nach ihrem Entdecker J. Ridley Stroop (1935) der Stroop-Effekt. Dieser Effekt besteht im Verhältnis mehrerer Reaktionsmuster im Rahmen verschiedener zusammengehöriger Benenn- bzw. Leseaufgaben. Er verdeutlicht besonders anschaulich, dass Buchstabenfolgen, die ein Wort ergeben automatisch als Wort verarbeitet werden und diese automatische Verarbeitung kaum aktiv unterdrückt werden kann.

Experiment Stroop-Effekt

›Stroop-Effekt‹ ist die zusammenfassende Bezeichnung für mehrere Reaktionsmuster bei den Aufgaben:

1. die Farbe einer Fläche zu nennen,
2. ein Farbwort in übereinstimmender Schriftfarbe zu lesen,

3. ein Farbwort in kontrastierender Schriftfarbe zu lesen und
 4. die Schriftfarbe einer kontrastierenden Farbbezeichnung zu nennen.
 Stroops Interesse galt vor allem dem Vergleich der Bedingungen 2 und 3 sowie dem Vergleich der Bedingungen 1 und 4. Der erste Vergleich sollte den interferierenden Effekt der Farbe auf das Lesen eines Wortes demonstrieren. Der zweite Vergleich sollte den Interferenzeffekt, der durch die Wortbedeutung bedingt ist auf das Benennen der Farbe zeigen. Der Vergleich zwischen den Bedingungen 2 und 3 zeigte keinen reliablen Unterschied. In der Originalstudie ergab sich nach dem Lesen von 100 Farbwörtern jeweils in beiden Bedingungen lediglich ein durchschnittlicher Unterschied von 2,3 Sekunden. Der Vergleich zwischen den Bedingungen 1 und 4 hingegen war enorm. Die Benennung von 100 farbigen Flächen dauerte durchschnittlich 63 Sekunden. Das Benennen der Schriftfarbe von 100 (kontrastierenden) Farbwörtern dauerte hingegen durchschnittlich 110 Sekunden.

Materialbeispiele und Angaben zur Experimentdurchführung enthält die Abbildung, die diesem Buch als loses Blatt beigelegt und auf der Webseite des Verlags im Internet zugänglich ist (ausführlichere Angaben enthält Miller 1991).

Die Erklärung für den Stroop-Effekt wird in einer Interferenz zwischen Wortbedeutung und Farbvorstellung gesehen. Die sprachliche Einheit des Wortes zwingt das Denken, sich mit seiner Verarbeitung zu beschäftigen und beeinträchtigt die Verarbeitung der Farbwahrnehmung. Die Theoriediskussion ist in Roelofs (2005) erörtert. Bemerkenswert sind allerdings Befunde zur neuronalen Lokation von Stroop-Mechanismen, die von Song/Hakoda (2015) berichtet werden. Die Daten weisen eher auf die sprachfernen Brodman-Areale 9, 10, 6 und 8 hin und zwar in beiden Hemisphären.

Lauterkennung im Wortkontext und Phonemrestauration: Auch die Verarbeitung von Lauten wird davon beeinflusst, ob der Laut als Bestandteil eines Wortes gehört wird oder nicht. Unter anderem wird die Fähigkeit davon beeinflusst, Urteile über auditiv Wahrgenommenes zu fällen. Wird in einem gehörten Wort ein Laut einmal mit Geräusch überlagert, ein andermal durch Geräusch vollständig ersetzt, so wird letzteres schlechter erkannt. Dieser Effekt tritt nicht auf beim Hören gleichartig unterschiedener Nicht-Wort-Paare (vgl. Samuel 1986).

Der Effekt der sog. Phonemrestauration (*phonemic restoration effect*) liefert Evidenz für dieselbe Annahme (vgl. Warren 1970; Sivonen/Maess/Lattner/Friederici 2006; Kashino 2006). Wird ein Laut eines Wortes vollständig von einem nicht-sprachlichen Laut überlagert, fällt dies Hörern bei der Wahrnehmung nicht notwendigerweise auf. Versuchspersonen geben sogar an, den Laut deutlich gehört zu haben. Je nach Kontext, ersetzen Hörer einen fehlenden Laut sogar in verschiedener Weise (s. Kapitel 5 für eine detaillierte Erläuterung).

Wortfrequenz: Auch der Frequenzeffekt stützt die Annahme von der kognitiven Realität des Wortes. Ein Wort, das in der täglichen Kommunikation zwischen Mitgliedern der Sprachgemeinschaft häufig verwendet wird,

wird vom Einzelnen schneller verarbeitet. Die Frequenz ist also eine Eigenschaft des Wortes und sie lässt sich nicht aus der Frequenz seiner Laute ableiten. Der Frequenzeffekt zählt mit zu den stabilsten Phänomenen lexikalischer Verarbeitung (vgl. Jescheniak/Levelt 1994; Jescheniak 2002; Graf/Nagler/Jacobs 2005; jedoch auch Jescheniak/Meyer/Levelt 2003b).

2.1.2 | Die Struktur der lexikalischen Einheit

Die sprachwissenschaftliche Analyse des Wortes lässt in erster Näherung eine Unterscheidung in folgende Bereiche erwarten:

- den Bereich der linear organisierten Ausdrucksform, also lautliches und graphematisches Wissen
- den Bereich der grammatischen Spezifikation, also morphosyntaktisches Wissen
- den Bereich der semantischen Information

Da kompetente Sprecher in der Lage sind, Wörter in Äußerungen in allen Hinsichten korrekt zu verwenden und Verletzungen aller Arten von lexikalischen Informationen zu bemerken, ist davon auszugehen, dass alle diese Informationen für jede lexikalische Einheit im mentalen Lexikon enthalten sind.

Vertikale Gliederung der lexikalischen Einheit: Da die Bedeutungsinformation mit der morphosyntaktischen und diese mit der phonologischen Information vernetzt sein muss, um die Zusammengehörigkeit dieser drei Informationskategorien zu sichern, ergibt sich für jede lexikalische Einheit eine Gliederung in drei Ebenen von lexikalischen Informationen:

Drei Ebenen von
lexikalischen
Informationen

- der **Konzept-Ebene** mit Informationen zur Bedeutung einer lexikalischen Einheit
- der **Lemma-Ebene** mit morphologischen und syntaktischen Informationen
- der **Lexem-Ebene** mit Informationen zu lautlichen Segmenten sowie metrischen Spezifikationen der lexikalischen Einheit

Diese Gliederung in Verarbeitungsebenen zeigt sich unter anderem in der zeitlichen Reihenfolge bei der Aktivierung eines Wortes beim Sprechen bzw. Verstehen (s. Kap. 4 und 5), bei Sprachstörungen auch darin, dass Informationen einer Ebene eines Wortes isoliert gestört sein können (s. Kap. 6). Man bezeichnet sie in Anlehnung an die raumzeitliche Metaphorik als die vertikale Gliederung der lexikalischen Einheit. Da alle lexikalischen Einheiten ebenfalls miteinander verbunden sind, resultiert aus der vertikalen Gliederung der Wörter eine – wiederum bildlich gesprochen – **horizontale Gliederung des gesamten mentalen Lexikons**.

Zum Begriff

Die **lexikalische Einheit** ist ein Eintrag im mentalen Lexikon, der aus einem Lemma und einem korrespondierenden lexikalischen Konzept sowie seinen Morphemen (eines oder mehrere) mit segmentalen und metrischen Informationen besteht.

Die **psycholinguistische Standardsicht** ist, dass der Lemma-Knoten der lexikalischen Einheit die Ausdrucksseite, das Lexem, mit den syntaktischen Informationen des Wortes verbindet, also solchen über Wortklasse, Genus, syntaktische Ergänzungen und andere. Nennen wir diese Verbindung unter Bezug auf das Bild von der horizontalen Gliederung die Verbindung nach unten. Eine weitere Verbindung geht vom Lemma nach oben zum Konzept, der semantischen Information, die je nach Theorie einen Bestandteil des Lemmas bildet oder eine außersprachliche Ebene von ihrerseits wieder horizontal verbundenen konzeptuellen Einheiten (s. Kap. 4). Alle Modelle, die diese Gliederung annehmen, nennt Caramazza (1997) **Syntactic Mediation-Modelle**, kurz SM-Modelle. Man beachte allerdings, dass Caramazza selbst die Notwendigkeit eines Lemma-Knotens bestreitet.

Empirische Evidenz für die vertikale Gliederung der lexikalischen Einheit stammen vor allem aus der Sprachproduktionsforschung. Zum einen wird hier das sogenannte *tip-of-the-tongue* (TOT)-Phänomen angeführt. Es bezeichnet eine spezifische Blockade beim Sprechen. Dabei wird die Bedeutung eines Wortes zwar aktiviert, aber es kann auf Informationen zur lautlichen Form nicht vollständig zugegriffen werden (vgl. Brown/Mc Neill 1966). Des Weiteren verdeutlichen Analysen von Versprechern (*Slips of the tongue*), dass im Sprachproduktionsprozess zunächst nur auf Be-

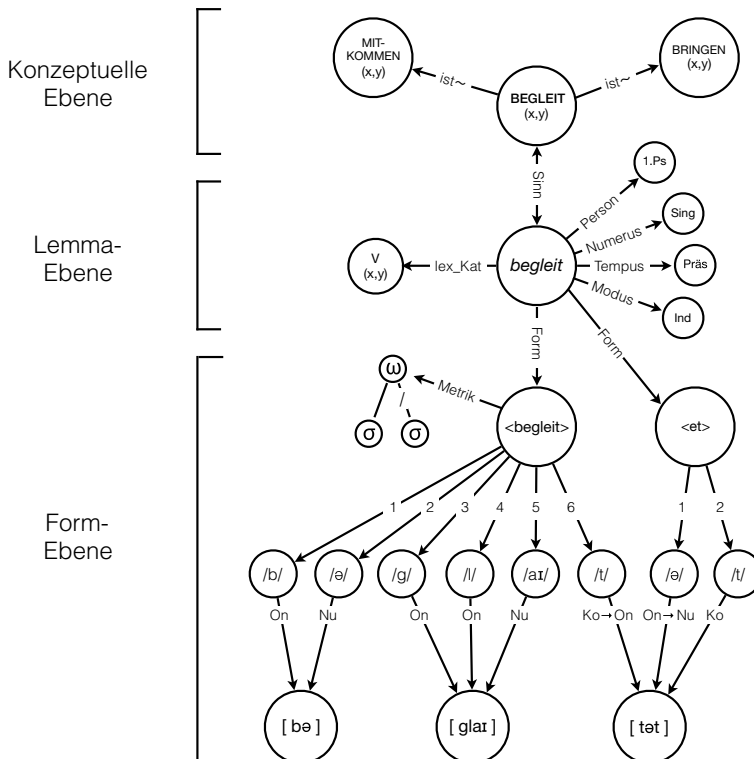


Abb. 2.2: Schematische Darstellung der Gliederung der lexikalischen Informationen des Verbs *begleitet* im mentalen Lexikon

deutungsinformationen und Informationen zur Syntax einer lexikalischen Einheit zugegriffen wird und erst in einem späteren Schritt auf die Spezifikationen zur Form (vgl. Garrett 1980). Des Weiteren lässt sich in Experimenten, in denen Versuchspersonen Objekte benennen sollen, der Abruf der semantischen und der phonologischen Information des Wortes, das für die Benennung verwendet wird, zu unterschiedlichen Zeitpunkten stören – semantische Information früh, phonologische Information spät (vgl. Schriefers et al. 1990). Letztlich kann ein früher Zugriff auf semantische und ein später Zugriff auf phonologische Informationen auch mit Hilfe neurowissenschaftlicher Methoden nachgewiesen werden (s. Kap. 4).

Zusammenhang zwischen phonologischer und graphematischer Repräsentation: Mit dem Stichwort ›lautlich‹ verbindet sich unmittelbar die Frage nach dem Ort der graphematischen Information im mentalen Lexikon. Für die Tatsache, dass Wissen über lautliche und graphematische Ausdrucksform prinzipiell voneinander getrennt gespeichert sein können, spricht auf den ersten Blick die Beobachtung, dass Hören und Sprechen ohne Lese- und Schreibfähigkeit uneingeschränkt möglich sind. Das besagt allerdings nicht viel über den Aufbau des lexikalischen Wissens bei Menschen, die sowohl die schriftliche als auch die mündliche Modalität beherrschen.

Direkte und indirekte Verbindung zwischen graphematischer und Lemma-Information: Anknüpfend an die Lemma-Lexem-Gliederung sind grundsätzlich zwei Wege der Verknüpfung von Graphem-Ebene (Schrift) und Lemma denkbar, eine indirekte vermittelt durch die phonologische Form des Lexems und eine direkte, zwischen Graphem- und Lemmaebene.

Da jeder, der lesen kann, auch ein visuell gegebenes Non-Wort aussprechen kann, muss es neben der direkten Beziehung auch die indirekte geben, denn Non-Wörter sind ja nicht als Lemmata im mentalen Lexikon vorhanden. Zusätzlich muss auch die Fähigkeit vorhanden sein, die eine Repräsentation in die jeweils andere regelgeleitet zu übertragen, ein Vorgang, der in der Leseforschung **Assembling** heißt. Auf die Frage nach der Natur der Verknüpfungen gibt es keine einfache Antwort, und verschiedene Beobachtungen zur Klärung der vermeintlich gleichen Frage deuten nicht alle in die selbe Richtung. Ganz abgesehen davon, dass die Art der Schrift (Buchstaben, Ideogramme, Silbenschrift) natürlich entscheidenden Einfluss hat. Interessant sind hierzu unter anderem Befunde von Zhu/Shu/Bi/Shi (1999), die eine Beteiligung der phonologischen Information beim Lesen und Worterkennen auch im Chinesischen nachweisen.

Dass es eine direkte Verknüpfung zwischen graphematischer und Lemma-Information gibt, wird daraus geschlossen, dass semantisches **Priming** das Erkennen einer Buchstabenkette als Wort beim Lesen beschleunigt. Bei semantischem Priming wird kurz vor dem Erscheinen des Zielwortes ein bedeutungsähnliches Wort gelesen (**unimodales Priming**). Auch lautliches Priming mit einem in diesem Fall gehörten Prime-Wort, das dem zu verarbeitenden, gelesenen Zielwort lautlich ähnlich ist (**cross-modales Priming**), beschleunigt und verbessert die Worterkennung, was dafür spricht, dass auch eine Verbindung zwischen lautlicher und graphe-

matischer Information auf der Ausdrucksebene besteht. Diese Verbindung wird nach Befunden von Seidenberg/McClelland (1989) beim Lesen generell aktiviert, was die Kernaussage ihrer *Parallel distributed processing-Theorie* (vgl. Seidenberg/McClelland 1989) darstellt, kritisch diskutiert von Besner (1990). Evidenz zugunsten dieses Modells liefern auch Beobachtungen des Verhaltens von Menschen mit Sprachkrankheiten oder Sprachstörungen, hier speziell mit verschiedenen Arten von Dyslexie (vgl. dazu Seidenberg 1995 und die Darstellung über Sprachstörungen in Kapitel 6).

2.1.3 | Beschaffenheit der lexikalischen Information

Informationen zur Form

Einem Vorschlag über den Gesamtablauf des lexikalischen Zugriffs von Levelt et al. (1999) folgend umfasst der Zugriff auf Forminformationen einer lexikalischen Einheit, also dem Abruf von Wissen, das im mentalen Lexikon gespeichert ist, die Aktivierung von drei Informationstypen:

- Informationen über den morphologischen Aufbau eines Wortes,
- Informationen über die metrische Form und
- Informationen über lautliche Segmente

Informationen über den morphologischen Aufbau von Wörtern: Wörter bestehen aus Stammmorphemen und Morphemen, mit deren Hilfe bestimmte Funktionen angezeigt werden, so z. B. Kasus, Genus und Numerus oder Tempus, Modus oder Aspekt. Das Wort *begleitet* etwa besteht aus einem Morphem <begleit> und einem Funktionsmorphem <et>. Dass genau diese beiden Morpheme aktiviert werden müssen, um eine ganz spezifische Bedeutung auszudrücken, ist im Lexikon-eintrag von <BEGLEIT> gespeichert. Die Aktivierung dieser beiden Morpheme liefert die Grundlage für die Auswahl der metrischen und segmentalen Informationen, die es ermöglichen, diese beiden Morpheme als das wahrnehmbare Ergebnis des Sprachproduktionsprozesses auszusprechen.

Informationen über lautliche Segmente: Für jedes Morphem, aus dem sich ein Wort zusammensetzt ist im mentalen Lexikon angegeben aus welchen phonemischen Segmenten es sich zusammensetzt und in welcher Reihenfolge diese Segmente zusammen das entsprechende Wort ergeben. In der Theorie von Levelt et al. (1999) wird diese Information zusammengefasst als **phonologischer Code** bezeichnet.

Informationen über die metrische Form: In vielen Sprachen lassen sich mehrsilbige Wörter auf der lautlichen Ebene dadurch charakterisieren, wo der Wortakzent liegt. Für Wörter, die vom Standardakzentmuster in einer bestimmten Sprache abweichen, ist als Teil der lautlichen Information, also als Teil des phonologischen Codes, auch gespeichert, wo der Wortakzent liegt. Das Wort *Kaffee* folgt dem kanonischen Format, wohingegen das Wort *Café* vom Standard abweicht. Für letzteres ist demnach als Teil der lexikalischen Information angegeben, wo der Wortakzent gesetzt werden muss.

Informationen über Silben: Man beachte, dass im Modell von Levelt et al. (1999) Informationen über die Silben, aus denen ein Wort zusammengesetzt ist, nicht eigens als Teil eines lexikalischen Eintrags gespeichert sind. Die Silbenstruktur für Wörter wird vielmehr durch einen darauf spezialisierten Prozess ›on-the-fly‹ generiert. Andere Modelle, etwa das von Dell (1988; 1986) hingegen nehmen an, dass als Teil der lautlichen Information im Lexikon phonemische Segmente bereits für Positionen innerhalb einer Silbenstruktur (Onset, Nukleus, Coda) spezifiziert ist.

Sprechsilbenlexikon: Des Weiteren wird im Rahmen der Sprachproduktionstheorie von Levelt et al. (1999) angenommen, dass es einen speziellen Wissensspeicher gibt, in dem sogenannte **Sprechsilben** gespeichert sind. Informationen über Sprechsilben sind also **unabhängig von spezifischen lexikalischen Einträgen** gespeichert, so die Theorie. Sprechsilben kann man sich als »Programme« vorstellen, in denen gespeichert ist, auf welche Art und Weise die an der Artikulation beteiligten Muskelgruppen zu bewegen sind, um eine bestimmte silbengroße Lautkette zu erzeugen. Man spricht in diesem Zusammenhang von artikulatorischen Gesten. Der Theorie entsprechend gibt es allerdings nicht für ausnahmslos jede mögliche silbengroße Lautkette einen Eintrag im Silbenlexikon, sondern nur für die in einer Sprache am häufigsten verwendeten. Für alle anderen werden die artikulatorischen Gesten, wenn sie benötigt werden, von darauf spezialisierten Prozessen generiert, also nicht aus dem Silbenlexikon abgerufen.

Syntaktische Informationen

Die syntaktische Information im mentalen Lexikon muss der Tatsache Rechnung tragen, dass Menschen die Fähigkeit zur Produktion syntaktisch wohlgeformter und zur Erkennung syntaktisch abweichender Sätze verfügen. In erster Näherung kann davon ausgegangen werden, dass im lexikalischen Wissen all die syntaktische Information anzunehmen ist, die sich nicht regulär aus lautlicher Information ableiten lässt.

Relevanz syntaktischer Informationen auf der Wortebene: Nicht aus lautlicher Information abzuleiten sind die Wortarteigenschaften des Wortes, ob es sich also im Satz verhält wie ein Verb, ein Adjektiv, eine Modalpartikel, ein Adverb, eine Präposition, ein Nomen, ein Eigennamen oder gleich wie ein ganzer Satz, wie z. B. *ja* oder *ach*. Man spricht in diesem Zusammenhang von Informationen zur **Wortkategorie**.

Ebenfalls auf der Wortebene relevant sind Informationen, die einen direkten Einfluss auf die Oberflächenform eines Wortes haben. Hierzu zählen etwa Genus, Numerus und Kasus, also Informationen, die sich in einem spezifischen Deklinationsschema äußern. Aber natürlich auch Tempus, Aspekt, Modus, in einigen Sprachen Evidenzialität, usw. In flektierenden Sprachen werden diese und weitere morphosyntaktischen Kategorien durch die jeweilige flektierte Wortform ausgedrückt. Die folgenden Beispiele illustrieren einige dieser sog. morphosyntaktischen Kennzeichnungen.

- (2–1) *Es war, als hätte der Himmel die Erde still geküsst.*
- (2–2) *Und in den Nächten fällt die schwere Erde aus allen Himmeln in die Ewigkeit.*

Im Vergleich dazu:

- (2–3) *Es ist, als hätte den Himmel die Erde still geküsst.*
 (2–4) *Und in der Nacht fiel aus dem Himmel zur Erde hin ein Licht.*

Damit die Wortformen in diesen Sätzen korrekt gebildet werden können, müssen die syntaktisch relevanten morphologischen Besonderheiten im mentalen Lexikon enthalten sein und zwar bei jedem Wort. Wörter unterscheiden sich also voneinander nicht nur hinsichtlich ihrer lexikalischen Kategoriezugehörigkeit wie Nomen, Verb, Adjektiv, Präposition, Konjunktion. Bei verschiedenen Wörtern derselben Kategorie können Unterschiede in den Formen bestehen, mit denen die oben genannten Flexionskategorien ausgedrückt werden. Die Notwendigkeit dieser lexikalischen Informationen zeigt sich in *der Himmel* aber *die Erde* (Genus), *Nacht* – *Nächten* (Dativ Singular – Dativ Plural) aber *Himmel* – *Himmeln*, *fallen* – *fällt* aber *küssen* – *küsst* (3. Pers., Sing., Präs., Indikativ, Aktiv) usw.

Subkategorisierung: Auch wie sich ein gegebenes Wort mit anderen zu einer Phrase verbinden lässt, ist bekanntlich nicht für alle Wörter gleich, wenngleich strukturell recht einheitlich. Kombinationsmöglichkeiten sowie -restriktionen müssen für jede lexikalische Einheit als syntaxrelevantes Wissen im Lexikon enthalten sein. In gewisser Weise prototypisch sind Informationen über die Subkategorisierung bei Verben. Ein Verb subkategorisiert eine bestimmte Anzahl sowie eine bestimmte Art von Argumenten. Dass ein Sprecher des Deutschen die Strukturen der Sätze (2–5) a.–c. ganz verschieden versteht, liegt an den Subkategorisierungseigenschaften der darin auftauchenden Verben. Da die Unterschiede auf die Verben zurückzuführen sind, die Information also nicht an der Wortart erkennbar ist, muss sie eigens als lexikalisches Wissen gespeichert sein.

Argumentstruktur

- (2–5) *Der Finanzminister...*
 a. ... *verspricht den Wählern, die Ausgaben zu senken*
 b. ... *scheint den Wählern die Ausgaben zu senken*
 c. ... *befiehlt den Wählern, die Ausgaben zu senken*

Aber auch für lexikalische Einheiten, die zu anderen Wortarten gehören können/müssen derartige Informationen zur Verfügung stehen. Beispielsweise hat auch eine Präposition ein Argument und die entsprechende Subkategorisierung muss in ihrem Lemma gegeben sein.

Spezifische **Evidenz für die Existenz syntaktischen Wissens im mentalen Lexikon** liefern Versprecher, neurobiologische Phänomene beim Satzverstehen und Äußerungen von Patienten mit syntaktischen Sprachstörungen. Genauer hierzu findet sich in den Kapiteln 4, 5 und 6.

Semantische Information

Grad der Spezifität semantischer Informationen: Eine der wohl größten Herausforderungen in der Sprachwissenschaft besteht in der Beschreibung von Bedeutung. Einer der Hauptgründe für die Schwierigkeiten, auf die man hier stößt, liegt darin, dass längst nicht klar ist, wo – salopp ge-

Bedeutung

sagt – Bedeutung aufhört und wo Interpretation oder Assoziation beginnt. Allein der Einfluss des Kontexts und der Intonation erlauben es, ein und denselben Satz auf vielerlei Weise zu interpretieren. Es ist – zur Illustration – wahrscheinlich ohne Weiteres möglich, für den Satz *Der Typ geht nicht* innerhalb weniger Minuten auf bis zu 10 oder mehr verschiedene Interpretation zu kommen. Dass dies möglich ist, scheint unmittelbar damit zusammenzuhängen wieviel bzw. wie wenig Bedeutung in den einzelnen Wörtern des Satzes enthalten ist, worin also die semantischen Informationen der lexikalischen Einheiten bestehen.

Semantik als Interface: Klar ist, es gibt nicht nur Wörter für Gegenstände, sondern auch für nicht gegenständliche Konzepte wie *Trauer*, *Recht*, *Hass*, *Primzahl* und für Relationen wie *Wurzel aus* oder *nah* und *fern*. Deren Bedeutung und zugleich den Wörtern mit gegenständlichen Bedeutungen wird die Annahme gerecht, Wörter seien Zeichen und deren Bedeutungen bestünden in ihrer Referenz auf **Begriffe**.

Zum Begriff

Begriffe sind eine spezifische Art von Konzepten. Ein Begriff ist Bestandteil eines Begriffssystems und sein Inhalt ist das Netz seiner Beziehungen zu den anderen Knoten des Systems.

Format

Aus dieser Sichtweise dient die semantische Information von lexikalischen Einheiten als Interface zwischen der sprachlichen und nicht-sprachlichen Repräsentation von Wissen.

Dekompositionelle Repräsentation: Die Bedeutung eines Wortes, also die Information, die es erlaubt, auf einen Begriff zu referieren, könnte theoretisch mental als gegliedertes Ensemble von Bedeutungskomponenten repräsentiert sein. Diese Annahme heißt in der Psycholinguistik die **Dekompositionsannahme** und sie weist deutlichen Bezug zur Komponentensemantik auf. Der Inhalt von *sterben* wäre z. B. repräsentiert durch *Wechsel: Leben(x) – tot sein (x)*, der Inhalt von *töten* würde den Inhalt von *sterben* enthalten und zusätzlich etwa *MACHEN (y, p)* und *VERURSA-CHEN ($p, sterben (x)$*). Evidenz für die Dekompositionsannahme kommt zum Beispiel aus einer Studie von Gennari/Poeppl (2003). Mit Hilfe von Messungen im Rahmen einer lexikalischen Entscheidungsaufgabe sowie durch Messungen der Lesezeit evaluierten die Autoren die Hypothese, dass semantisch komplexere Verben (solche, mit denen auf Ereignisse referiert werden kann) mehr Verarbeitungszeit benötigen als semantisch weniger komplexe Verben (solche, mit denen auf Zustände referiert werden kann). Die Ergebnisse aus den Experimenten deuteten auf die Richtigkeit dieser Hypothese. Zu ähnlichen Befunden kommt auch Gerwien (2011).

Holistische Repräsentation: Der Dekompositionsannahme kritisch gegenüber steht die Annahme, dass der Wortinhalt nicht komponentiell repräsentiert ist und somit durch eine Definition nicht angemessen modelliert ist. Vielmehr sei der Wortinhalt eines jeden Wortes elementar und die Verschiedenheit des Inhalts eines Wortes von dem eines anderen sei lediglich ein Unterschied in dem Muster von Relationen des einen Wortes zu

seinen Nachbarn gegenüber dem Verknüpfungsmuster des anderen Wortes. Dies kann als **Holismusannahme** bezeichnet werden. Auch diese Konzeption ist auf eine Semantiktheorie bezogen; sie weist Ähnlichkeit zu Grundannahmen der formalen Semantik auf, die letztlich auf Richard Montague (1970) zurückgeht.

Semantische Netze: Hand in Hand mit der Holismusannahme geht die Vorstellung, dass sich eine spezifische Wortbedeutung aus dem Verknüpfungsmuster mit der Bedeutung anderer Wörter ergibt. Diese Vorstellung wird allgemein **Semantische-Netze-Theorie** genannt (vgl. Aitchison 1997, S. 93 f.; Helbig 2013). Eine psycholinguistische Adaption findet sich z. B. in den Grundannahmen des Wortproduktionsmodells von Levelt et al. (1999). Hier argumentieren die Autoren wie folgt: Eine Herausforderung an eine Theorie der Bedeutungsrepräsentation im mentalen Lexikon besteht darin, zu erklären, wieso beim Selektieren eines Wortes in der Sprachproduktion genau das spezifische Wort gefunden wird, z. B. *Apfel*, wenn das Bild eines Apfels gezeigt, und nicht ein ebenso zutreffendes Hyperonym, z. B. *Obst*. Diese Frage stellt sich deshalb, weil die Repräsentation von Apfel, wäre sie komponentiell, alle Komponenten von Obst enthalten müsste (und darüber hinaus weitere spezifischere). Levelt (1989) bezeichnet dieses Problem als das **Hyperonym-Problem** des lexikalischen Zugriffs in der Produktion.

Im **Zwei-Ebenen-Modell des Bedeutungswissens** wird zwei Beobachtungen Rechnung getragen. Zum einen unterscheiden sich die Sprachen in der semantischen Gliederung der Welt im Lexikon. Bedeutungen, für die es in einer Sprache ein einfaches Wort gibt, werden in einer anderen Sprache durch einen zusammengesetzten Ausdruck wiedergegeben, z. B. *cotton* – *Baumwolle*. Mithin muss zwischen den universalen Konzepten und den sprachspezifischen Wortbedeutungen unterschieden werden. Zum zweiten erfährt ein Wort durch seine Einbettung in einen Satz eine spezifische Interpretation. So wird *verlassen* in den beiden Sätzen *Hans hat das Institut vor fünf Minuten verlassen* und *Hans hat das Institut vor drei Jahren verlassen* Verschiedenes verstanden. Und das geschieht nicht zufällig. Also muss noch eine Verarbeitung stattfinden, in die die lexikalische Bedeutung eines Wortes eingeht, ferner relevante Teile von Bedeutungseinheiten im aktuellen Kontext mit dem Ergebnis der aktuellen begrifflichen Interpretation, ausgedrückt in einer Konfiguration von Konzepten.

Eine komplexere Modellierung dieser Vorstellung ist in Bierwisch/Schreuder (1991) vorgeschlagen. Sie trägt der Beobachtung Rechnung, dass ein und dasselbe Wort benutzt werden kann, um Entitäten sehr unterschiedlicher Natur zu bezeichnen. Die Feststellung *Das ist kein guter Weg* kann auf ein Stück unwegsamen Pfades bezogen sein oder auf ein aussichtsloses Verfahren. Wörter sind, wie man sagt, **mehrdeutig**. Bierwisch/Schreuder (1991) tragen dieser Eigenschaft von Wörtern Rechnung, indem sie zwei Bausteine der Wortbedeutung unterscheiden, eine Kernbedeutung und an diese Kernbedeutung angegliederte alternative »Argumentstrukturen«. Erst die Kombination einer Kernstruktur mit einer für die jeweilige Äußerung gewählten Argumentergänzung stellt die jeweils aktuelle Bezeichnung des Konzepts im Kontext dar. Als eine weitere Illustration von vielen mag der Satz *The office is closed* dienen. Die Äußerung kann zu ver-

stehen geben, (a) dass die Tür des Betriebs geschlossen ist oder (b) dass der Betrieb aktuell nicht/nicht mehr arbeitet. Um diese Mehrdeutigkeit zu erklären, wird mit dem Element closed in (a) das Argument *door (of the office)* und in (b) das Argument *business (of the office)* verbunden.

Beziehungen zu semantischen Nachbarn: Wer eine Sprache beherrscht, verfügt über Wissen darüber, dass ein Wort den anderen Wörtern seines mentalen Lexikons auf unterschiedliche Weise benachbart ist. Wer Deutsch kann, wird ohne langes Nachdenken der Behauptung zustimmen, dass das Wort *Bruder* mit dem Wort *Schwester* inhaltlich enger verbunden ist, als mit *Stein* oder *Tinte* und *grün* enger mit *blau* als mit *groß*. Die semantische Erklärung für diese Intuition ist, dass zwischen Wortbedeutungen Sinnrelationen bestehen und dass es zwischen *Bruder* und *Schwester* mehr oder andere sind als zwischen *Bruder* und *Stein*. *Bruder* und *Schwester* stehen zueinander in der Sinnrelation der Ko-Hyponomie. Sie sind beide Hyponyme, Unterbegriffe, zu *Kind* und sie sind im Deutschen die beiden einzigen Verwandtheitsbezeichnungen, die zu *Kind* hyponym sind.

Evidenz

Experimentelle Evidenz für Nachbarschaftsbeziehungen: Spielen semantische Relationen zwischen Wörtern auch eine Rolle in der Realität des mentalen Lexikons? Hinweise darauf liefern unter anderem Versprecher sowie semantische Effekte beim Lexikonzugriff während der Sprachproduktion und -rezeption:

- **Versprecher:** Eine strukturelle Besonderheit aller Versprecher ist, dass das nicht geplante, also falsch produzierte Element immer in einer sprachlichen Beziehung zu dem beabsichtigten steht. Laute kollidieren mit Lauten, Silben mit Silben und wenn ein Wort mit einem inhaltlich anderen Wort kollidiert, dann eher mit einem solchen, das per Sinnrelationen mit ihm in Verbindung steht als mit einem solchen, das das nicht tut: *der SWF-3-Leser*, *ah -Hörer* (vgl. Dell 1995, S. 193 ff.; zu den Beispielen vgl. Leuninger 1993).
- **Lexikalische Erschließung:** Beim Verstehen einer sprachlichen Äußerung finden mehrere verschiedene kognitive Prozesse der Informationsverarbeitung statt. Einer davon besteht darin, dass zu der lautlichen Form der gehörten Wörter die syntaktischen und die semantischen Angaben abgerufen werden, die zu dem Wort gehören, fachsprachlich gesagt: Zu einem identifizierten Lexem wird die Lemma- und Konzeptinformation im Lexikon gesucht, die in die weiteren Verstehensvorgänge eingeht. Dieser Vorgang heißt lexikale Erschließung (*lexical retrieval*) und es ist bekannt, dass er nicht für jedes Wort und nicht unter allen Umständen gleich schnell abläuft. Kurze Wörter werden schneller gefunden als lange, häufige schneller als seltene. Zudem wird die Erschließung beschleunigt, wenn zu einem bestimmten Zeitpunkt ein Wort präsentiert wird, das dem zu suchenden inhaltlich benachbart ist. Diese Wirkung lässt sich in einem lexikalischen Entscheidungsexperiment mit Einzelwörtern bzw. Non-Wörtern zeigen, indem in visueller Präsentation dem Ziel-Item ein semantisch verwandtes Wort als Prime-Wort vorangestellt wird. Das Zielwort wird unter dieser Bedingung schneller als Wort erkannt als bei einem semantisch neutralen Vorgängerwort (vgl. Meyer/Schvaneveldt 1971).

- **Semantische Interferenz:** Ein anderer Effekt, der ebenfalls auf die Existenz semantischer Relationen zwischen Wortinhalten hindeutet, tritt in der Sprachproduktion auf. Das ist der Effekt der semantischen Interferenz. Wie der Name erkennen lässt, wirkt er sich verzögernd auf die Geschwindigkeit des Zugriffs auf ein Wort im Lexikon aus. Er ist in dem Wortentscheidungsexperiment mit eingebauter Objektbenennungsaufgabe von Levelt/Schriefers/Vorberg/Meyer (1991) nachgewiesen worden. Er wird damit erklärt, dass die Aktivierung eines lexikalischen Konzepts z. B. durch Zeigen eines Bildes des Objekts beim Benennen dieses Objekts den Zugriff auf ein gelesenes Wort mit semantisch benachbarter Bedeutung deaktiviert und damit die Entscheidung, ob dieses gelesene Item ein Wort ist oder nicht, verzögert; der Effekt tritt nicht auf, wenn vor dem zu entscheidenden Item ein semantisch neutrales Konzept aktiviert wird (s. Kap. 4).

2.1.4 | Komplexe Wörter

Dekompositionale Verarbeitung: Zwar nicht sehr häufig, aber meistens erheiternd sind Versprecher wie *Stanislaw Lem verglühe ich ehrend* (Leuninger 1993, S. 69). Sie heißen auf Englisch **Strandings** und weil sich in der deutschsprachigen Literatur noch keine andere Bezeichnung eingebürgert hat, wird dieser Name auch hier verwendet. Strandings sind Vertauschungen lexikalischer Morpheme, die Bestandteile komplexerer Wortformen sind; dabei stranden die zugehörigen Bestandteile an ihren ursprünglichen Positionen. *Strandings* betreffen also Bestandteile von **komplexen Wörtern**, sowohl Stämme von Flexionsformen wie in ... *in fremden Fischen wässern* (ebd., S. 44) als auch in Komposita ... *herrenlosen Ärmelpullover* (ebd., S. 26), in Ableitungen ... *das gemeinste Kleinsame* (ebd., S. 70) und in Kombinationen ... *die venige Lästensucherei* statt ... *die lästige Venensucherei*. Die Existenz solcher Versprecher lässt vermuten, dass die Einzelteile, aus denen die ursprünglich beabsichtigten Wörter also *ver-ehr-en*, *glüh-end*, *klein-st-e*, *ge-mein-sam* etc. im mentalen Lexikon je eine eigene Einheit bilden und dass sie bei Bedarf jedes Mal neu zusammengesetzt werden, wobei eben Vertauschungen vorkommen.

Komplexe
Bedeutungen

Entsprechend wird vermutet, dass im Verstehen solcher komplexen Wörter eine Analysephase abläuft, deren Ergebnis dann mit den Einträgen im mentalen Lexikon verglichen wird. Diese Konzeption von der Speicherung komplexer Wörter heißt **Full-Parsing-Hypothese**.

Verarbeitung als ganze Einheit: Nun schließt aber die Existenz von Strandings eine alternative Hypothese nicht aus. Da die Produktion eines komplexen Wortes mehrere Schritte umfasst, kann die Vertauschung auch in einem späteren Schritt, z. B. der phonologischen Kodierung, geschehen sein. Diese Erklärung wäre dann durchaus mit der Hypothese vereinbar, dass ein komplexes Wort als Ganzes eine lexikalische Einheit bildet und mit seiner Strukturinformation im Lexikon gespeichert ist. Diese Konzeption heißt die **Full-Listing-Hypothese**. Zwar finden sich Strandings, die die Full-Listing-Hypothese unwahrscheinlich machen: ... *gemonatete Arbeiten*. Hier ist offenbar nicht einfach *Monat* mit *Arbeit* vertauscht wor-

den, sondern es ist nach der Vertauschung genau die Pluralendung erzeugt worden, die zu dem Vertauschungsergebnis, nämlich *Arbeit* + *Plural* = *Arbeiten*, passt, nicht aber zu *Monat* + *Plural* = *Monate*. Solche Belege sind aber sehr selten.

Ökonomie und Transparenz: Allein anhand von Versprechern ist die Entscheidung zwischen den beiden Hypothesen nicht zu treffen und auch theoretische Argumente lassen sich für bzw. gegen beide Hypothesen anführen. Ein generelles Argument, das in der Kognitionspsychologie gut bestätigt ist, ist das Ökonomieprinzip. Angesichts von oberflächlich äquivalenten Organisationsalternativen bevorzugt der menschliche Geist die ökonomischere, im vorliegenden Fall also die Full-Parsing-Strategie, die weniger von dem kostbaren Speicherplatz im Gedächtnis verbraucht. Dagegen ist einzuwenden, dass diese Ökonomie mit einer unökonomischen Verschwendung von »Rechenaufwand« und Speicherzugriffen erkaufte ist, was wiederum für die Full-Listing-Hypothese spricht. Zwingender und tatsächlich auch unbestritten ist ein anderes theoretisches Argument. Komplexe Wörter, deren syntaktische und/oder semantische Eigenschaften nicht regulär aus der Oberflächenform hervorgehen, können nicht durch Analyse verstanden bzw. beim Sprechen nach Wortbildungsregeln gebildet werden, weil sie eben unregelmäßig oder, wie es heißt, nicht transparent sind. Umgekehrt gefolgt: als Ganzes wird ein komplexes Wort im Lexikon gespeichert, wenn es nicht transparent ist.

Wortfrequenz

Verwendungshäufigkeit: Ein weiteres theoretisches Argument wird aus dem erwähnten Frequenzeffekt abgeleitet. Es gibt komplexe Wörter, die als ganze in der Sprache häufiger verwendet werden, als die Bestandteile zusammengenommen. Ein Beispiel im Deutschen ist der Plural *Perlen* von *Perle*. Die Pluralform ist nach Ortmann (1978, S. 141: IV) 103 Mal in elf Millionen Wortformen laufenden Texts belegt, *Perle* gar nicht. Solche Wörter heißen **pluraldominant**, die mit umgekehrter Häufigkeitsverteilung **singulardominant**. Wenn der Frequenzeffekt allgemein gilt, würde man erwarten, dass *Perlen*, sofern es als Ganzes gespeichert ist, schneller verarbeitet, z. B. als Wort des Deutschen erkannt wird, als *Perle*. Unter Annahme der Full-Parsing-Hypothese würde man die umgekehrte Beobachtung erwarten. Da nun aber die Wörter in schriftlichen Texten nicht gleich häufig sind wie im Mündlichen, muss auch noch mit dem Faktor »Verarbeitungsmodalität« gerechnet werden. Zusammengenommen wirken sich also Transparenz, Frequenz, Modalität und schließlich doch auch das Ökonomieprinzip auf die Entscheidung aus, ob ein Sprecher ein komplexes Wort als Ganzes im Lexikon speichert oder nicht.

Experimentelle Untersuchung: Wie die Anwendung des Ökonomieprinzips zwischen Frequenz und Transparenz entscheidet oder ob beide Faktoren berücksichtigt werden, ist eine empirische Frage, der in Experimenten mit Komposita, Derivationen und Flexionsformen nachgegangen worden ist. Einen informativen und verständlichen Überblick über die Gesamtsituation bietet Schriefers (1999). Als experimentelle Aufgabe wurde in fast allen Untersuchungen die lexikalische Entscheidungsaufgabe angewendet; variiert wurden eben die Frequenz der beteiligten Form und die Modalität.

Fast alle Theorien stimmen in der Annahme überein, dass zwischen

zwei Arten von Repräsentationen von Wörtern unterschieden werden muss, der modalitätsspezifischen **Zugriffsrepräsentation**, also einer schriftlichen und einer lautlichen, und der Repräsentation im zentralen Lexikon. Auch die Prozeduren, die – bei der Worterkennung – die beiden Typen von Repräsentationen verarbeiten, werden unterschieden, nämlich in **prälexikalische** und lexikalische. Eine modalitätsspezifische Zugriffsrepräsentation ist natürlich nicht notwendigerweise komplex, sondern eben modalitätsspezifisch. Man kann theoretisch aber auch mit unterschiedlichen Strategien für beide Repräsentationen rechnen: Full-Listing in beiden Speichern, Morphemspeicherung in beiden und Full-Listing prälexikalisch neben Einzel-Morphemspeicherung im zentralen Lexikon. Da die weitaus meisten Untersuchungen zu dieser Frage in der schriftlichen Modalität durchgeführt worden sind, liegen hierzu die verlässlicheren Erkenntnisse vor.

Vier Modelle: Gegenstand experimenteller Untersuchungen waren hauptsächlich vier Modelle, die sich zwischen den beiden radikalen Hypothesen anordnen lassen. Der Full-Listing-Hypothese am nächsten steht das Modell von Butterworth (1983) und – auf Aphasiedaten gestützt – das von Seidenberg/McClelland (1989). Der Full-Parsing-Hypothese am nächsten stehen Pinker/Prince (1991). Ein Zusammenspiel beider Organisationsprinzipien nehmen Caramazza/Laudanna/Romani (1988) und Chialant/Caramazza (1995) sowie Schreuder/Baayen (1995). Erstere vertreten die Annahme, dass sowohl für Vollformen als auch für Morpheme Zugriffsrepräsentationen gespeichert sind und alle nichttransparenten nur als Vollformen. Auf welche Einheiten prälexikalisch zuerst zugegriffen wird, hängt von der Häufigkeit der Vollformen gegenüber der der Bestandteile ab. Generell, so die zentrale Annahme dieses sog. **Augmented Addressed Morphology**, kurz AAM-Modells gilt, dass auf Vollformen prälexikalisch schneller zugegriffen wird als auf die entsprechenden Einzeleinträge (vgl. Caramazza et al. 1988; Chialant/Caramazza 1995). Speicherungen auf drei Ebenen nehmen Schreuder/Baayen (1995) an, einen Speicher von Wortformen, einen mit Zugriffsrepräsentationen und eine Ebene der Lemmata, über die Einzel-Morphembedeutungen mit ihren syntaktischen Informationen verbunden sind.

Parallel Dual Route-Modell: In einer Serie von lexikalischen Entscheidungsexperimenten zeigen Schreuder/Baayen (1995), dass ihr sog. **Parallel Dual Route-Modell** die Reaktionszeitbeobachtungen am besten erklärt. Demnach gibt es einen signifikanten Einfluss der Frequenz der Vollform (pluraldominante werden schneller gefunden als singulardominante Vollformen), der Frequenz des Numerus (Singular schneller als Plural) und einen signifikanten Einfluss der Frequenz der Stammelemente in komplexen Formen (hochfrequente schneller als niedrigfrequente – bei sonst gleichen Bedingungen). Die Faktoren interagieren miteinander, was hier nur so weit ausgeführt werden kann, als die Tatsache, dass pluraldominante Vollformen trotz des höheren Analyseaufwandes (Singular schneller als Plural) schneller erkannt werden, damit erklärt wird, dass die höhere Frequenz im Speicher der Oberflächenformen die Verarbeitung stark beschleunigt. Bemerkenswert ist an diesem Ergebnis u. a. die langsamere Verarbeitung der Pluralformen. Sie könnte mit der Mehrdeutigkeit des Plu-

Daten

ralsuffixes (-en) (im Holländischen) erklärt werden, das alle Testwörter aufwiesen. Mit dieser Erklärung zusammen passt das Ergebnis einer chinesischen Untersuchung zur Verarbeitung von Komposita solche mit einfacher koordinativer Struktur (zwei koordinierte Morphemzeichen) nach der Häufigkeit der Bestandteile erkannt werden, bei der Verarbeitung solcher mit einer komplexeren inneren Struktur (Modifizierer plus Stamm) die Frequenz der Vollform die Verarbeitungsdauer bestimmt (Zhang/Danling 1992).

Fazit: Gemeinsam ist den Befunden der Flexionsformenstudie von Schreuder/Baayen (1995) und der Komposita-Studie von Zhang/Danling (1992), dass die Entscheidung zwischen Vollformen- und Einzelmorphemspeicherung letztlich von Oberflächenfrequenz und Analyseaufwand bestimmt wird. Leicht zu analysierende komplexe Wörter werden weniger wahrscheinlich als Ganze gespeichert, solche die wegen Mehrdeutigkeit oder komplexer Struktur mehr Analyseaufwand verursachen, werden eher als Ganze gespeichert.

2.2 | Strukturelles Wissen

Am Anfang dieses Kapitels wurde gesagt, dass das sprachliche Wissen eines Menschen neben Wissen über die sprachlichen Elemente (Wissen im mentalen Lexikon) auch Wissen über Strukturbildungsprozesse umfasst. Ohne diesen Teil des sprachlichen Wissens, der »losgelöst« vom Wissen über spezifische lexikalische Einheiten vorhanden ist, wäre es dem Menschen nicht möglich, Sätze nach den Strukturregeln seiner Sprache zu bilden, zu verstehen und auch nicht, grammatische »Fehler« und Versprecher in ihnen zu erkennen und zu korrigieren. Den Inhalt dieses Wissens bildet also das grammatische Regelwerk einer Sprache.

Als Struktur wollen wir hier zunächst einmal alle komplexen Einheiten verstehen, die sich aus kleineren Einheiten zusammensetzen. Da Strukturen auf allen linguistisch analysierbaren Ebenen zu finden sind – Phoneme ergeben Silben, Silben ergeben Morpheme, Morpheme ergeben Wörter, Wörter ergeben Sätze, usw. –, kann man annehmen, dass Wissen über das Bilden solcher Strukturen eben auch Möglichkeiten und Beschränkungen zur Kombination einfacher Elemente zu komplexen auf allen Ebenen betrifft.

Zur Vertiefung

Prinzip der Kompositionalität

Zur Systemhaftigkeit der natürlichen Sprache gehört als fundamentale Eigenschaft die Kompositionalität, auch »Frege-Prinzip« genannt. Sie findet sich auch in nicht-sprachlichen kombinatorischen Systemen, z. B. im Zahlensystem. Für einen zusammengesetzten Ausdruck, der nach den Regeln eines kompositionellen Systems gebaut ist, gilt, dass sich seine Bedeutung aus der Bedeutung seiner Teile und der jeweiligen strukturellen Verbindung zwischen den Teilen ergibt. Da eine natürliche Sprache in diesem Sinne systematisch und kompositionell ist, ist zu schlussfolgern, dass die mentale Grammatik diese beiden Eigenschaften repräsentieren muss. Die

psycholinguistische Untersuchung des grammatischen Wissens hat zum Ziel, herauszufinden, wie das grammatische Wissen des Menschen beschaffen ist, der seine Sprache normal beherrscht. Die Forschung ist begreiflicherweise eng auf die Ergebnisse linguistischer Analysen der Sprachstruktur bezogen, also auf grammatische Beschreibungen einzelner Sprachen und auf Theorien über die Form möglicher Grammatiken generell, also Behauptungen der Grammatiktheorie.

Strukturelles Wissen auf der phonologischen Ebene

Wissen über die Kombinierbarkeit und Distribution von Phonemen: Wenn Menschen gesprochene Sprache rezipieren, müssen sie in einem ersten Schritt aus einem mehr oder weniger kontinuierlichen Schallstrom einzelne Wörter identifizieren. Wie viele psycholinguistische Theorien zur Sprachrezeption annehmen, wird dabei auf strukturelles Wissen über die Kombinierbarkeit von lautlichen Elementen zurückgegriffen. In erster Linie zählen dazu die Erkennung akustischer Muster von Frequenzspektren und die Abbildung auf die phonologischen Informationen im mentalen Lexikon.

Ferner zählt zum Wissen über die Kombinierbarkeit und Distribution von Phonemen einzelsprachspezifisches Wissen über die Möglichkeiten zur **Besetzung der verschiedenen Positionen innerhalb der Silben** (Onset, Nukleus, Koda; s. Kap. 1.1). Im Deutschen sowie in vielen anderen Sprachen kann zum Beispiel der Onset einer Silbe komplex besetzt sein, d. h. durch die Kombination von bis zu drei phonemischen Segmenten ausgefüllt werden. Welche Segmente allerdings kombiniert werden können, ist nicht beliebig, sondern restriktiv. Zwar kann der Onset beispielsweise durch /f-p/ besetzt sein wie in *Sport*, aber nicht umgekehrt (/p -f/). Die Koda einer Silbe kann hingegen durchaus mit der Kombination /p -f/ besetzt sein wie in *grapsch*, aber nicht durch /f-p/. Derartiges distributives Wissen wird für die Segmentierung des Lautstroms genutzt, nämlich dadurch, dass für den Onset und die Koda zulässige Segmentkombinationen potentiell als Grenze zwischen einzelnen Silben identifiziert werden, womit der Prozess der Worterkennung prinzipiell beginnen kann (Zwitserslood/Schriefers/Lahiri/Van Donselaar 1993). In ganz ähnlicher Weise wird strukturelles einzelsprachspezifisches Wissen über den Wortakzent für die Segmentierung des Lautstroms genutzt (vgl. Cutler/Norris 1988). Einen Überblick über die Ebenen der phonetischen, phonologischen und silbenstrukturellen Modelle bieten Daham/Magnuson (2006).

Empirische Evidenz: Dass Strukturwissen über lautliche Eigenschaften der Sprache unabhängig von Spezifikationen im mentalen Lexikon existiert, lässt sich auch durch die Performanz von Versuchspersonen in Experimenten mit lexikalischer Entscheidung (*lexical decision*) nachweisen. Zur Erinnerung: In Experimenten mit lexikalischer Entscheidung sollen Versuchspersonen entscheiden, ob eine Folge von Buchstaben ein Wort in ihrer Sprache ist. Zum experimentellen Material gehören dementsprechend Wörter und Non-Wörter. Bei den Non-Wörtern kann man zwischen Pseudowörtern und echten Non-Wörtern unterscheiden (s. Kapitel 2.1.1 und z. B. Eiter/Inhoff 2008).

Phonologische
Regeln

Strukturelles Wissen auf der morphologischen Ebene

Wortbildung

Silben sind die lautlichen Realisierungen von Morphemen. Morpheme ihrerseits können in Kombination komplexe Wörter ergeben. Es gibt Morpheme, die die Basis eines Wortes bilden und solche, die eine spezifische Funktion anzeigen. Von ersteren wird in der Sprachwissenschaft angenommen, dass sie eine offene Klasse bilden, d. h. es können prinzipiell immer neue in einer Sprache entstehen bzw. in die Sprache aufgenommen werden. Von Funktionsmorphemen auf der anderen Seite wird angenommen, sie bildeten eine geschlossene Klasse. Strukturelles Wissen darüber, wie Basismorpheme mit Funktionsmorphemen verbunden werden können, erlaubt es einem Hörer, wichtige Informationen zu gewinnen, zum Beispiel Informationen über die Wortart eines gegebenen Elements. Dies gilt sowohl für Wortbildungs- als auch Flexionsmorpheme (Schreuder/Baayen 1995).

Strukturelles Wissen auf der syntaktischen Ebene

Satzbau

Wohlgeformtheit: In der spontanen, alltäglichen Sprachverwendung findet man immer wieder ungrammatische Sätze. Das verdeutlicht, dass wir als kompetente Sprecher einer Sprache die Fähigkeit besitzen, diese von den grammatisch wohlgeformten zu unterscheiden und im Fall, dass wir selbst eine ungrammatische Struktur gebildet haben, diese zu korrigieren. Überlegt man, aus welchen Gründen man einen Satz als grammatisch nicht wohlgeformt empfinden kann, ergibt sich eine ganze Liste:

- **Die lineare Abfolge von Konstituenten** kann zu Nicht-Wohlgeformtheit führen: *Er ihr eine Einladung gibt.*
- **Die Verletzung der Phrasenintegrität** kann zu Nicht-Wohlgeformtheit führen: *Er ihr eine gibt Einladung.*
- **Die Verletzung von Kongruenzkriterien** kann zu Nicht-Wohlgeformtheit führen: *Er gibst ihr eine Einladung.*
- **Fehler bei der Markierung von Kasus** kann zu Nicht-Wohlgeformtheit führen: *Er gibt ihr einer Einladung.*
- **Fehler bei der Abbildung semantischer Rollen auf syntaktische Funktionen** kann zu Nicht-Wohlgeformtheit führen: *Er gibt ihr eine Einladung* (wenn eigentlich gemeint ist, *Sie gibt ihm eine Einladung*)

Mehrdeutigkeit: Abgesehen von der Fähigkeit, wohlgeformte von nicht wohlgeformten Sätzen zu unterscheiden, gibt es noch ein weiteres Phänomen, an dem sich gut veranschaulichen lässt, worin strukturelles Wissen auf der syntaktischen Ebene prinzipiell bestehen könnte. Das sind **temporär und endgültig mehrdeutige Strukturen**.

Dass es in einem Satz eine temporär ambige Teilstruktur gibt, bemerkt man meistens nur, wenn man die Verarbeitung eines (gehörten) Satzes im Verlauf der Zeit betrachtet. Folgende Beispiele illustrieren temporäre syntaktische Mehrdeutigkeit (Ambiguität):

- (2–6) *Oskar glaubte Anna...*
- a. ... *kein Wort*
 - b. ... *wäre gar nicht zu Hause.*

- (2–7) *Oskar erwartete Anna...*
- a. ... *am Eingang*
 - b. ... *würde das schon richten*
 - c. ... *befiehlt den Wählern, die Ausgaben zu senken*

In beiden Beispielen ist die syntaktische Struktur bis inklusive *Anna* mehrdeutig, denn erst die nachfolgenden Elemente bestimmen wie *Anna* zu interpretieren ist, nämlich in 2–6a) als Dativ-Objekt von *glaubte*, in 2–6b) als Subjekt des Objektsatzes zu *glaubte*, in 2–7a) als direktes Objekt von *erwartete* und in 2–7b) wiederum als Subjekt des Objektsatzes zu *erwartete*. Zur Auflösung temporärer Ambiguität ist syntaktisches Regelwissen notwendig.

Temporäre Ambiguitäten lösen sich immer spätestens am Ende eines Satzes auf. Endgültig strukturelle Ambiguität hingegen bleibt auch nach der vollständigen Verarbeitung eines Satzes bestehen.

- (2–8) *Oskar schlug den Einbrecher mit dem Hammer.*
(Entweder Oskar oder der Einbrecher hatte einen Hammer)

Syntaktisches Wissen umfasst also Wissen über die Kombinierbarkeit von Wörtern zu Phrasen, von Phrasen zu Sätzen – damit ist in erster Linie die Möglichkeit der Abfolge von Phrasen gemeint – sowie Wissen über die Möglichkeiten, thematische Rollen (Agens, Patiens, Thema, Benefaktiv, Object u. a.) mit syntaktischen Funktionen (Subjekt, direktes Objekt, indirektes Objekt, etc. zu verknüpfen.

Empirische Evidenz: Der spektakulärste Nachweis dafür, dass dem Menschen syntaktisches Regelwissen zur Verfügung steht, ist ein Phänomen des Scheiterns bei dem Versuch, dieses Wissen beim Sprachverstehen zu aktivieren. Dieser Effekt tritt u. a. auf, wenn eine Person einen Satz hört, der im Anfangsteil eine syntaktisch mehrdeutige Phrase enthält, deren Mehrdeutigkeit sich erst spät im Satz auflöst, aber auf unerwartete Weise. Da der Prozess des Satzverstehens auch eine laufend aktive syntaktische Analyse umfasst, die mehrdeutige Phrase dabei in die gerade bearbeitete sofort eingliedert wird und sich dies später als Fehlanalyse erweist, scheitert der menschliche Parser. In dem Satz *Peter versprach Maria in Zukunft keine Zigaretten mehr anzubieten* wird die Dativphrase *Maria* zunächst als Objekt der Verbalphrase unter *versprach* analysiert. Erst später im Satz wird deutlich, dass sie die Funktion des indirekten Objekts von *anbieten* hat. Die Analyse bricht zusammen und muss neu gestartet werden.

Evidenz

Dass das syntaktische Wissen im Kern Wissen über die hierarchische Gliederung von syntaktischen Phrasen, ihre Funktionen im Satz, ihre Beziehung zu thematischen Rollen und ihre Präzedenz im Zusammenspiel mit Kategorien von Kontextwissen darstellt, gilt als gut bestätigt durch Beobachtungen der Aufmerksamkeitsverteilung mittels Blickbewegungsmessungen beim Lesen von Sätzen, besonders bei der Verarbeitung temporär syntaktisch mehrdeutiger Teile im Satz (einen Überblick geben Tanenhaus/Trueswell 1995; zu dem syntaktischen Unterschied zwischen Argument- und Adjazenzbeziehungen, vgl. Frazier/Clifton 1996). Zusätzliche Evidenz liefern neurolinguistische Beobachtungen aus Hirnstromab-

leitungen von ereignisrelatierten Potentialen (ERP) bei der Verarbeitung von Äußerungen mit syntaktisch überraschenden Fortsetzungen von Satzanfängen wie *Die Bahn wurde fahre*; einen Überblick gibt Hahne (1998: Kap. 5), exemplarische Experimente sind in Friederici/Gunter/Hahne/Mauth (2004) dargestellt.

Auch Prinzipien bei der Gestaltung der sprachlichen Form von Selbstkorrekturen liefern Evidenz. Levelt (1983) beobachtete, dass Selbstkorrekturausdrücke mehr Teile enthalten als lediglich die Korrektur des Fehlers, dass dabei vielmehr auch nicht korrekturbedürftige Bestandteile wieder aufgenommen und dabei syntaktische Phrasengrenzen berücksichtigt werden (vgl. zum Prinzip der syntaktischen Wohlgeformtheit von Selbstkorrekturen Levelt 1989).

2.2.1 | Strukturelles Wissen auf der semantischen Ebene

Unter strukturellem Wissen auf semantischer Ebene soll hier das Wissen verstanden werden, das sich auf die Verknüpfung von verschiedenen Bedeutungskomponenten bezieht. So lässt sich zum Beispiel die Bedeutung, die durch das Wort *Kiste* ausgedrückt werden kann, in semantischer Hinsicht modifizieren, nämlich dadurch, dass sie mit der Bedeutung, die durch ein Adjektiv wie *groß*, *schwarz* oder *mein* kombiniert wird. Durch eine derartige Kombination entsteht eine Bedeutungsstruktur. Die Möglichkeiten, wie Bedeutungskomponenten zu Bedeutungsstrukturen verknüpft werden können, sind vielfältig aber eben auch restringiert. Sie unterliegen Regeln. Des Weiteren sind derartige Regeln zumindest teilweise sprachspezifisch.

Ein eindrucksvolles Beispiel dafür, dass strukturelles Wissen auf der semantischen Ebene online, also im Moment der rezeptiven Sprachverarbeitung, von einem Hörer angewendet wird, bietet eine Eye-Tracking-Studie, die Altmann/Kamide (1999) veröffentlichten. Die Ergebnisse dieser Studie verdeutlichen unter anderem, dass Wissen über Strukturbildung auf semantischer Ebene genutzt wird, um Diskursreferenten während der Sprachrezeption vorherzusagen.

Experiment Vorhersagbarkeit von Referenten durch semantisches Strukturwissen

In einer Studie von Altmann/Kamide (1999) sahen Versuchspersonen dieselben visuellen Stimuli in zwei verschiedenen Bedingungen. Die visuellen Stimuli bestanden aus ganz einfachen Szenen, in denen verschiedene Objekte enthalten waren. Ein Stimulus zeigte z. B. einen Jungen, eine Torte, einen Ball, eine Spielzeugeisenbahn und ein Spielzeugauto. Die entscheidende experimentelle Manipulation betraf die Torte. Während die Versuchspersonen diesen Stimulus sahen, hörten sie in einer experimentellen Bedingung den Satz *The boy will eat the cake* und in einer anderen *The boy will move the cake*. Gemessen wurde in beiden experimentellen Bedingungen, wann ein Hörer seine visuelle Aufmerksamkeit auf die Torte lenkt. Es zeigte sich, dass Hörer die Torte signifikant früher anschauten, wenn der Satz *eat* enthielt im Vergleich zu *move* (Differenz ca. 200ms). Interessanterweise erfolgte die erste Augenbewegung zur Torte sogar durchschnittlich 85ms bevor das Wort *cake* überhaupt gehört wurde.

2.3 | Repräsentation sprachlichen Wissens

Die Frage nach der Repräsentation gilt dem formalen Apparat, durch den das sprachliche Wissen kognitiv kodiert ist und es konkurrieren hier hauptsächlich zwei Theorien. Die ältere beruht auf linguistischen Modellen, die eine Architektur aus symbolverarbeitenden Systemen annehmen, einem semantischen, einem syntaktischen und einem phonologischen. Jedes ist für sich genommen ein eigener Formalismus mit je spezifischen Einheiten und strukturellen Eigenschaften und der Zusammenhang des Ganzen wird im lexikalischen System sowie in gesonderten Verarbeitungsalgorithmen, einem Produktions- und einem Rezeptionssystem gesehen. Ein derartiges Modell wird als **Symbolverarbeitungsmodell** bezeichnet.

Modelle

In **Symbolverarbeitungsmodellen** wird als Symbolsprache der **Bedeutungsstrukturen** eine Logiksprache angenommen mit Konzeptkonstanten und Variablen verschiedener Typen, Quantoren, Konnektoren und Operatoren, sowie Regeln, nach denen einfache Ausdrücke zu zusammengesetzten verbunden werden können. Als Symbolsprache des **Syntaxwissens** und der Morphologie wird eine Variante einer Algebra postuliert mit einem Vokabular von Kategoriensymbolen für lexikalische und phrasale Kategorien, definiert als Bündel syntaktischer Merkmale und eine Form von Unifikationsregularitäten zur Repräsentation hierarchisch zusammengesetzter Strukturen. Am weitesten verbreitet sind Modelle mit Ersetzungsregeln, basierend auf Formalismen der Mengentheorie (vgl. Sternefeld 2008).

Das **phonologische Wissen** ist im Symbolverarbeitungsmodell repräsentiert mittels eines hybriden Formalismus mit einer begrenzten Menge universaler phonologischer Merkmale, lexikalisch gegebener Phoneme als Merkmalkombinationen in Silbenstrukturen, Regeln zur Manipulation von Phonemen und Silben in wechselnden Umgebungen, sowie Akzentkategorien und Wort-, Phrasen- und Satzakkzentberechnungsverfahren. Die Prozesse der Sprachverwendung sind kognitive Prozeduren der Informationsverarbeitung.

Die Sprachproduktion ist, grob gesprochen, eine Folge von Vorgängen zur Erzeugung von komplexen Ausdrücken innerhalb eines Teilsystems und der Übertragung der resultierenden Strukturen in Ausdrücke des nächst oberflächennäheren Moduls. Beim Sprachverstehen findet entsprechend eine Folge von Strukturanalyseschritten in Richtung von der lautlichen zur semantischen Repräsentation des gegebenen Ausdrucks statt. Für ein kleines Fragment des Holländischen ausgearbeitet ist das Syntaxmodul eines Produktionsmodells der mentalen Grammatik in Kempen/Hoenkamp (1987) und für weitere Teile der Grammatik ausgebaut in De Smedt (1990).

Konnektionismus: Von diesem Modell formal wesentlich verschieden sind Modelle vom Typ **konnektionistischer** Repräsentationen sprachlichen Wissens. Ein konnektionistisches Modell ist insgesamt ein mehrdimensionales Netz aus Knoten und Verbindungen zwischen ihnen. Ursprünglich in der Physik für die Modellierung von magnetischen Zuständen und Vorgängen in Festkörpern entwickelt, sind damit dann auch Ei-

enschaften des Zentralnervensystems beschrieben worden und – daran mehr oder weniger direkt orientiert – das völlig immaterielle grammatische Wissen. Ein konnektionistisches Modell der mentalen Grammatik ist durch drei Gruppen von Eigenschaften charakterisiert: die Knoten, die es umfasst, die Verbindungen zwischen ihnen und die Aktivationswerte dieser Knoten und Verbindungen. Diese muss man sich als abstrakte Eigenschaften vorstellen, die das ganze System sozusagen vital machen. Sie kodieren die frequenzbedingten Unterschiede in der Zugreifbarkeit auf Knoten und die Stärke und damit Geschwindigkeit des Informationsflusses durch die Verbindungen. Kategoriales Wissen ist einheitenweise durch die Knoten repräsentiert und Strukturwissen durch die Art ihrer Vernetzung im Modell.

Was im Symbolverarbeitungsmodell durch die verschiedenen Symbolsprachen von Semantik, Syntax und Phonologie beschrieben ist, ist im konnektionistischen Modell durch unterschiedliche Inhalte der Knoten erfasst und durch eine Verbindungsstruktur, die gleichartige Knoten wie etwa alle Konzeptknoten, alle Knoten syntaktischer Phrasen etc. zu je einer Ebene oder Schicht (engl. *layer*) des Netzes strukturiert.

Netzmodelle sind bewusst als Alternativen zu Symbolverarbeitungsmodellen der Äußerungsproduktion entwickelt worden (vgl. Stemberger 1982 in Auseinandersetzung mit Garretts (1980) Modell; für das Deutsche vgl. das Modell von Schade 1999; über konnektionistische Modellierungen verschiedener Prozesssysteme informiert der Sammelband von Christiansen/Chater 2001).

Die Sprachverarbeitung wird im konnektionistischen Modell als Kette von Aktivationen und Inhibitionen im Netz repräsentiert. Die wesentlichen Unterschiede zwischen den beiden Modellklassen liegen also in ihrer Architektur und der Rolle der Probabilistik in den Systemen. Symbolverarbeitungsmodelle sind modular, hybrid, indem sie verschiedene formale Teilsysteme enthalten, und deterministisch. Konnektionistische Modelle sind weniger hybrid und typischerweise probabilistisch. Sie lassen sich relativ gut mit Computerprogrammen simulieren, was die Möglichkeit bietet, ihr Produktionsverhalten anhand ihres Zeitbedarfs, der qualitativen Ergebnisse und ihrer Häufigkeiten mit dem menschlichen Verhalten anhand von Experimentdaten zu vergleichen. Konkurrenz besteht zwischen beiden Theorien hinsichtlich der Erklärungsstärke. Repräsentative Arbeiten pro und contra Konnektionismus sind z. B. Goldblum (2001: Kapitel 12) und Mayor/Gomez/Chang/Lupyan (2014).

2.4 | Die Lokalisationsfrage

Ziel psycholinguistischer Arbeit ist, wie eingangs ausgeführt, die kognitiven Zustände und Vorgänge aufzudecken, die die menschliche Sprachfähigkeit ausmachen. Kognitive Zustände und Vorgänge sind geistiger Natur, und das wirft die Frage auf, wie ihre Untersuchung von Einsichten in Aufbau und Funktionsweise des Gehirns gefördert werden könnte. Denkbar sind zwei Wege, und sie werden auch in der Forschung gegangen:

Ziele neurophysiologischer Untersuchungen: Es kann ein Ziel sein, die

Denkvorgänge insgesamt zu beschreiben, indem die zugehörigen physiologischen Zustände und Vorgänge im Gehirn angegeben werden. Das Denken an ein Wort wird also z. B. beschrieben als genau die Aktivität in der Hirnrinde, die diesen kognitiven Vorgang ausmacht. Die Sprachfähigkeit des Menschen wird unter dieser Zielsetzung also letztlich neurologisch und nicht mehr kognitionswissenschaftlich beschrieben. Wie immer man den Erkenntniswert dieses reduktionistischen Ansatzes beurteilt, das Ziel liegt in weiter Ferne. Näher liegt ein anderes. Der neurolinguistische Befund, dass beispielsweise beim Zugriff auf ein Wort im mentalen Lexikon nacheinander benachbarte aber dennoch separate neuronale Areale besonders aktiviert sind, kann als eine unabhängige Bestätigung dafür gelten, dass die lexikalische Information im mentalen Lexikon gegliedert ist und in zeitlicher Reihenfolge abgerufen wird.

Zusammenhang zwischen kognitivem und neuronalem System: Beiden Bemühungen liegt jedenfalls die Annahme zu Grunde, dass es einen unmittelbaren Zusammenhang zwischen kognitivem und neuronalem System gibt, und diese Annahme ist schwerlich zu bestreiten. Eine annähernd zwingende Veranlassung ergibt sich zudem schon aus Beobachtungen von Neurologen des 19. Jahrhunderts: Die Befunde von Post-Mortem-Sektionen von Gehirnen von Patienten mit Sprachstörungen zeigen Verletzungen des neuronalen Gewebes dominant linkshemisphärisch und markanterweise in einer eher vorderen Region des Schläfenlappens und andererseits eher im posterioren Bereich. Aus den Pionierarbeiten dieser Forschung, Paul Broca (1861), Carl Wernicke (1874) und Ludwig Lichtheim (1885), ergibt sich ein schon stark differenziertes Modell der neuronalen Areale sprachlicher Teilsysteme (zur historischen Aufarbeitung vgl. Levelt 2012: Kapitel 3). Dies alles legt grundsätzlich den Versuch nahe, sie detaillierter zu untersuchen.

Möglichkeiten der Beobachtung: Die Erfolgsaussichten werden natürlich wesentlich von den Beobachtungsmöglichkeiten bestimmt, unter denen die klinisch-chirurgische die älteste und direkteste ist. Das Abbilden von Hirnaktivitäten mittels lokaler Konzentrationen im Blutfluss ist ein indirektes und relativ langsames Verfahren (zwischen 1 u. 5 Sek.), und das Messen von elektrischen bzw. magnetischen Spannungs- bzw. Zustandsveränderungen (Elektroenzephalogramm, Magnetenzecephalogramm), ein direktes und schnelles Verfahren (1 ms) mit allerdings geringerer Auflösung. Knapp informieren über den hirnochirurgischen Zugang Dronkers/Ludy (1998), über MEG, EEG und über bildgebende Verfahren die Artikel in den einschlägigen Handbüchern von Stemmer/Whitaker (2008) und Gazzaniga (2014), speziell über EEG-Methoden Rugg/Coles (1997) und Luck (2005).

Infolge rascher technischer und methodischer Verbesserungen ändern sich die Beobachtungsmöglichkeiten derzeit schnell. In einem Einführungstext wie diesem kann nur dargestellt werden, was mehrere Jahre Bestand hat. Das sind hirnphysiologische Grundlagen, die leitenden Ziele der Neuropsycholinguistik und die wichtigsten bisher gut bestätigten Befunde. Die Zahl und die thematische Breite von Arbeiten im neurolinguistischen Paradigma haben mit der Verbesserung der Beobachtungs- und Messverfahren seit Beginn des 21. Jahrhunderts beträchtlich zugenom-

men. Sowohl die räumliche Verteilung des sprachlichen Wissens im kortikalen Netz als auch die raumzeitlichen Muster des Sprechens und Verstehens sind damit differenzierter beschrieben worden.

Neurophysiologische Grundlagen

Das Gehirn, so weiß jeder, befindet sich im Kopf, unter den Knochen der Schädeldecke, fünf gewölbten Knochenplatten, dem Stirnbein, dem Scheitelbein, den beiden seitlichen Schläfenbeinen und dem Hinterhauptbein. Es nimmt etwa das Volumen im Kopf ein, das durch eine eng sitzende Wollmütze angezeigt wird, die man sich vorne bis auf die Augenbrauen und hinten in den Nacken zieht. Mit dem übrigen Skelett ist es nach unten hin durch den Hirnstamm verbunden. Anatomisch ist das Gehirn ein Komplex aus vier miteinander verbundenen Hauptgruppen, Stammhirn, Kleinhirn, Zwischenhirn und Großhirn. Letzteres umwölbt wie eine Kappe die ersten drei Gruppen, was ihm die Bezeichnung Hirnrinde eingetragen hat.

Entwicklungsgeschichtlich ist das Gehirn im Lauf eines vieltausendjährigen Prozesses entstanden, kurz und einfach beschrieben in Kapitel 3, »Biologische Grundlagen« von Zimbardo (1995). Morphologisch implementiert sind seine Funktionen in einem in Material und Gliederung diversifizierten, ca. 1,5 kg schweren System von vernetzten Nervenzellen, versorgt mit Glukose und Sauerstoff über ein verästeltes Blutsystem und nicht erregbare sog. Gliazellen und durchzogen von größeren und kleineren Hirnkammern (Ventrikeln). Allein das Großhirn umfasst bei der Geburt ca. 20 Milliarden Nervenzellen (Neuronen), deren jede über Zuleitungen (Dendriten) und Ableitungen (Axone) direkt oder indirekt mit allen anderen verbunden ist.

Verortung der Sprache im Gehirn: In dieser Gesamtarchitektur nehmen kognitive **Funktionen der Sprache** nach heutiger Kenntnis bestimmte Areale der Großhirnrinde und subkortikalen Knoten ein, letztere jedoch ausschließlich automatisierte Routinen (mehr dazu im nächsten Abschnitt). Die Großhirnrinde ist, wie gesagt, ein kappenförmiges, stark gefaltetes Gewebe aus Nervenzellen, zwischen 4 mm (frontal) und 1,5 mm (okzipital) dick, im Äußeren walnussähnlich. Sie ist in Längsrichtung in zwei Hälften (Hemisphären) geteilt, die durch einen dichten Strang von Nervenfasern, Corpus callosum, im Inneren des frontalen Bereichs verbunden sind. Längs- und Querfurchen in den Faltungen geben jeder Hemisphäre eine zusätzliche Gliederung im sog. Hirnlappen, die im Deutschen in Analogie zu den Schädelknochen nach ihrer Lage bezeichnet sind (s. Abb. 2.3).

Eine anschauliche und für psycholinguistische Belange gut geeignete Einführung in die Neurophysiologie des Gehirns und seine Funktionen ist das im positiven Sinne populärwissenschaftliche Werk von Greenfield (1996). Eine hervorragende Internetpräsentation bietet das Stuttgarter Tutorial »Sprache und Gehirn« von Dogil/Mayer (2014), der auch Abbildung 2.4 entnommen ist.

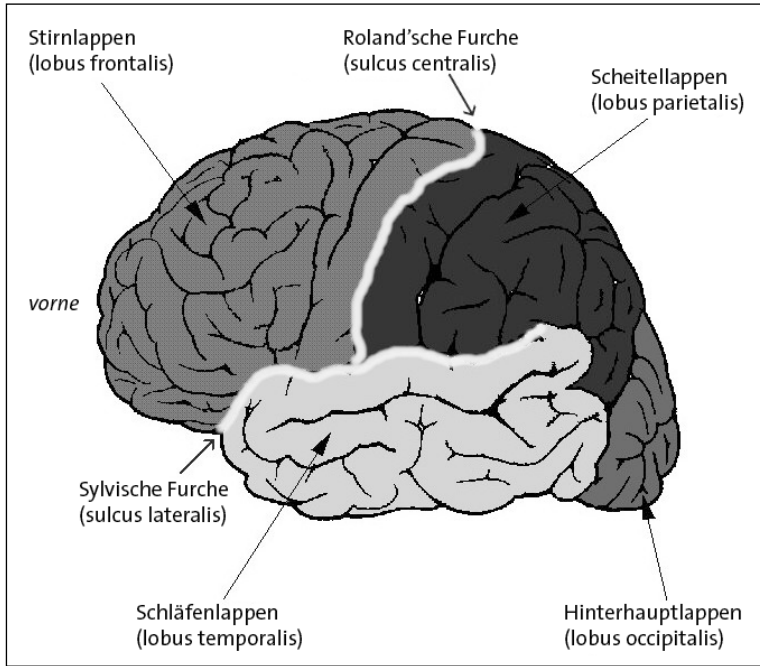


Abb. 2.3: Abbildung der Großhirnrinde, linke Hemisphäre (Quelle: Dogil/Mayer 2014)

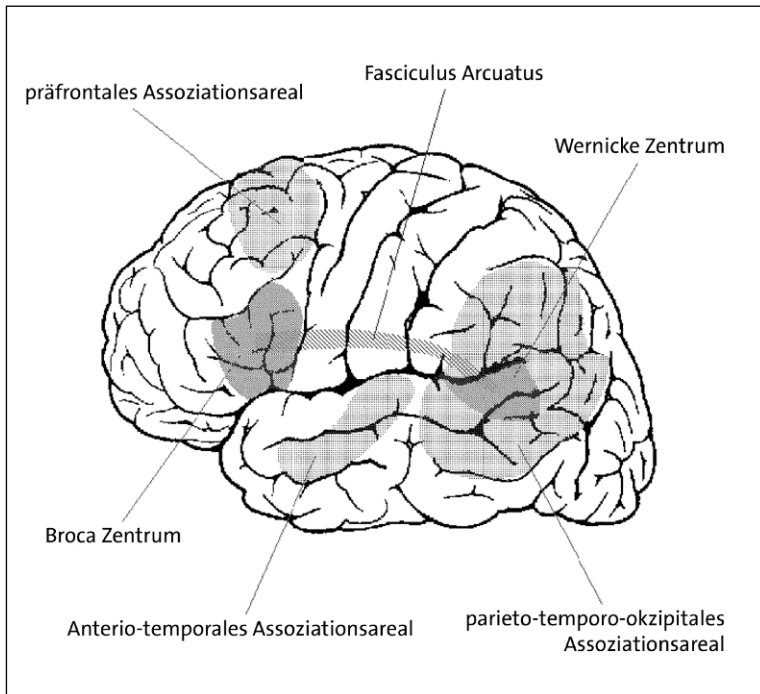


Abb. 2.4: Schema der kognitiven Funktionen in der linken Hemisphäre des menschlichen Kortex (Quelle: Dogil/Mayer 2014).

Die funktionale Anatomie der Sprache

Lokalisationstheorie: Seit Mitte des 19. Jahrhunderts wird mit hirneingebundenen Methoden der Frage nachgegangen, ob es eine Topologie der Hirnfunktionen im Kortex gibt. Für eine große Zahl motorischer Funktionen (wie die Steuerung von Muskelgruppen in den Gliedmaßen und für die neuronalen Seiten visueller und auditiver Wahrnehmung) ist diese sog. Lokalisationstheorie gut bestätigt, in neuerer Zeit auch durch Untersuchungen mit bildgebenden Verfahren (vgl. Crystal 1997: Kapitel 45). Auch für sprachliche Wissenssysteme sind spezifische Hirnregionen in besonderer Weise für relevant gehalten worden. Bis in die heutige Forschung maßgebend sind zwei Grundannahmen aus der Neurologie des 19. Jahrhunderts. Von den drei »Baugruppen« des Gehirns (Stammhirn, Kleinhirn, Großhirn) ist das **Großhirn der Sitz sprachlicher Funktionen**. Es ist lateral durch die Längsspalte (*fissura longitudinalis*) in zwei Hemisphären und – in den Rindenfeldern der Kappe – ventral/dorsal durch die quer laufende zentrale Roland'sche Furche (*sulcus centralis*) in einen vorderen und hinteren Teil gegliedert. Neurologische Befunde sprachen dafür, dem vorderen Teil motorische, dem hinteren sensorische Funktionen zuzuschreiben und die sprachlichen Funktionen der linken Hemisphäre, wobei diese jeweils sowohl das deklarative sprachliche Langzeitwissen als auch das prozedurale Wissen, die Programme, umfasst. Forschungsleitend ist das Modell noch heute insofern, als es weiterhin den Bezug für die Erklärungen und Revisionen infolge neuer Beobachtungen unter Einbeziehung grammatischer Funktions- und Strukturbeschreibungen darstellt (wie exemplarisch in Poeppel/Hickok 2004). Damit erweist es sich allerdings auch zunehmend als zu grob und unzutreffend.

Funktionale Topographie der Sprachfähigkeit: Nach heutigem Kenntnisstand kann als sicher gelten, dass:

- motorisches und sensorisches Speicher- und Prozedurwissen in vorderen und hinteren Rindenfeldern verteilt lokalisiert ist,
- auch andere Teile des Gehirns als nur das Großhirn bei der Sprachverarbeitung aktiviert sind,
- schließlich die rechte Hemisphäre des Kortex nicht unbeteiligt ist.

Gow (2012) konnte nachweisen, dass lexikalisches Wissen kortikal verteilt gespeichert ist, in einer eher dorsalen, inferior parietalen Region und einer ventralen, eher vorderen Region der linken Hemisphäre, der posterior superior temporalen Furche und dem mittleren temporalen Gyrus. In der ventralen Sektion ist das Interface zwischen lautlicher und artikulatorisch-motorischer Information installiert, in der temporalen Region das Interface zwischen phonetischer und semantischer Repräsentation.

Befunde aus Untersuchungen mit bildgebenden Verfahren, durchgeführt von Arbeitsgruppen um Gregory Hickok und David Poeppel (vgl. Hickok/Poeppel 2000; Poeppel/Hickok 2004; Okada/Hickok 2006), lassen keinen Zweifel daran, dass bei der Produktion von Sprache auch Areale im superioren posterioren Temporallappen aktiv sind und dass umgekehrt bei Perzeptionsprozessen frontale Areale der linken und der rechten Großhirnrinde Aktivität zeigen (vgl. Friederici 2011).

In Okada/Hickok (2006) z. B. wurden Experimententeilnehmern Zeichnungen von Objekten gezeigt und gleichzeitig der Objektamen lautlich präsentiert. Die Aufgabe war einmal lediglich, aufmerksam auf den Namen zu achten. In einer zweiten Phase wurden dieselben Objekte mit der Aufgabe präsentiert, sie lautlos zu benennen. Beobachtet wurde die kortikale Aktivationsverteilung mit funktionaler Magnetresonanztomographie (fMRT). Die Ergebnisse belegen weitgehende und signifikante Überlappungen von Aktivationen sowohl im Wernicke- als auch im Broca-Zentrum und links- wie rechtshemisphärisch bei Rezeptions- und Produktionsprozessen.

Experiment

Neurophysiologische Unterscheidung der Sprachproduktion und -rezeption: Der Befund, dass sich die Aktivierung bei Sprachproduktion und -rezeption im Wernicke- und Broca-Zentrum weitgehend überlappt (vgl. Okada/Hickok 2006), wirft natürlich die einfache Frage auf, wie sich dann wohl Produktion und Rezeption neurophysiologisch unterscheiden. Nahelegend wäre die Annahme, dass sie sich, wenn schon nicht räumlich, so eben im Zeitablauf unterscheiden. Das ist der Kern des Modells von Hickok/Poeppe (2000). Ihnen zufolge fließen die Informationen zwischen den beteiligten neuronalen »Arbeitsplätzen« in verschiedenen »Strömen«, einem ventralen, also vorderen, bei Produktionsprozessen und einem dorsalen Strom bei Rezeptionsprozessen (s. Abb. 2.4 und schematisch 2.5).

Modularisiertheit

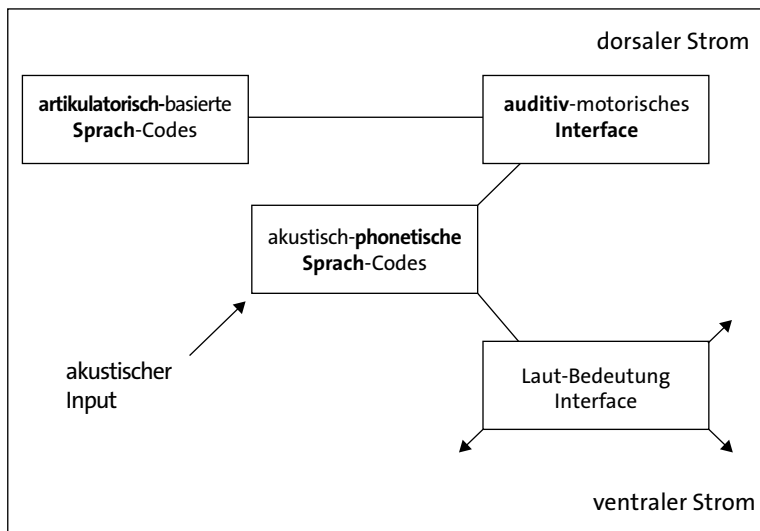


Abb. 2.5: Schema des neuroanatomischen Systems der phonologischen Verarbeitung nach Poeppel/Hickok (2004).

Evidenz durch Mehrsprachigkeit im Gehirn: Dass neben der Großhirnrinde (Kortex) weitere Module des Zentralnervensystems in die Sprachverarbeitung involviert sind, haben Untersuchungen der Mehrsprachigkeit erkennbar gemacht.

Experiment **Gehirnaktivität bei Mehrsprachigen**

Werden mehrsprachigen Personen paarweise Folgen von Substantiven aus beiden Sprachen präsentiert, deren jeweils zweites mit dem ersten entweder bedeutungsverwandt ist oder nicht, und wird dazu die Aufgabe gestellt, Eigenschaften (Größe, Farbe, etc.) des jeweiligen Zielwortes (zweites Wort) anzugeben, so wird die Bearbeitung der Aufgabe bei bedeutungsverwandten Wörtern beschleunigt. In einem solchen Experiment haben Crinion/Turner/Grogan/Hanakawa/Noppeney/Devlin/Aso/Urayama/Fukuyama/Stockton/Usui/Green/Price (2006) zusätzliche Bedingungen eingebaut, indem die Sprache der Wortpaare variiert wurde: (a) beide Wörter in der Muttersprache oder (b) Fremdsprache, oder (c) nur das zweite Wort in einer anderen Sprache. Gemessen wurden mittels Positronenemissionstomographie (PET) Aktivationen in subkortikalen Knoten. Es zeigte sich ein Priming-Effekt (schwächere Aktivität bei bedeutungsverwandtem Zielwort); im linken nucleus caudatus; der trat allerdings nur unter den Bedingungen (a) und (b) auf, wenn also beide Stimuli derselben Sprache angehörten. Sprachverschiedenheit, Bedingung (c), führte zu keiner Aktivierung.

Die Autoren interpretieren diese Ergebnisse dahingehend, dass der nucleus caudatus Sprachwechsel bzw. Beibehalten signalisiert. Friederici/Grodzinsky/Amunts (2006) erklären den Befund aufgrund von weiteren experimentellen Daten damit, dass der Sprachensensitivität des nucleus caudatus der Mechanismus zugrunde liegt, dass dieses automatisch und nicht bewusst operierende Modul durch seine Aktivität höhere kognitive Funktionen aktiviert, die den Sprachenwechsel vorbereiten.

Methodologische Herausforderungen und Lösungsversuche

Ein methodisches Hauptproblem beim Experimentieren mit Sprachverarbeitungsaufgaben besteht darin, dass bei der Bearbeitung der Aufgabe durch die Versuchspersonen eben eine Äußerung gesprochen bzw. verstanden werden muss und das ist ein komplexer kognitiver Vorgang, der lexikalische, syntaktische und lautliche Aktivitäten umfasst. Bei der geringen zeitlichen Auflösung der bildgebenden Verfahren ist es somit nicht möglich, die Aktivationsverteilung den verschiedenen, in Millisekunden aufeinander folgenden bzw. sogar parallel stattfindenden Vorgängen zuzuordnen.

Subtraktionsmethode: Einen Ausweg aus diesem Problem bildet dieses nachträgliche rechnerische Analyseverfahren. Es wird dabei mit verschiedenen sprachlichen Aufgaben experimentiert. Die Aufgaben werden so gewählt, dass sie übereinstimmende aber zusätzlich je verschiedene sprachliche Teilsysteme aktivieren. Zum Beispiel sind im lauten und leisen Wortlesen das Buchstabenerkennen und der Zugriff auf das Lemma enthalten, im lauten aber zusätzlich die phonologische Kodierung. Legt man nun die Aktivationsverteilungsabbildungen beider Experimente übereinander, so sollte im Idealfall das Laute-Lesen-Tomogramm ein Aktivationsareal zeigen, das im Leise-Lesen-Tomogramm nicht enthalten ist. Im großen Maßstab ist dieses Verfahren von Peter Indefrey und Willem J. M. Levelt ange-

wendet worden, um die Lokalisation der am Wortproduktionsvorgang beteiligten sprachlichen Teilsysteme zu bestimmen (vgl. Indefrey/Levett 2000; 2004). Dazu mehr in Kapitel 4.3.5.

Zweifel an der Lokalisationsannahme

Zur Vertiefung

Beobachtungen, die die Lokalisationsannahme grundsätzlich in Zweifel ziehen, diskutieren Bates/Dick (2000). Diese sehr kritische Sicht fassen die Autoren in einem Kommentar, dessen Schärfe am besten das wörtliche Zitat erkennen lässt:

»Likewise, in almost every laboratory that has conducted language activation studies, the first round of experiments was designed to uncover discrete and dedicated »language areas,« or at least »language-relevant areas« which (by definition) activate selectively for specific language tasks. Language-relevant areas have been duly found.

Dozens of them. Hundreds of them. Depending on the task, experimental design, population in question, and/or the protocols used in a given laboratory, virtually every region of the human brain has been implicated in at least one language activation study. The usual left perisylvian »regions of interest« show up in most studies, but the literature also abounds with positive results for right-hemisphere homologues of these zones (Hervorh. von RD und JG), as well as prefrontal regions, parietal regions, temporal areas of various kinds (basal, medial, ventral), and (where they are possible to detect) robust cerebellar and subcortical findings« Bates/Dick (2000).

Wie ist darauf zu reagieren? Die Logik dieser Bewertung wäre vernichtend, wenn man nicht zugleich zeigen würde, dass die beobachteten neuronalen Aktivationen tatsächlich sprachspezifisch sind und sich von nicht-sprachlichen kognitiven Aktivitäten in Stärke und/oder lokaler Verteilung signifikant unterscheiden lassen. Dahingehende jüngere Studien gelten genau diesem Ziel. Poeppel/Guillemin/Thompson/Fritz/Bavelier/Braun (2004) führen zu diesem Zweck eine Untersuchung zur auditiven Wahrnehmung durch.

Die Probanden hören drei Arten von phonetischem Material: Reine Tonhöhenverläufe, ansteigend und absteigend, sprachliche Silben und Wörter des Englischen. Gemessen wurde die Verteilung neuronaler Aktivitäten mit dem PET-Verfahren, welches auf die Stärke von Positronenemissionen im kortikalen Netz reagiert. Unterschiede in der Stärke ihrer Verteilung lassen sich für die separaten Hirnareale rechts und links isolieren, quantifizieren und statistisch bewerten.

Ein durchgehend bestätigter Befund ist, dass sprachspezifische Stimuli, dominant mit linkshemisphärischen Aktivitäten einhergehen, anders als die Verarbeitung purer Tonunterschiede und Silben. Ein weiteres Ergebnis betrifft die Dauer der Aktivationen. Aktivationen von längerer Dauer finden sich eher links- als rechtshemisphärisch. Studien dieser Art bestätigen und differenzieren bekannte Modelle der Neurolinguistik und sind somit geeignet, als Bestätigung kognitionspsychologische Befunde zur modularen Struktur der Sprachverarbeitung betrachtet zu werden.

3 Spracherwerb

- 3.1 Die Herausforderung
- 3.2 Die Voraussetzungen
- 3.3 Methoden – historisch
- 3.4 Die Sprachentwicklung
- 3.5 Erklärungsansätze
- 3.6 Der bilinguale Erstspracherwerb
- 3.7 Zweitspracherwerb
- 3.8 Gebärdenspracherwerb

Das Phänomen: Dass ein Kind sprechen lernt und wie es das bewerkstelligt, löst bei Erwachsenen in seiner Umgebung vielerlei Empfindungen aus. Der Vorgang wird als Faszinosum empfunden, als Selbstverständlichkeit, als Naturphänomen, als ein eigenwilliger Kampf mit dem Ziel der Angleichung an die erwachsenen Mitmenschen, als Spiel und Lust und Triebbefriedigung, kurzum als all das, was man als Erwachsener angesichts der menschlichen Sprachfähigkeit auch empfindet. Zusätzlich gewinnt die Bemühung des Kindes um die Sprache etwas Rührendes und Rätselhaftes eben dadurch, dass sie von einem Kleinkind unternommen wird, einem Lebewesen mit dem der Erwachsene in erster Linie energische, zielstrebige Tollpatschigkeit assoziiert, kaum aber die kognitive Bewältigung eines anspruchsvollen Informationsverarbeitungsproblems. Für alle Sichten auf den Spracherwerbsvorgang liefert der kindliche Umgang mit der Sprache Anhaltspunkte und Material und die folgende knappe Einführung in die Welt der Kindersprachentwicklung hätte ihr Ziel nicht erreicht, wenn die Erklärung dieser Phänomene das nicht erkennen ließe.

Die **wissenschaftliche Untersuchung des Spracherwerbs** hat die längste Geschichte unter den psycholinguistischen Arbeitsfeldern und trotzdem sind die wesentlichen Befunde aus der Frühzeit der Kindersprachforschung, die in den vierzig Jahren um die Wende vom 19. zum 20. Jahrhundert gewonnen worden sind, bis heute nicht ungültig. Ertragreich war die Arbeit von rund 150 Jahren auch insofern, als immer differenziertere Befunde herausgearbeitet wurden und als durch Entwicklungen neuer Methoden und der erforderlichen Geräte immer mehr Fakten erkannt worden sind. Auch insofern, als wissenschaftliche Arbeit letztlich danach strebt, Befunde zusammenhängend zu erklären, ist die Kindersprachforschung ertragreich gewesen, da sie im Lauf der Zeit eine gute Handvoll an Erklärungsansätzen erarbeitet hat, die die Fakten und Befunde in ein Bild bringen und auf zu Grunde liegende Prinzipien beziehen, die ihrerseits jeweils ein zusammenhängendes Ensemble darstellen. Wer sich ein differenziertes und umfassendes Wissen über die Sache verschaffen will, als es dieses Kapitel vermittelt, hat dazu reichlich Möglichkeit, es sind nämlich über die Zeit hinweg immer wieder Überblicksdarstellungen aus unterschiedlichen Perspektiven und unterschiedlichen Umfangs geschrieben worden. Eine jüngere deutschsprachige ist Kauschke (2012). Informative Handbücher

Geschichte

Überblick

sind Fletcher/MacWhinney (1995), Grimm (2000) und Grimm (2012). Die umfassendste Sammlung von Befunden aus verschiedenen Sprachen sowie von theoretischen Erörterungen darüber ist Slobin (1985). Mehrere wissenschaftliche Zeitschriften veröffentlichen aktuelle Ergebnisse. Weit verbreitet sind *Language Acquisition*, *Journal of Child Language*, *Child Development*. Für Zwecke einer Einführung in das Themengebiet lässt sich der Gesamtkenntnisstand nach einigen Merkpunkten ordnen.

- **Die Herausforderung:** Eine generelle Orientierung für den Weg durch die Welt des Spracherwerbs gewinnt man, indem man sich das Problem, vor dem der Säugling angesichts der sprechenden Erwachsenen um ihn herum steht, aus seiner Perspektive bewusst macht.
- **Die Voraussetzungen:** Die knappen verhaltensbiologischen Betrachtungen über die Verständigungsarten bei Enten und Gänsen am Anfang des Buches haben es schon deutlich gemacht: Nicht jedes Lebewesen kann jedes Verständigungssystem lernen. Die jeweilige biologische, geistige und soziale Ausstattung bzw. Umgebung der Lebewesen prägt und limitiert den Erwerbsvorgang selbst und auch Art und Umfang des Erwerbsergebnisses.
- **Methoden und Daten:** Welche Vorstellungen auch immer man sich von der Herausforderung des Säuglings durch die Sprache im Einzelnen macht, klar ist, dass die Lösung darin besteht, das (unbewusste) sprachliche Wissen installiert und die psychischen und motorischen Routinen zu seiner Verwendung beim Sprechen und Verstehen verfügbar gemacht zu haben. Der Problemlösevorgang ist also ein Lernvorgang und der lässt sich nicht unmittelbar anschauen oder messen. Aussagen darüber können letztlich nur von Verhaltens-, neuerdings auch anhand von Messwerten neurophysiologischer Vorgänge abgeleitet werden. Um die Aussagen gut zu verstehen und beurteilen zu können, ist es unabdingbar, dass man die Verfahren kennt, mit denen man die zu Grunde liegenden Beobachtungen gewonnen hat, auf denen sie beruhen.
- **Befunde:** Die Frage, wie das Kind die Sprache lernt, veranlasst einen naiverweise zu dem Gedanken, dass man, um sie zu beantworten, doch einfach hinreichend vielen Kindern einmal dabei zuschauen und Sprechen und Verstehen des Kindes über die Zeit hin festhalten und die Entwicklungsmuster in linguistischen Kategorien beschreiben sollte. Das Ergebnis eines solchen systematischen Hinschauens müsste eine Art chronologische Dokumentation der zeitlichen Reihenfolge von Einheiten und Strukturen im sprachlichen Verhalten sein. Und genau das ist der wesentliche Schritt. Soweit an dem Vorschlag überhaupt etwas naiv ist, ist es vielleicht die Annahme, er sei einfach zu realisieren.
- **Erklärungen:** Die Erwerbsvoraussetzungen, besonders die einschlägigen Anteile der geistigen und biologischen Ausstattung des Säuglings und Umstände der Auseinandersetzung mit den Menschen um es herum, liefern die Ansatzpunkte für die Erklärungen des Spracherwerbs. Diese sind nämlich Behauptungen über Beziehungen zwischen, sagen wir, spezifischen oder artspezifischen Reaktionsweisen des Kindes auf sprachlichen Input und eben dem Gang und den Ergebnissen des Spracherwerbs.

- **Bilingualismus:** Von den Voraussetzungen des Spracherwerbs haben einige mehr, andere weniger spektakuläre Auswirkungen; denken wir an den sprachlichen Input. Eine verbreitete Vorstellung ist, dass das Kind die Sprache seiner Eltern hört und die von gelegentlichen Besuchern – sonst keine, und dass die Eltern diese Sprache auch miteinander sprechen. So verbreitet diese Vorstellung ist, so wenig zutreffend ist sie. Das ergibt sich alleine aus der Verteilung der Sprachen über die Länder der Erde. Crystal (1997, S. 362) gibt dazu an, dass sich die rund 5000 Sprachen auf ungefähr 200 Länder verteilen und dass schon daraus folge, dass für eine Gesellschaft nicht Einsprachigkeit sondern Mehrsprachigkeit das Normale sei. Wie geht ein Kind mit der Erfahrung um, dass von Anfang an Sätze aus zwei verschiedenen Sprachen auf sein Ohr treffen?
- **Gebärdenspracherwerb:** Ein anderer, gleichermaßen für die Betroffenen lebenswichtiger wie für die Wissenschaft herausfordernder Fall unter den Erwerbsvoraussetzungen ist gegeben, wenn ein Kind gehörlos geboren wird. Der akustische Kanal ist für die Aufnahme sprachlicher Ausdrucksmittel nicht zu nutzen. Die sprachliche Verständigung muss also auf den visuellen Kanal verlagert werden und sich eines dafür geeigneten Ausdruckssystems bedienen, der Gebärdensprache; s. Kap. 3.9.

3.1 | Die Herausforderung

Die Dramatik der Herausforderung, der das Kind beim Erwerb der Sprache begegnet, lässt sich sehr unterschiedlich beurteilen, je nachdem, aus welcher Perspektive man es betrachtet. Lapidar: Das Kind steht vor der Aufgabe, den systematischen Zusammenhang zwischen den Lautformen der Sprache und ihren Bedeutungen zu erkennen. Pompös: Das Kind steht vor der lebensentscheidenden Herausforderung aus mehr oder weniger kryptischen Schallgebilden und sonstigen Sinneseindrücken die Bestandteile und die Kombinatorik eines abstrakten und hoch komplizierten Zeichensystems zu dechiffrieren, und das ohne nennenswerte Hilfe.

Artspezifisch

Natürlich kann man sich auch fragen, ob die Vorstellung überhaupt zutrifft, dass der Spracherwerb für das Kind eine Herausforderung darstellt. Die Aufgabe, zu lernen, wie der geeignete Platz und die beste materielle Umgebung zu bestimmen sind, ein tellergroßes Spinnennetz zu weben, die hauchdünnen Fäden zu produzieren und in wachsenden, aber regelmäßig gleichen Abständen zu verknüpfen, eine Aufgabe, vor die jede Spinne in ihrem Leben gestellt ist, würden wir vielleicht nicht als Herausforderung begreifen, weil das Tier diese Fähigkeit instinktiv entfaltet. So sieht Pinker (1994) den Spracherwerb. Und mindestens eine Erfahrung spricht stark für diese Sicht. Würde es sich um eine Herausforderung handeln, würden wir doch in einer nennenswerten Zahl von Fällen ein Scheitern erwarten. Kinder scheitern aber nicht, jedenfalls nicht, wenn sie gesund sind und unter normalen Bedingungen aufwachsen. Andererseits ist aber Scheitern auch nicht ausgeschlossen. Wird die Spracherwerbsaufgabe nämlich von einem Erwachsenen angegangen, der eine zweite Sprache lernt, ist Scheitern das Normale und wer es einmal versucht hat, wird die Herausforderung dieser Situation lebhaft nachvollziehen können.

Sprachfähigkeit im Detail

Das Ergebnis: Im Ergebnis verfügt der Mensch über die Kenntnis der komplexen Struktureigenschaften seiner Sprache, die Grammatik und einen Wortschatz zwischen 15.000 und 100.000 Wörtern. Um dahin zu kommen, muss das Kind den akustischen Reizen, die auf sein Ohr treffen, entnehmen,

- welche endlich vielen Lauteinheiten sich in ihnen finden,
- welche phonetischen Realisierungen zu einer funktionalen lautlichen Einheit gehören,
- zu welchen komplexen Mustern wie Silbe oder artikulatorische Geste sich die lautlichen Funktionseinheiten kombinieren lassen und zu welchen nicht,
- welche Lautform sich mit welchem Begriff verbindet,
- welche formalen kombinatorischen Eigenschaften ein Wort hat,
- welche endlich vielen Konstruktionsmuster den theoretisch beliebig vielen Sätzen zu Grunde liegen,
- wie sich einer komplexen syntaktischen Konstruktion aus Wörtern systematisch die entsprechende Gesamtbedeutung entnehmen lässt,
- wie die Mittel zum Bezeichnen von Sachen und Sachverhalten je nach Sprachsituation verwendet werden,
- wie sprachliche Muster sozial und nach anderen Größen variieren können
- und wie unter mehrdeutigen Äußerungen durch Bezüge zum Kontext und zur Situation die zutreffende bestimmt werden kann; Beispiel: *Der Gewinner trank das letzte Bier auf dem Tisch aus.*

Mit erfolgreichem Abschluss dieser analytischen Arbeit ist die Erwerbsaufgabe aber nicht erledigt. Zusätzlich und zwar Hand in Hand mit dem Wissensaufbau erwirbt das Kind die motorischen und kognitiven Prozeduren zur reibungslosen Aktivierung jeweils gerade derjenigen Kenntnisse, die für das Verstehen oder Produzieren einer Äußerung in einer Rede erforderlich sind. Es sind die schon einleitend ausgesprochenen Routinen,

- Schallsignale fast zeitgleich mit ihrem Auftreffen im Gehör zu segmentieren und als Phoneme und Silben zu klassifizieren,
- Schritt haltend mit dem raschen Redestrom, die Wörter im mentalen Lexikon zu identifizieren,
- die syntaktische Gliederung der Äußerung zu erkennen, um
- die Bedeutung zu »errechnen«
- und auf das Weltwissen zu beziehen und den bezeichneten Sachverhalt mental zu repräsentieren und schließlich
- mit dem Wissen, das sonst noch vorhanden ist, zu verknüpfen, sowie
- ein entsprechendes Repertoire an Produktionsprozeduren.

3.2 | Die Voraussetzungen

In Kapitel 1 ist auf biogenetische Erkenntnisse über den Zusammenhang zwischen biologischer Ausstattung einer Art und den Besonderheiten des Verhaltens der ihr angehörenden Exemplare hingewiesen worden. Immer wieder ist von Vergleichen zwischen Mensch und Tier die Rede gewesen

und so stellen wir der Beschreibung der Spracherwerbsfakten und ihren Erklärungen in den folgenden Abschnitten eine kurze Zusammensicht der Ausstattungsmerkmale des Menschen voran, die mit seiner Fähigkeit, eine natürliche Sprache zu lernen, nach heutiger Kenntnis in einem Voraussetzungenzusammenhang stehen.

Soziale Interaktion: Die erste Gruppe sind die **sozialen Voraussetzungen**. Dazu gehören ein stabiler Bezug zu und die Interaktion mit einer Brutpflege-Person, im Normalfall den Eltern. Nicht so sehr ob, sondern wie sich die Sprache eines Kindes entwickelt, hängt mit dem Gang der Sozialisation im Allgemeinen zusammen. Zwar ist der Zusammenhang zwischen Sprachverhalten und Sozialverhalten des Kindes kein direkter; der Gang der Sozialisation bestimmt aber vieles an Inhalt und Komplexität der Interaktionen mit den Mitmenschen und beides bestimmt Inhalt und Komplexität des sprachlichen Wissens (vgl. Bernstein 1962). Implizit schon angesprochen, aber doch unbedingt ausdrücklich zu erwähnen ist die Voraussetzung sprachlichen Inputs in Verwendungssituation und zwar von Geburt an. Ohne sprachliche Kommunikation entwickelt das Kind keine Sprache. Drastische Evidenz für diese Feststellung liefern die Schicksale der sog. **Wilden Kinder** (vgl. zum berühmten Fall Curtiss 1977; zu einer umfassenderen Behandlung Benzaquén 2006).

Wahrnehmungs- und Sprechorgan: Dann sind da die **biologischen Voraussetzungen**. Mit der Spracherwerbsfähigkeit direkt verbunden ist der Sprachapparat; das sind die Atemorgane (Lunge, Zwerchfell und Luftweg), der Artikulationstrakt mit Kehlkopf, Rachen, Mund- und Nasenraum und das Hörorgan mit seinen mechanischen Feinstbestandteilen im Innenohr und seinen Anschlussstellen zur neuronalen Weiterverarbeitung durch das Gehirn. Kein anderer Organismus als der Mensch verfügt artspezifisch über diese biologische Ausstattung und der Mensch verfügt darüber auch nicht vom ersten Lebenstag an, sondern lediglich über die Erbanlage, die z. B. Ansatzrohr und Gehirn sich so entwickeln lassen, wie sie es nach Abschluss des Wachstums sind. Ebenfalls zur biologischen Ausstattung zählt das Gehirn, wobei anzumerken ist, dass nicht die Masse weder des Bindegewebes noch der eigentlichen Nervenzellen, sondern die außerordentlich hohe **Dichte an Vernetzung** für die Bewältigung des Spracherwerbs und für das Sprechen und Verstehen wesentlich sind.

Die Tatsache, dass alle diese Bestandteile der biologischen Ausstattung nicht allein für Erwerb und Gebrauch der Sprache eingesetzt werden, sondern auch für die Sauerstoffversorgung des Blutes, die Nahrungsaufnahme und die Aufnahme und Produktion sonstiger Geräusche, hat Anlass gegeben, zwischen **Primär- und Sekundärfunktion** der Ausstattung zu unterscheiden, eine Unterscheidung, die hier nicht weiter erörtert wird (klug und unterhaltsam handelt darüber Jean Aitchison (1998), besonders in Kap. 3. *Grandmama's teeth*).

Sprachspezifische geistige Ausstattung: Eine dritte Gruppe von artspezifischen Voraussetzungen für Erwerb und Gebrauch der Sprache könnte man die **kognitiven Voraussetzungen** nennen, die wiederum nicht alleine für die Sprachfähigkeit zuständig sind. Dazu zählt zunächst die besondere Wahrnehmungsfähigkeit, hauptsächlich die schon von Preyer beobachtete Fähigkeit des Kleinstkindes, zwischen sprachlichen und nicht-

sprachlichen Geräuschen zu unterscheiden. Zum Zweiten – in der Reihenfolge des Erwerbs – die Denkfähigkeit, die die Fähigkeit umfasst, Begriffe zu entwickeln und Begriffe zu verarbeiten. Ebenfalls zur kognitiven Kapazität im weiteren Sinne zählt die genetisch verankerte kognitive Disposition, sprachliche Wahrnehmungen gerade so zu verarbeiten, wie es durch das Kleinkind eben geschieht; Beispiele für diese Anlage sind vermutlich angeborene Mechanismen der Segmentierung des sprachlichen Schallstroms anhand bestimmter akustischer Eigenschaften, die Fähigkeit zur Kategorisierung sprachlicher, etwa formal und/oder funktional identischer Eigenschaften wahrscheinlich unter Anwendung distributioneller Analyseverfahren, die Voraussetzung für die Entwicklung von Strategien zur Analyse struktureller Eigenschaften von Äußerungen aus den Morphemen und ihrer linearen Anordnung etc. Man wird diesen Teil der kognitiven Ausstattung in der Literatur auch unter der Bezeichnung **Angeborenes sprachliches Wissen** antreffen. Weniger direkt als alle bisher erwähnten kognitiven Voraussetzungen, aber vielleicht für die Sprachfähigkeit am ehesten eine Schlüsselfähigkeit stellt das leistungsfähige **Gedächtnis** des Menschen dar. Eine Schlüsselrolle kommt ihm insofern zu, als es die Begriffsbildung und die komplizierten Analysevorgänge überhaupt erst möglich macht.

3.3 | Methoden – historisch

Evidenz

Von der Antike bis heute sind Beobachtungen zur sprachlichen Entwicklung des Kindes aus unterschiedlichen Beweggründen angestellt worden. Sie sollten Aufschluss geben über die Entwicklung der Sprachen, über die Entwicklung des Menschen im Vergleich zum Tier, über Unterschiede zwischen der geistigen Verfassung des Kindes gegenüber der des Erwachsenen, über linguistische Annahmen über den Bau der Sprache und – seit Mitte des 20. Jahrhunderts – über die Vorgänge beim Spracherwerb und bei seinen Störungen. Diese Aufzählung spiegelt zugleich grob eine Chronologie.

Die Antike: Die Geschichtsschreibung enthält Berichte über Projekte von Herrschern, die mittels heute grausam und unmenschlich erscheinenden Methoden herauszufinden versuchten, welche die zeitlich erste unter den Sprachen der Erde war. Der nicht unplausiblen Annahme folgend, die Entwicklung der Sprache beim Kinde spiegele die Entwicklung der Sprachen in der Menschheit, wurden Kinder dabei beobachtet, welche Sprache sie als erste verwenden, wenn Sie ohne Einfluss durch sprachlichen Input der Umgebung zu sprechen beginnen. Natürlich waren für solche Isolationsexperimente mindestens zwei Kinder erforderlich, weil anzunehmen war, dass ein einzelnes Kind in Isolation mangels Veranlassung gar keine Sprache entwickeln würde. Über ein positives Ergebnis berichtet lediglich Herodot; er erwähnt Erzählungen von Priestern in Memphis, die ihrerseits über ein etwa zweihundert Jahre zurückliegendes Experiment des ägyptischen **Königs Psammetit I** berichten. Dieser soll im 7. Jahrhundert v. Chr. einen Hirten damit beauftragt haben zu beobachten, welche Wörter zwei ihm zur Aufzucht anvertraute Kinder, mit denen er nicht sprechen durfte,

zuerst verwendeten. Dem Bericht zufolge war es das Wort *becos*, das auf phrygisch soviel wie *Brot* bedeutet. Spätere Versuche sind ergebnislos geblieben; die Kinder haben den Berichten nach gar keine Sprache entwickelt (vgl. dazu die kurzen Ausführungen bei Crystal 1997, S. 230).

Die Neuzeit: Die meisten Spracherwerbsuntersuchungen haben den Charakter von Fallstudien. Entweder sind die beobachteten Kinder die eigenen Kinder der Wissenschaftler und Wissenschaftlerinnen (vgl. Preyer 1989; Stern/Stern 1920; Bloom 1973; 1993), die von Bekannten (vgl. Kaltenbacher 1990), die eigenen und die von Bekannten (vgl. Miller 1976), oder fremde, die über institutionelle Kontexte (Kinderkrippe) vermittelt sind wie die Harvard-Kinder (vgl. Brown 1973). Größere Gruppen wie in Piaget (1923) oder Bloom (1973) werden zumeist nicht **longitudinal** (Langzeit-Design) sondern **cross-sectional** (Querschnitt-Design) untersucht; Weissenborn (2003) bildet hier eine Ausnahme.

Die Methoden der Longitudinalstudien unterscheiden sich über die Zeit hin in den Medien, die für die Datenaufnahme eingesetzt werden konnten. Solange dafür keine Apparate zur Verfügung standen, war das einzige und übliche Medium das Tagebuch und das Verfahren das der teilnehmenden Beobachtung. Der oder die Beobachter nehmen hierbei eine Rolle in der Beobachtungssituation ein und dokumentieren in festgelegten Abständen das sprachliche Verhalten der beobachteten Kinder für festgelegte Zeiten in vorher bestimmten Tätigkeitsbereichen. Einen Überblick über solche Tagebuchstudien des 19. und frühen 20. Jahrhunderts gibt Richter (1927). Die bis heute umfassendsten **Tagebuchdokumentationen**, deren Ergebnisse auch noch auf die heutige Forschung wirken, sind die von Preyer (1989), dem eine Pionierrolle in dieser Methode zukommt, und die des Ehepaars Stern (1920), sowie für die Bilingualismusforschung die von Ronjat (1913) und (Leopold 1939–49). Diese und exemplarische Studien mit anderen Methoden werden im Folgenden zur Illustration ihrer Stärken und Schwächen kurz vorgestellt.

3.3.1 | Longitudinalstudien und Methoden der Verhaltensbeobachtung

Die Preyer-Studie: William Preyer hat sein eigenes Kind, Axel, während der ersten drei Lebensjahre systematisch beobachtet und sein Verhalten in einem Tagebuch dokumentiert. Um zu prüfen, wie typisch Axels Verhalten für das von Kindern generell war, hat er seine Beobachtungen mit denen aus anderen Tagebuchstudien verglichen. Zusätzliche Vergleiche mit dem Verhalten von **Jungtieren** sollten Aufschluss darüber geben, wie artspezifisch die beobachteten Verhaltensweisen sind (vgl. zu den Vergleichen im Einzelnen die ausführlichen Kommentare von Eckhardt in Preyer, 1989). Zur Beobachtungsdichte schreibt Preyer selbst (Preyer 1882, S. V–VI):

»Gerade die chronologische Untersuchung der geistigen Fortschritte im ersten und zweiten Lebensjahr bietet große Schwierigkeiten dar wegen der täglichen Registrierung von Erfahrungen, welche nur in der Kinderstube gewonnen werden können. Ich habe jedoch ein Tagebuch durchgeführt von der Geburt meines Sohnes an

Tagebücher

bis zum Ende seines dritten Lebensjahres. Da ich mit zwei unerheblichen Unterbrechungen fast täglich mindestens dreimal, Morgens, Mittags und Abends mich mit dem Kinde beschäftigte [...]«

Die Untersuchung galt nicht ausschließlich dem Spracherwerb, nicht einmal in erster Linie, sondern, wie der Titel *Die Seele des Kindes* andeutet, der Entwicklung von Wahrnehmung, Willen und Verstand; den Spracherwerb zählt Preyer zur Verstandesentwicklung. Ihm sind nur die zwei Kapitel 17 und 18 gewidmet und die Hälfte des ersten Kapitels, einer Typologie der Sprachstörungen.

Zur Vertiefung

Die wichtigsten Erkenntnisse Preyers

Aufsehen erregte hauptsächlich seine Behauptung, dass die Sprachfähigkeit vom Kind nicht durch Imitation erworben, sondern im wesentlichen Erbgut sei: »[...] erblich ist auch das Artikulieren beim Menschen und angeboren die Anlage irgendeine artikulierte Sprache zu erwerben.« (Preyer 1989, S. 280).

Zum zweiten äußert sich Preyer dezidiert zu der Frage über den Bedeutungszusammenhang zwischen der Fähigkeit begrifflichen Denkens und dem Erwerb der Sprache: »Wer die geistige Entwicklung des Säuglings gewissenhaft beobachtet hat, muss zu der Überzeugung kommen, dass die Bildung von Vorstellungen nicht an die Erlernung von Wörtern gebunden, sondern notwendige Vorbedingung für das Verstehen der ersten zu erlernenden Wörter, also für das Sprechenlernen ist.« (ebd., S. 268).

Zum dritten schließlich ist Preyer der erste, der auf die wesentliche Bedeutung des Gedächtnisses des Menschen für die Entwicklung des Verstandes hingewiesen hat (vgl. ebd., S. 225).

Die Stern-Tagebücher: Mehr als über die knappen Kapitel in Preyers Tagebüchern wissen wir über diejenigen des Ehepaares Clara und William Stern. Zum einen haben sie selbst ihre Methode im Vergleich zu dem Vorgehen von Preyer und anderen Tagebuchprojekten vor ihnen kommentiert (vgl. Stern/Stern 1909), zum Zweiten sind die Stern-Tagebücher, nachdem sie als Teil des von den Kindern Günther und Eva verwalteten Nachlasses an der Hebrew-University in Jerusalem auf Initiative der Psychologen Behrens und Deutsch wieder zugänglich gemacht worden waren, in den 1980er Jahren vollständig auf Datenträger übertragen und für psycholinguistische Auswertungen nutzbar gemacht worden (vgl. Behrens/Deutsch 1991). Motiviert waren die Aufzeichnungen durch die Absicht, Beobachtungen für kinderpsychologische Untersuchungen zu sammeln. Wie bei Preyer, richtete sich das Interesse von C. und W. Stern nicht nur auf die Sprachentwicklung. Geplant waren Studien über das Spielen des Kindes, Denken und Weltanschauung, Willens- und Gemütsleben u. a.

Das Vorgehen der Sterns unterschied sich aber in einigen Punkten von dem ihrer Vorgänger. Das Projekt war erheblich größer. Es wurde die **Entwicklung von drei Kindern** beobachtet, über lange Zeiträume hin parallel und in individuellen Tagebüchern mit insgesamt 5000 Seiten festgehalten: Hilde von 1900 bis 1913 (2474 S.), Günther von 1902 bis 1918 (1605 S.) und Eva von 1905 bis 1915 (755 S.). Die Beobachtung hat sich im Prin-

zip über den ganzen Tag erstreckt und nicht auf vorher festgelegte, immer gleiche Ausschnitte. Den Sterns war es wesentlich, dass Beobachten und Deuten von den Eltern durchgeführt wurde, nicht von Außenstehenden; nur die Eltern, besonders die Mutter, sind so kontinuierlich mit den Kindern zusammen, dass Wiederholungen von Verhalten und Zusammenhänge zwischen Episoden erkannt werden können. Clara und William Stern arbeiteten zusammen, sowohl bei den Beobachtungen selbst, wobei die Protokolle fast vollständig von Clara Stern angefertigt wurden, die wörtlichen Mitschriften der Äußerungen der Kinder von William Stern, als auch bei der Interpretation und Auswertung. Zur Illustration des Protokollstils und der Themenvielfalt hier ein paar Eintragungen in Hildes Tagebuch Band II vom 3. Nov. 1901; Hilde war damals 1; 6, 26 (1 Jahr; 6 Monate; 26 Tage):

Aus dem Stern-Tagebuch

Beispiel

3.11. – Heute Nachm. sah sich H. ein Catalogbüchlein an und ich beobachtete, wie sie plötzlich, gleichsam versuchend den rechten Arm in die Höhe streckte und ihn dann aufmerksam beobachtete, mit der linken Hand das aufgeschlagene Buch haltend. Sie nahm den Arm wieder herunter, guckte wieder ins Buch und streckte ihn dann wieder in die Höhe. Ich schlich mich an und sah zu meiner Überraschung, dass H. ein Bild nachzuzahlen suchte. [...] Wirkliche Bewegungen ahmt das Kind ja schon seit langem nach; dies ist aber das erste Mal, dass eine bildliche Situation reproduziert wurde. ...

4.11. – Gestern, als H. die Flasche getrunken hatte, sagte sie: alle, alle Milch. Wortverbindungen werden jetzt häufiger. ...

7.11. – Gestern als H. Kuchen aß, nahm sie ihren Holzmann und hielt ihm ein paar Sekunden lang den Kuchen an den Mund. [...] Der dem Kind bisher fremde Löscher von Vater's Schreibtisch wurde gestern ›Lampe‹ genannt; gleich darauf aber zur Gaskrone hinaufgesehen und ›Lampe‹ gesagt.

9.11. – Wenn die Kaffeemaschine auf dem Tisch in Thätigkeit ist, fragt H. stets: isn das? Wir antworten: Kaffee; H. spricht nach: Backe, gestern erst Backe, dann Batze. [...] Alle Flaschen heißen Mils; die Arzneiflaschen (leere) auf dem Paneel, abgebildete Liqueurflaschen in der Zeitung etc.

Stärken und Schwächen der Tagebuchdaten liegen auf der Hand. Tagebuchdaten sind **Longitudinal-Daten**; sie spiegeln das Verhalten desselben Kindes über die Zeit hin. Eltern sind mit ihren Kindern vertraut und ihre Beobachtungen sind entsprechend verständnisvoll und zuverlässig. Eltern sind mit den Kindern dauernd zusammen und nehmen Veränderungen im Verhalten sicher wahr. Andererseits sind Eltern, auch wenn sie wissenschaftlich arbeiten, nicht in allen Lebensbereichen des Kindes Fachleute. Entsprechend lückenhaft oder undifferenziert sind Daten aus solchen Verhaltensbereichen. Weder Clara noch William Stern waren Linguisten und sie weisen am Anfang ihres Buches über die Kindersprache darauf hin, dass sie über wenig phonetisches Fachwissen verfügen. Schließlich haben

Probleme alle Autoren erkannt, dass anhand von Tagebuchdaten, die ja immer nur Einzelfälle dokumentieren, nicht zwischen individuellem und allgemeinem Verhalten zu unterscheiden ist. Dem lässt sich durch Vergleiche mit anderen Beobachtungen entgegenwirken, freilich nur, soweit über die Vergleichsfälle geeignete Beobachtungen berichtet sind. In jedem Fall: Tagebuchdaten sind keine experimentellen Daten; sie eignen sich nicht für quantitative Analysen, sondern eben für die qualitative Datenexploration.

Alternativen **Ton- und Bildaufnahmen:** Die Beschränkung der Tagebuchmethode auf das Medium schriftlicher Protokolle konnte überwunden werden, sobald handliche Geräte für Ton- und später Bild- und Filmaufnahme und Wiedergabe verfügbar waren. Eine Kombination aus Tonaufzeichnung und schriftlichen Notizen – auch wörtlicher Mitschnitte von Äußerungen von Kindern und Gesprächsteilnehmern – ist das Verfahren in dem Projekt von Roger Brown, Ursula Bellugi, Colin Fraser und Mitarbeitern (vgl. Brown 1973).

Die Harvard-Kinder: Drei Kinder, Adam, Eve und Sarah genannt und zusammen später unter Bezug auf den Projektort als die Harvard-Kinder bezeichnet, wurden aus über dreißig Vorschulkindern ausgesucht, wobei der Sprachstand, die Reichhaltigkeit ihres Sprechens und die Schichtzugehörigkeit maßgebend waren. Sie waren zu Beginn der Aufnahmen 1 Jahr und 6 Monate (notiert als 1;6) (Eve) bzw. 2;3 (Adam und Sarah) Jahre alt. Wie auch Tagebuchdaten, werden Sprachaufnahmen in vertrauter Umgebung der Kinder aufgenommen, die sich übrigens bis zum Alter von 5 bis 6 Jahren durch die technischen Geräte nicht gestört fühlen. Im Vergleich zu teilnehmender Beobachtung durch die Eltern, sind Tonaufzeichnungen zeitlich natürlich viel seltener, in der Harvardstudie alle zwei Wochen.

Im Verfahren ähnlich aber in der Zeiteinteilung variabler hat Miller (1976) gearbeitet. Die Beobachtungsfrequenz ist an den Gang der Sprachentwicklung angepasst – in Zeiten rascher Entwicklung wurde häufiger aufgezeichnet. Ferner wurden alle sechs Wochen regelmäßig ausgedehnte Aufnahmen von rund sechs Kindern durchgeführt; so sollte gesichert werden, dass in regelmäßigen Intervallen der Sprachstand der Kinder umfassend dokumentiert ist. Miller hat auch außerhalb der Wohnung aufgenommen und Kommentare und Feldnotizen mündlich auf Tonband aufgenommen.

Zur Vertiefung

Vor- und Nachteile der Ton- und Videodaten

Die Tonaufnahme macht einen zusätzlichen Arbeitsschritt in der Datendokumentation erforderlich, mitunter auch zwei. Für die Analyse müssen die sprachlichen Äußerungen schriftlich wiedergegeben werden. Diese sog. Transkription muss für alle Datenteile einheitlich sein. Es muss also festgelegt sein, was in der Transkription festzuhalten ist (sprachliche Äußerung, vollständig oder ohne Versprecher und Selbstkorrekturen, nicht-sprachliche Anteile wie Pausen, Lachen, Klatschen etc.). In der Harvardstudie wurde ein gemischtes Verfahren verwendet. Die Daten eines der Kinder wurden phonetisch transkribiert, die der anderen lediglich in literarisierter, orthographischer Transkription. Nach der Transkription, für die man einen Zeitaufwand von mindestens 10:1 ansetzt, ist eine Kontrolle

der Transkripte durch eine weitere Person üblich. So soll gewährleistet werden, dass die Transkription nicht durch individuelle Hörgewohnheiten verzerrt ist.

Eignungen und Nicht-Eignungen von Tonbanddaten sind wiederum leicht zu erschließen. Der Fokus der Beobachtung liegt hier auf dem sprachlichen Verhalten des Kindes; Gesichtsausdruck, Muskelaktivitäten, Bewegungen werden nicht oder nur kursorisch erfasst. Die Transkription und ihre Kontrolle sind zusätzliche, äußerst zeitintensive Arbeitsvorgänge. Dafür liefert das Verfahren detailliertere Daten-Eigenschaften wie Satzaufbau, Morphemstruktur der Wörter, lautliche Eigenschaften, aber auch das Gesprächsverhalten (langsam, rasch, interaktiv oder eher monologisierend) ist dokumentiert.

Videoaufzeichnung ist in longitudinalen Datenerhebungen der Sprachentwicklungsforschung selten. Die Gründe dafür liegen auch in dem großen Arbeitsaufwand, der für die Transkription der visuellen Information geleistet werden muss, und zum zweiten in der Unklarheit, welche Ausschnitte des nicht-sprachlichen Verhaltens mit welchen Aspekten des Sprachverhaltens und besonders des Erwerbs systematisch zusammenhängen. Longitudinale Projekte mit Video- und Tonaufzeichnungen sind Kaltenbacher (1990) im Bereich des Erstspracherwerbs und das longitudinale und vergleichende Zweitspracherwerbsprojekt der Europäischen Forschungsgemeinschaft (ESF) (vgl. Perdue 1993).

Kaltenbacher wertet von den visuellen Daten hauptsächlich die Beobachtungen über Zeigegeesten des Kindes aus sowie – kursorisch – Informationen über die Situation des Kindes, seine aktuelle Beschäftigung und mitunter Informationen über das Verhalten von anderen Personen »auf der Bühne«. In Kaltenbacher ist übrigens auch ein typisches Beispiel für Transkriptionsrichtlinien enthalten (vgl. Kaltenbacher 1990, S. 53–60). In dem ESF-Projekt dienten die Videodaten dazu, sprachliche Äußerungen mit Raumreferenzen für die Auswertung besser verständlich zu machen. Hierbei handelt es sich im Wesentlichen um stärker formalisierte Aufgaben wie etwa Regieanweisungen der Lerner an Dritte zum Nachspielen eines zuvor beobachteten kleinen stummen Bühnensketchs. In derartigen Datensätzen liefert die Videoaufzeichnung Informationen über sprachbegleitende Umstände und Aktionen.

Einen ganz anderen Stellenwert hat die Videoaufzeichnung in Untersuchungen zum Gebärdenspracherwerb. Das Lernen und Verwenden von Gebärdensprache geschieht im visuellen Feld und um diese Vorgänge zu dokumentieren sind Videoaufzeichnungen von Nöten. In vollständiger Analogie zur Tonaufzeichnung werden auch Videodaten für die Analyse transkribiert. Das kann mit den visuell dargestellten Zeichen der Gebärdensprache geschehen. Für die Transkription deutscher gebärdensprachlicher Daten liegt eine computerbearbeitbare Umschrift vor (vgl. Prillwitz/Leven/Zienert/Hanke/Henning 1989). Für spezifische Analysezwecke werden auch spezifische Transkriptionsverfahren entwickelt, die lediglich die Verteilung spezifischer Merkmale der Daten erfassen, z. B. die Gestalt der Hand bzw. der Hände, Richtung und Ort einer Gebärde oder andere. Typische Beispiele von vollständig bildlich dargestellten Transkriptionen von gebärdensprachlichen Zeichen enthält Leuninger (2000), eine Sammlung verschiedener Einzelstudien Volterra/Erting (1998).

Datenbank **CHILDES:** Longitudinal angelegte Studien sind, wie gesagt, wegen des hohen Beobachtungsaufwandes und angesichts begrenzter Kapazitäten durchweg Fallstudien. Allerdings soll man diese Feststellung nicht so verstehen, dass das Augenmerk der Beobachtung auf den Besonderheiten des individuellen Falles liegt. Ein Weg, im gegebenen Fall zwischen den typischen und den individuellen Verhaltensweisen zu unterscheiden, ist der Vergleich mit den Daten benachbarter Untersuchungen. Ein Instrument, das den Vergleich von Daten sehr erleichtert, ist die international zugängliche Datenbank mit Daten vieler Spracherwerbsprojekte unter einem einheitlichen Transkribierformat, CHAT, zusammengestellt. Die Datenbank heißt **CHILDES** (Child Language Data Exchange System). Sie geht zurück auf eine Initiative von Brian MacWhinney; eine Beschreibung findet sich in MacWhinney (1995); eine eindrucksvolle Nutzung von CHILDES-Daten ist die Analyse von Snyder (2007).

Querschnittsdesign: Ein alternativer Weg, die Allgemeinheit der beobachteten Phänomene und damit die Reichweite der Ergebnisse zu sichern, ist natürlich, das Sample zu vergrößern und zudem, um auch den zeitlichen Aufwand zu minimieren, die Beobachtung der longitudinalen Entwicklung durch eine Querschnittsanordnung zu simulieren d. h. vergleichbare Versuchspersonen aus unterschiedlichen Entwicklungsstufen zusammenzustellen. Mit wachsender Zahl der Fälle wird dabei im Allgemeinen die Breite der Beobachtungen eingeschränkt, jedenfalls, wenn die Detailiertheit der Beobachtungen nicht aufgegeben werden soll. Testungen bestimmter Ausschnitte des Sprachverhaltens werden typischerweise zur Früherkennung von Sprachstörungen durchgeführt. Ein Beispiel ist Templin (1973). Mit dem Ziel, Ausspracheabweichungen, die im Alter von 6 bis 7 Jahren therapiebedürftig sein würden, schon im Vorschulalter zu diagnostizieren, führte Mildred Templin Tests mit mehreren tausend Kindern verschiedener Altersstufen durch. Zudem geben die Daten aus solchen Beobachtungen Aufschluss über die nicht-pathologische, normale Entwicklung z. B. des Lispelns im Zuge des Spracherwerbs.

3.3.2 | Experimentelle Methoden der Spracherwerbsforschung

Die experimentelle Beobachtung setzt generell eine begründete Annahme über das zu beobachtende Verhalten voraus. Es muss also eine **Theorie** vorliegen, aus der sich eine solche Annahme ableiten lässt und die auch Anhaltspunkte dafür liefert, mit welcher sprachlichen Aufgabe unter welchen Verarbeitungsbedingungen welche Reaktion herbeigeführt und damit das Zutreffen der Hypothese statistisch beurteilt werden kann. Untersuchungen mit experimenteller Datengewinnung gelten bisher hauptsächlich der sprachlichen Wahrnehmung des Kindes. Die generelle Überlegung, die dahingehenden Hypothesen zu Grunde liegt, geht von der gesicherten Beobachtung aus, dass die passive Sprachbeherrschung des Kindes früher entwickelt ist als die aktive. Evidenz dafür liefert die Entwicklung des lexikalischen Wissens. Wie im nächsten Abschnitt ausführlich dargelegt, kennen Kinder mehr Wörter als sie in eigenen Äuße-

rungen verwenden. Das legt rein theoretisch die Frage nahe, ob es sich nicht in anderen Bereichen der Sprache genauso verhält, z. B. in der Beherrschung der Lautstruktur oder der grammatischen Struktur des Satzes. Da man nun einen Säugling nicht nach seiner passiven Sprachbeherrschung fragen kann, muss man andere Wege finden, Aufschluss über das unbewusste Wissen des Kindes zu erlangen.

Das Nuckel-Paradigma: Von entwicklungspsychologischen und klinischen Befunden ist bekannt, dass beim Menschen schon bald nach der Geburt Verhaltensvariation zu beobachten ist, die auf ein Zusammenspiel zwischen Vertrautheit und Aufmerksamkeit zurückgeht. Ein klassisches Beispiel ist das Nuckelverhalten des Säuglings. Nuckeln ohne Nahrungsaufnahme (*non-nutritive sucking*) ist ein frühes motorisches Verhalten des Säuglings. Die Intensität dieses Saugvorgangs hängt offenbar damit zusammen, ob die Aufmerksamkeit des Säuglings durch Wahrnehmungen angeregt ist oder nicht. Diesen Zusammenhang haben Eimas/Siqueland/Juszyk/Vigorito (1971) ausgenutzt, um die linguistisch basierte Annahme zu prüfen, dass Kinder genetisch die Fähigkeit besitzen, kontinuierlich variierende Lauteindrücke zu Kategorien zusammenzufassen, eine Fähigkeit die für die Erklärung des Phonemerwerbs bedeutsam ist.

Das head-turn-preference-Paradigma: Wie das Nuckeln, lässt sich auch die Kopfbewegungen des Kindes, die Veränderung seines Gesichtsausdrucks, Stirnrunzeln, Augenbewegung u. a. für den Zugang zu seiner sprachlichen Wahrnehmungsfähigkeit ausnutzen. Nicht nur das Ein- bzw. Aussetzen von Aufmerksamkeit lässt sich erfassen, sondern auch die Aufmerksamkeitsdauer. Dazu wird anstelle der Nuckelreaktion eine andere automatische Reaktion des Kindes ausgenutzt, nämlich die, den Blick und ggf. den Kopf dorthin zu wenden, wo es Interessantes wahrzunehmen gibt. Interesse weckt u. a. sprachliches Material, das dem Kind bekannt ist und das ihm in neuem Kontext wieder begegnet. Diesem wendet es dann auch länger seine Aufmerksamkeit zu, als Material, das es zuvor nicht aufgenommen hatte. So kann man anhand der Verteilung seiner Blickrichtung ermitteln, ob es den Unterschied zwischen ausgewählten sprachlichen Proben erkennt oder nicht (vgl. Fodor/Garrett/Britt 1975; Nelson/Juszyk/Mandel/Myers/Turk/Gerken 1995). Ausführlicher und mit zahlreichen Literaturhinweisen schreiben über solche Reaktionszeitverfahren Hennon/Hirsh-Pasek/Michnick-Golnikoff (2000, S. 48 ff.) und – mit Bildmaterial aus dem Labor – Kuhl (2004).

Das ERP-Paradigma: Mit der Entwicklung schneller Messverfahren von elektrophysiologischen Zuständen und Veränderungen ist seit den 1980er Jahren auch der experimentellen Kindersprachforschung ein zusätzlicher, unmittelbarer Zugang zu Reaktionen des Kindes auf sprachlichen Input eröffnet worden. EEG-Messwerte zeigen bei geeigneter zeitlicher Kopplung mit sprachlichem Inputmaterial spezifische Muster sog. Ereignisrelatierte Potenziale (*event-related potentials, ERP*) von positiven bzw. negativen Spannungsveränderungen. Evozierte Potenziale als Daten ergänzen die Beobachtungsmöglichkeiten der Kindersprachforschung, indem sie zum Einen unter Umständen gewonnen werden können, unter denen Kinder z. B. aus Altersgründen noch kein Kopfwende-Verhalten erkennen lassen (vgl. Molfese/Molfese 1979), zum Zweiten, indem sie neurokognitive

Baby's Gehirn

Erklärungen zum Modell des sprachlichen Wissens beitragen und zum Dritten, indem sie gleichermaßen mit Tieren und Menschen durchgeführt werden und so zu Erkenntnissen über die Artspezifität von »Wissen« führen können. Einen Überblick über frühe EEG-Experimente zur kategorialen Wahrnehmung bei Erwachsenen, Säuglingen und Tieren gibt Molfese (1987); ein Überblick über jüngere Arbeiten zur Sprachbeherrschung von Säuglingen und Kleinkindern im Allgemeinen findet sich in Molfese/Narter/Van Matre/Ellefson/Modglin (2001) und Jusczyk/Pisoni/Mullernix (1992), in den letzten Jahren besonders die Arbeiten aus dem Leipziger Max Planck Institut »Cognitive and Brain Sciences« (vgl. Friederici 2005; Friederici et al. 2006; Friederici 2011).

Zur Vertiefung**Entwicklung der funktionalen Anatomie**

Die elektrischen Potentiale des Hirns, wie sie mit EEG-Methoden aufgezeichnet werden, haben, wie alle Methoden, ihre spezifischen Anwendungsgebiete. Infolge des engen zeitlichen Zusammenhangs der kognitiven Aktivität und der Veränderung der elektrischen Spannungsverhältnisse ermöglichen sie eine zeitliche Auflösung des beobachteten Geschehens im Millisekundenbereich. Die räumliche Auflösung ist hingegen geringer. Sie erlaubt zwar Unterscheidungen hinsichtlich größerer Areale der Hirnrinde, lässt aber keine Schlüsse auf die genaue Lokalisation der elektrophysiologischen Aktivität innerhalb der Komponenten des Gehirns zu. Soweit aber Erkenntnisse über die funktionale Anatomie der Sprache bei Erwachsenen als gesichert gelten, wird auch die Frage spannend, wie sich diese entwickelt hat. Arbeitet das Gehirn des Säuglings und Kleinkindes ebenso oder durchläuft die Entwicklung der Sprache im Gehirn eine lange Entwicklung bis zur Lokalisation im erwachsenen Kortex?

Eine darauf gerichtete Untersuchung müsste also ein elektrophysiologisches Modell sprachlicher Aktivitäten des Erwachsenen zugrunde legen, dann vergleichbare Beobachtungen zum Spracherwerb des Kindes anstellen und sichern, dass auch das Kind dabei zwischen sprachlichem und nicht-sprachlichem Material unterscheidet. Eben solche Befunde sind in ERP-Experimenten mit Kindern im Alter von zwei Monaten bis 2 Jahren durchgeführt worden. Getestet wurden phonetische, lexikalische und syntaktische Entwicklungen mit dem Gesamtbefund, dass der Erwerb von einem im Prinzip Erwachsenen ähnlichen Initialzustand ausgeht und anscheinend kontinuierlich verläuft (vgl. Friederici 2005).

Bildgebende Verfahren: Eine genauere räumliche Auflösung lässt sich mit Messverfahren erreichen, die gewissermaßen in das Innere des Gehirns schauen können. Ein solches ist die funktionale Magnetresonanztomographie (fMRT). fMRT-Befunde aus entsprechenden Untersuchungen bestätigen die ERP-Ergebnisse von Friederici (vgl. Dehaene-Lambertz 2004).

Fazit: Schon dieser knappe Überblick lässt die Fortschritte erkennen, die in den letzten zehn Jahren errungen worden sind. Sie bereichern die Möglichkeiten des experimentellen Forschens beträchtlich. Gerade deshalb sollte nicht aus dem Blick geraten, dass das Experiment ein Verfahren

ist, Hypothesen und ggf. Modelle und Theorien zu überprüfen. Aufgestellt werden müssen Theorien, soweit noch keine vorliegen, nach wie vor durch kreatives Durchdenken von vielen, vielen Einzeldaten.

3.4 | Die Sprachentwicklung

Es ist in den bisherigen Ausführungen öfter von dem Reichtum und der Komplexität der natürlichen Sprache die Rede gewesen. In diesem Abschnitt wird beschrieben, welchen Weg das Kind geht, um vom ersten Schrei zur Beherrschung der Sprache im Alter von ca. 3 Jahren und zum fließenden Verstehen und Produzieren von längeren Gesprächsbeiträgen im Alter zwischen 10 und 12 Jahren zu gelangen. Als Leitfaden und Gedächtnisstütze eignet sich die tabellarische Übersicht in Abb. 3.1.

Vom Schreien zum Sprechen: Es ist an dieser Stelle sicher nützlich, sich noch einmal diesen oder jenen Ausschnitt aus den Gesprächen im Flugzeug in Kapitel 1 zu vergegenwärtigen, um sich bewusst zu werden, was für das Produzieren und Verstehen von Äußerungen im Kontext beherrscht werden muss.

Mit diesem differenzierten Bild von möglichst vielem, was beim Sprechen und Verstehen einer normal schnellen sprachlichen Äußerung geschieht, bringe man nun die Vorstellung eines drei bis vier Kilo schweren und 50 cm großen Säuglings in Verbindung. Als spontane Reaktion auf den Kontrast zwischen diesen beiden Bildern wäre der Gedanke nicht unplausibel, dass den Säugling eigentlich das Geschäft des Spracherwerbs nicht interessiert, vielleicht sogar überhaupt nichts anderes als Nahrungsaufnahme, Stoffwechsel und Schlafen.

Schein und Wirklichkeit: Kaum eine Annahme könnte von der Realität weiter entfernt sein als diese. Von Geburt an ist der gesamte Organismus des Menschen darauf ausgerichtet, sich seiner Umwelt anzupassen, sich motorisch, kognitiv und emotional zu entwickeln und die oben beschriebenen Voraussetzungen zum Erwerb der Sprachbeherrschung nach dem Vorbild der Erwachsenen in seiner Umgebung einzusetzen. Dieser Eindruck ist weniger darauf gestützt, was das Kleinkind zu welchem Zeitpunkt an Sprache produziert, als darauf, zu welchen Wahrnehmungen es ab der Stunde der Geburt, in Bezug auf akustische Reize sogar schon vorgeburtlich in der Lage ist. Diese Einsicht ist noch nicht sehr alt. Für lange Zeit, etwa die ersten hundert Jahre der Kindersprachforschung, galten als maßgebend für die Einschätzung seiner Sprachkenntnis die Äußerungen des Kindes, nicht seine Verstehensfähigkeit. So schreibt Jusczyk (1997, S. 265) »*when Eimas, Siqueland, Jusczyk, and Vigorito (1971) undertook their study, there was some doubt as to whether infants could perceive contrasts before they began producing them in their own babbling*«.

Das geborene Sprachgenie: Heute ist zweifelsfrei, dass das Kleinkind schon in der sog. vorsprachlichen Zeit, also im ersten Lebensjahr zum einen sehr viel »weiß« und zum zweiten, mit seinem Ausgangswissen dem Input viel weiteres Wissen über die Sprache entnimmt. Ein Bild des Spracherwerbs ist nur vollständig, wenn es mindestens dreierlei enthält: die kategorialen Unterscheidungen, mit denen das Kind die Wahrneh-

Alter	Sprachperzeption / Sprachverstehen	Sprachproduktion
1.–3. Monat	a. Universelle Lautunterscheidung (b/p) b. Wahrnehmung unterschiedlicher sprachspezifischer Betonungstypen (Französisch vs. Japanisch vs. Deutsch)	Schreien Entwicklung der Feinmodulation
4.–5. Monat	a. Erkennen des eigenen Namens im Redefluss b. Präferenz für trochäisches Betonungsmuster	Beginn des kanonischen Lallens
6. Monat	a. Erkennen von Satzgrenzen aufgrund prosodisch-rhythmischer Merkmale (u. a. Silbenlänge, Pausen) b. Erkennen der Wörter <i>Mama</i> / <i>Papa</i>	Ausbildung des zielsprachlichen Phoneminventars (zuerst Vokale)
7. Monat	a. Erkennen von Wörtern mit dem vorherrschenden zielsprachlichen Betonungsmuster, d. h. im Deutschen betont-unbetont, z. B. <i>Háse</i> . Endbetonte Zweisilber wie <i>Alárm</i> werden noch nicht erkannt. b. Erkennen einsilbiger, unbetonter Funktionswörter, z. B. <i>das</i> , <i>von</i> , <i>sein</i>	Kanonisches, reduplizierendes Lallen (<i>reduplicated babbling</i>): <i>ba-ba</i>
9. Monat	Erkennen von syntaktischen Grenzen innerhalb eines Satzes, z. B. <i>Der Junge / spielt Ball</i> Anfänge des allgemeinen Wortverstehens: ca. 60 Wörter	Kanonisches, buntes Lallen: <i>ba-da</i> Entdecken der zielsprachlichen Silbe
10.–12. Monat	a. Verlieren der Fähigkeit, alle Laute zu unterscheiden. Es werden nur noch Laute unterscheiden, die zur Differenzierung von Wörtern in der Zielsprache dienen. b. Erkennen von Wörtern mit atypischem Betonungsmuster: wie <i>Alárm</i> c. Verstehen von etwa bis zu 150 Wörtern	Übergang zum frühen Lexikon: Substantive, z. B. <i>Hund</i> , <i>Ball</i> und erste Verben, z. B. <i>auf-</i> (<i>machen</i>)
13.–15. Monat	Syntaktische Kategorisierung mithilfe von Artikelformen	
16. Monat	Erkennung von syntaktischen Funktionen wie Subjekt und Objekt	Erste Mehrwortäußerungen (Objekt-Verb): <i>Ball spielen</i>
17. Monat	Unterscheidung von Eigennamen und Gattungsnamen aufgrund des Artikels: <i>Zeig mir Dax</i> vs. <i>Zeig mir einen Dax</i> .	AUCH- und NICHT-Konstruktionen, wie <i>Simone auch Flasche haben</i>
18.–20. Monat	Entdecken von Wortstellungsregeln und grammatischen Abhängigkeiten, z. B. zwischen der Präsenz, bzw. dem Fehlen einer Nebensatzkonjunktion und der Stellung des finiten Verbs im Deutschen: <i>Hans sagt, dass er Hunger hat</i> vs. <i>Hans sagt, er hat Hunger</i> .	Schnelle Vergrößerung des Wortschatzes
21.–30. Monat	Kinder können die syntaktische Struktur eines Satzes zur Erschließung der Bedeutung von unbekannten Verben (<i>gorpen</i>) heranziehen: z. B. <i>Hans und Lisa gorpen</i> vs. <i>Hans gorpt Lisa</i>	Erwerb von Possessivstrukturen wie <i>Vaters Haus</i> ; Erwerb der Betonungsregeln für Wörter wie <i>Mandarine</i> , <i>Elefant</i> , <i>Suppenlöffel</i>
30.–36. Monat	Verständnis von Fragen wie <i>Wer..?</i> , <i>Was..?</i> , <i>Wo..?</i> , usw., und Mengenausdrücken wie <i>jeder</i> , <i>alle</i> , usw.	Erwerb der Grundregeln der Artikelverwendung, der Verbstellungsregel im Hauptsatz, sowie der Regeln der Nebensatzbildung

Abb. 3.1: Zeitliches Muster sprachlicher Erwerbschritte in den ersten drei Lebensjahren (vgl. Weissenborn 2005, S. 1–12)

mung von sprachlichem Input organisiert und verarbeitet, die Analysemechanismen, die es auf die gewonnenen Inputdaten anwendet, um zu sprachlichen Regularitäten zu kommen, und die sprachlichen Kenntnisse, die es in der Perzeption und in seinen Äußerungen zu erkennen gibt.

Da die Analysemechanismen nicht direkt beobachtet werden können, und auch experimentell nur punktuell, kommt den Äußerungsdaten methodisch zum einen der Stellenwert zu, zu den zu Grunde liegenden Sprachkenntnissen zu führen und zum Zweiten durch deren longitudinalen Vergleich zu den systembildenden Mechanismen, in der Literatur häufig *boot strapping*-Vorgänge genannt. Die Metapher soll zu verstehen geben, dass das Kind Inputdaten, die es schon verarbeiten kann, z. B. prosodische Information, als Ansatzpunkte für Mechanismen nutzt, etwa dem, die ersten Wörter im Schallstrom zu erkennen, und sich so den Stiefel der Sprache anzieht. Bei diesem kognitiven Vorgang werden lautliche, positionale und schließlich lexikalische Informationen integriert. Ein Beispiel: Kindern werden lautliche Konturen einer Wortfolge innerhalb eines Satzes durch mehrmaliges Präsentieren vertraut gemacht, etwa: *Kaninchen fressen Blattgemüse*. In der späteren Testphase werden ihnen dieselben Wortfolgen einmal mit der zuvor familiarisierten Satzkontur und – zur Kontrolle – in anderer Struktur und Kontur präsentiert, z. B. *Die Kinder sehen, wie Kaninchen essen. Blattgemüse mögen sie besonders*. Die Aufmerksamkeitszeiten zeigen, dass Säuglinge im Alter von sechs Monaten diesen Unterschied schon verarbeiten (vgl. Soderstrom/Nelson Kemler/Juszyk 2005). Es ist wichtig, sich dieses integrative Vorgehen der Analysemechanismen immer im Bewusstsein zu halten, gerade angesichts der Gliederung der folgenden Ausführungen, in denen die Entwicklung nach den sprachlichen Strukturbereichen Phonologie, Lexik, Morphologie, Syntax und Kommunikationsaufbau der Übersichtlichkeit wegen getrennt präsentiert werden.

3.4.1 | Die lautliche Entwicklung

In der ersten, der sog. vorsprachlichen Phase der Verarbeitung von Schallereignissen in seiner Umwelt, liegt die Aufmerksamkeit des Säuglings auf einer Reihe von akustischen und phonetischen Eigenschaften der Geräusche, die auf sein Ohr treffen (vgl. Fifer/Moon 1989). Durch Unterscheidungen, zu denen das Kind in den ersten Monaten nach der Geburt fähig ist, und möglicherweise bestimmte angeborene oder frühzeitig erworbene Präferenzen für bestimmte Eigenschaften des sprachlichen Signals, etwa seine rhythmische Struktur, wird seine Aufmerksamkeit auf die menschliche Stimme, auf bestimmte Sprechweisen, sprachliche Gliederungen und schließlich auf die sprachlichen Besonderheiten seiner Umgebungssprache gelenkt. Die Ergebnisse dieser Unterscheidungen und darauf beruhender Lernprozesse spiegeln sich schon in der vorsprachlichen und dann natürlich in der sprachlichen Produktion des Kleinkindes wider.

Lautliche Entwicklung – rezeptiv

Geräuschwahrnehmung pränatal und postnatal: Zwischen der 24. und 28. Schwangerschaftswoche entwickelt das Kind die Fähigkeit zur Unterscheidung von Geräuschkonturen aus der Umwelt (vgl. Brazelton/Cramer 1991; Gervain 2015). Rezeptiv unterscheidet das Kind **Stunden nach der Geburt**, zu einer Zeit, in der es visuell kaum mehr als hell und dunkel wahrnimmt, zwischen sprachlichen und nicht-sprachlichen Schallereignissen; es wendet Kopf und Gesicht in Richtung der Schallquelle, aus der das interessierende Geräusch kommt. Ebenfalls gleich nach der Geburt kann es zwischen der Stimme der Mutter und der anderer Menschen unterscheiden. Innerhalb der ersten vier Wochen nach der Geburt lässt das Kind Vorlieben für bestimmte Lautstärken und Tonhöhen erkennen, etwa die einer Frauenstimme in normaler Lautstärke. Auch in der ersten Woche ist eine Präferenz für eine spezifische Sprechhaltung entwickelt, die in der englischsprachigen Fachliteratur *infant directed speech* heißt.

Cooper/Aslin (1990) haben in einem Aufmerksamkeitsverteilungsexperiment (Dauer von Blickfixierung in einem Präferenzparadigma) 2 Tage alte und einen Monat alte Babys mit zwei Arten von sprachlichen Registern konfrontiert: Das ID-Register (Infant directed) enthält kürzere Sätze, deutliche intonatorische Konturen, deutliche Markierungen der Akzente und der Pausierung. Es wird von Erwachsenen auch als an ein Kleinkind gerichtet empfunden. Es ist nicht mit Baby-Talk zu verwechseln, verstümmelten Äußerungen und Nonsens-Wörtern wie *eideidei*. Eine von beiden Gruppen von Babys nicht präferierte Redeweise heißt *adult directed speech* (AD) und weist eben alle Merkmale des Registers zwischen Erwachsenen auf.

Alle diese Präferenzen gelten ersichtlich nur den Modalitäten der Rede-Verwendung, nicht aber sprachlich relevanten Strukturphänomenen. Aber auch hier ist das Kleinkind schon sehr früh zu Unterscheidungen in der Lage. So akzeptiert es u. a. ab der zweiten Woche phonetisch ähnliche aber doch auch geringfügig verschiedene konsonantische Varianten als äquivalent, wenn sie sich innerhalb eines zeitlichen Limits des Stimmeinsatzes nach Beginn der Artikulation (Voice Onset Time, VOT) unterscheiden, wie mit dem schon erwähnten Nuckelexperiment (*High Amplitude Sucking*) nachgewiesen worden ist. Ebenfalls ab Geburt reagieren Kinder auf Unterschiede zwischen trochäischem und jambischem Silbenpaar, also betont-unbetont (*Vater*) vs. unbetont-betont (*Gehalt*) (vgl. Höhle/Weissenborn 1999).

Zum Begriff

Infant directed speech ist ein von Babys präferiertes sprachliches Register, welches sich unter anderem durch kurze Sätze mit einfachen grammatischen Strukturen, ein langsames Sprechtempo, eine höhere Tonhöhe, deutliche Markierung von Pausen, eine stärkere Akzentuierung und eine hohe Wiederholungsrate der verwendeten Phrasen und Sätze auszeichnet.

Vokalunterscheidung: Ab dem sechsten Monat erkennt das Kind Kontraste zwischen vokalischen Stimuli. Hierzu muss man wissen, dass – anders als konsonantische Kontinua – vokalische nicht in kategoriale Äquivalenzklassen zerlegt werden, weder vom Kind noch vom Erwachsenen. Zu den früh wahrgenommenen Kontrasten zählen [a] vs. [i] und – etwas später [a] vs. [o] (vgl. Kuhl 1979; 1983). Die Kontraste werden unabhängig von der Stimme wahrgenommen (vgl. Jusczyk et al. 1992). Das wurde in HAS-Experimenten (*High Amplitude Sucking*) beobachtet. Die Entdeckungen von Kuhl sind in einem Reaktionszeitexperiment im *head-turn preference*-Paradigma bestätigt worden (Kuhl 2004) und mittels **neurophysiologischer Messungen** präzisiert in Cone (2015).

EEG auch schon beim Kleinkind

Seit den 1980er Jahren sind auch **EEG-basierte Untersuchungsmethoden** entwickelt worden, um die rezeptiven sprachlichen Fähigkeiten von Kleinkindern zu prüfen. Eine auf Lautverschiedenheit reagierende elektrophysiologische Aktivität ist die sog. **Mismatch-Negativity** (MMN). Es handelt sich dabei um eine signifikante Negativveränderung der Spannungslage im Gehirn in einer Distanz von ca. 200 ms nach dem Reiz. Die MMN reagiert auf die Wahrnehmung von unerwarteten Lautveränderungen nach einer Folge als gleichartig empfundener Schallsignale. Sie ist zunächst bei Erwachsenen beobachtet worden, dann – unter Ausnutzung dieser Befunde – bei Kleinkindern. Beobachtungen mit dem MMN-Paradigma bestätigen die Fähigkeit zur kategorialen Lautwahrnehmung schon bei frühgeborenen Kindern gleich nach der Geburt (vgl. Cheour/Ceponiene/Lehtokoski/Luuk/Allik/Alho/Näätänen 1998). Mit demselben Verfahren wurde der Befund von Werker/Lalonde (1988) bestätigt, dass die Fähigkeit zur phonematischen Diskriminierung ab dem 6. Lebensmonat zunehmend auf die Merkmale reduziert ist, die für die Phonemunterscheidung der Muttersprache des Kindes funktional sind; kategoriale Kontraste aus anderen als aus der Muttersprache werden im Schnitt ab dem 12. Lebensmonat nicht mehr wahrgenommen (vgl. wiederum Cheour et al. 1998). Die durchaus nicht einheitliche Forschungslage zu diesem Schritt in der Entwicklung referiert Werker/Goodman/Nusbaum (1994). Die Autoren weisen nach, dass die Einschränkungen nicht auf reduzierte Wahrnehmungsfähigkeit zurückgehen. Erklärungen für die Lautdiskriminierungsfähigkeit des Kindes diskutiert Best (1994).

Zur Vertiefung

Anfänge des Verstehens: Um das **Alter von 6 Monaten** herum sind auch Anfänge von sprachlichem Verstehen beobachtet worden; das heißt, dass das Kind Lautfolgen im Gedächtnis mit Inhalten verbindet. Auch diese Entwicklung umfasst mehrere Schritte, die in einem späteren Abschnitt über Theorien des Spracherwerbs dargestellt werden (s. Kap. 3.5). Deshalb hier nur kurz so viel: In der zweiten Hälfte des ersten Jahres stellt sich das Kind von der Wahrnehmung universal vieler Lautunterschiede nach und nach auf die Wahrnehmung derjenigen seiner Umgebungssprache um. Um die möglichen Ursachen dafür zu verstehen, muss man drei Phasen des komplexen Wahrnehmungsvorganges unterscheiden. Die

Vom Hören
zum Sprechen

erste ist die der (sensorischen) Empfindung, die zweite die der perzeptuellen Organisation der empfundenen Reize zu einer Wahrnehmungseinheit; dieser Prozess wird als Wahrnehmung im engeren Sinne bezeichnet. Der dritte Einzelvorgang ist der der Klassifikation. Eine Wahrnehmungseinheit wird als Element einer Klasse erkannt. Die Anpassung des Kindes an die Lautmuster seiner Umgebungssprache kann nun entweder in einer Einschränkung seiner Empfindungsmöglichkeiten bestehen oder in dem Aufbau von gerade den Wahrnehmungseinheiten seiner Umgebungssprache in seinem Wissen. Aus der Beobachtung des Zweitspracherwerbs bei Erwachsenen ist Letzteres bekannt (vgl. Flege 1991). Auch ist gezeigt worden, dass Erwachsene neue Lautsysteme lernen können, was die Annahme verlorengegangener Empfindungsmöglichkeiten als Ursache schwächt.

Phonologisches Wissen: Wie auch immer die Anpassung an die lautlichen Muster einer Sprache sich erklären, ein wesentlicher weiterer Schritt in der Entwicklung des phonologischen Wissens ist die Entdeckung der **Laut-Bedeutungs-Relation**, d. h. der Erwerb von lexikalischem Wissen, dem Gegenstand des nächsten Abschnitts (s. Kap. 3.4.2). Die Kenntnis der phonologischen Gliederung seiner Sprache entsteht im Wissen des Kindes nicht schlagartig von einem Moment zum nächsten und auch nicht Phonem für Phonem. Entwickelt werden sukzessiv Kenntnisse von zielsprachlich relevanten phonologischen Kontrasten und zwar in bestimmten Positionen von Silben. Am Anfang der phonologischen Entwicklung steht also die Silbe. In welcher Reihenfolge welche Kontraste vom Kind erworben werden, wird aus den frühen Äußerungen des Kindes abgeleitet und wir stellen diesen Teil der Sprachentwicklung deshalb anschließend im Rahmen der Produktionsentwicklung dar. Dass Kinder aber schon davor über Kenntnisse der Silbeneinheit und auch der in ihrer Umgebungssprache dominanten Verteilung der Akzente in Zweisilblern verfügen, zeigen wiederum Rezeptionsuntersuchungen mit der *head-turn-Preference-Methode* (vgl. Mattys/Juszyk/Luce/Morgan 1999).

Lautliche Entwicklung – produktiv

Schreiben und Sprache

Die lautliche Produktion des Kindes, soweit sie zur Entwicklung der Sprache zählt, folgt der Entwicklung der rezeptiven Fähigkeit – notwendigerweise. Diese hat ja auch eine pränatale Periode, jene nicht. Die Entwicklung der Lautproduktion ist Bestandteil der Gesamtentwicklung des Kindes, seines körperlichen Wachstums, der Ausdifferenzierung der Psychomotorik der kognitiven Fähigkeiten und schließlich seiner sozialen Anpassung an die Umwelt. Der Gang der Entwicklung ist ein Zusammenspiel mehrerer parallel stattfindender Entwicklungsstränge, deren Zustände ineinandergreifen. Die Hauptstränge sind das Wachstum des Körpers und die damit einhergehenden Veränderungen der Artikulationsorgane, das Wachstum des Gehirns und die Ausdifferenzierung des vegetativen und des Zentralnervensystems, das Wachstum des sprachlichen Wissens und die Ausdifferenzierung in der sozialen Interaktion, die Entwicklung der sozialen Interaktion und der Emotionen. Die in zeitlicher Reihenfolge von Geburt bis etwa zum 6. Lebensjahr beobachtbaren

Schritte des Äußerungsverhaltens des Kindes lassen eine Abfolge von Ausdifferenzierungen erkennen. Mit dem nun schon mehrfach verwendeten Ausdruck ›Ausdifferenzierung‹ soll zu verstehen gegeben werden, dass nichts verloren geht, sondern im Wesentlichen neues Wissen, neue Fertigkeiten hinzukommen, die vorhandenen Möglichkeiten aber nicht entfallen. Auch Erwachsene könne noch schreien, gurren, lallen, aber eben auch artikuliert sprechen. Nebenbei sei daran erinnert, dass Fertigkeiten mangels Training natürlich ›einrosten‹ können.

Schreien, Gurren, Babbeln: Am Anfang der Ausdifferenzierung steht das Schreien. Schreien ist ein neuronal gesteuerter Vorgang, der das motorische Zusammenspiel von Atmung, Kehlkopfaktivitäten und Ansatzrohr (die Atemwege oberhalb der Stimm lippen) erfordert. Auf die wesentliche Beteiligung der neuronalen Aktivitäten weist hin, dass zerebrale Störungen sich im Schreien zeigen und entsprechend auch das Schreien ein Indikator für Unauffälligkeit der Hirnentwicklung des Säuglings und damit der neuronalen Voraussetzung für den gesunden Spracherwerb ist (Wermke 2008; vgl. Mende/Wermke 1998).

Die lautliche Ausdifferenzierung in den ersten drei Monaten ist von der physiologischen Entwicklung bestimmt; hierzu und zur Lautentwicklung im Deutschen vgl. Elsen (1991), zum Erwerb des phonologischen Wissens Penner (2000). Der Zustand des Brustkorbs und das flache Ansatzrohr mit hohem Kehlkopf schränken die Artikulationsmöglichkeiten auf Schreien, **Gurren** und **Babbeln** ein. Im dritten Monat ist melodische Differenzierung der Stimmgebung zu beobachten, auch Modulation der Stimmstärke, erste Versuche von Lautnachahmungen; am häufigsten sind Konsonanten, besonders [h, r, k] im vierten Monat zunehmend Vokale, eher die mittleren. Im vierten bis sechsten Monat, der auch *vocal play-Phase* genannten Zeit, versucht das Kind, eben die Unterscheidungen zu realisieren, die ihm sein Artikulationsorgan erlaubt, besonders vordere Konsonanten. Im Zuge der Absenkung des Kehlkopfes vergrößert sich die Mundhöhle, eine Voraussetzung für die Differenzierung des Vokalismus. Die Laute zeigen erste Annäherung an die Umgebungssprache, Laute werden zu Silben zusammengefügt, die CV-Struktur (Consonant-Vocal) aufweisen. Es beginnt die Phase des **Lallens**, in der englischen Literatur *reduplicated babbling* genannt, deutsch auch Plappern.

Akzentmuster der Muttersprache: Im Deutschen zeigen sich Kombinationen von zwei gleichen Silben und sehr bald das trochäische Akzentmuster mit Akzent auf der ersten Silbe, das die meisten Wörter der Erwachsenensprache aufweisen. Ab dem zehnten Monat werden verschiedene Silben kombiniert, auch zu längeren Ketten. Das Repertoire der Laute differenziert sich, es treten Liquide auf, untere Konsonanten, Reibelaute und gerundete Vokale.

Gegen Ende des ersten Jahres weisen die geplapperten Silbenketten zielsprachennahe Intonationskonturen auf. Gurren, Babbeln und Plappern sind begleitet von Entwicklungen des Gesichtsausdrucks (Lächeln!) und der Motorik der Gliedmaßen. Gegen Ende des ersten Jahres erscheint das **erste Wort**. Eine Lautform wird mit einer kommunikativen Funktion verbunden; meist der Kundgabe von Affekten. Die Verbindung ist nicht immer klar und nicht konstant und Babbeln und Plappern vorsprachlicher

Schreien ist
sprachspezifisch

Lautketten dauern weiter an. Die ersten Wörter haben die Form der späten Plapperausdrücke, also CV-Silbe und Kombinationen davon, untere, mittlere Vokale, eben die ersehten [mama], [papa]. Informative tabellarische Übersichten über die Lautentwicklung ihrer Tochter Annalena gibt Elsen (1999, S. 50–53). In der Ausdifferenzierung der Sprechphase, vom neunten Monat bis etwa zum Ende des dritten Jahres beobachtet sie zehn Entwicklungsstufen.

Konsonantischer und vokalischer Strang: In dieser Entwicklung zeigen sich mehrere parallel laufende Stränge von Ausdifferenzierungen (vgl. die ausführliche Langzeitstudie von Elsen 1991, S. 58–63). Die markantesten sind die folgenden:

- Der **konsonantische Strang** von den vorderen zu den hinteren Konsonanten, von den Sonoranten über Liquide zu den Obstruenten, von den Plosiven zu den Frikativen.
- Der **vokalisches Strang** von den mittleren und unteren zu oberen, vorderen und hinteren, von ungerundeten zu gerundeten, z. B. zunächst [a] für <Öl>, und von einfachen Vokalen zu Diphthongen, z. B. [na] für <nein>.
- Der auffälligste Strang ist der der **Silbenstrukturentwicklung**. Er weist, grob gesprochen, einen Verlauf von der offenen Silbe zur geschlossenen auf, z. B. [bu] für <Buch>. Geschlossene Silben treten anfangs in Zweisilbern mit Gelenk-Konsonant wie /nanə/, <Banane> auf.

Die Komplexität der Silbenränder nimmt zu mit initialen, später dann finalen Konsonantenclustern. Die Silbenstruktur des Wortes schließlich wächst vom Einsilbler bzw. Zweisilber mit Anfangsakzent, das minimale Wort im Deutschen, das einen Fuß umfasst, zum Jambus und zum Mehrsilber mit Pänultimaakzent (vorletzte Silbe).

Bei der Deutung dieser Daten muss man stets berücksichtigen, dass die Entwicklung der Perzeption und der Lautungsmöglichkeiten nicht synchron verlaufen (Ende des ersten, Anfang des zweiten Jahres).

Dass und wie weit die von Elsen gefundenen Phasenmuster verallgemeinerbar sind, zeigt der Vergleich mit der Beschreibung der Entwicklungen in anderen Sprachen (vgl. Locke 1993; Penner 2000). Es ist nicht auszuschließen, dass die produzierten Formen gar nicht die lexikalische Repräsentation der Wortformen wiedergeben, dass vielmehr die Reduktionen nicht ein Ergebnis der Perzeption, sondern der Produktion sind.

3.4.2 | Erwerb des Wortschatzes

Die Aufgabe, vor die das Kind beim Aufbau lexikalischen Wissens gestellt ist, veranschaulicht man sich am besten, wenn man das Kapitel 2.2 über das mentale Lexikon noch einmal durchdenkt. Der Wortschatzerwerb ist nämlich gerade in dem Maße erfolgreich, wie sein Ergebnis ein solches mentales Lexikon mit 50.000 oder mehr lexikalischen Einheiten ist, einem ausgetüftelten Netzwerk zu seiner Repräsentation und mit Millisekunden schnellen Zugriffszeiten auf im Prinzip jede Einzelinformation. An dieser

Aufgabe arbeitet der Mensch letzten Endes lebenslang. Ein kommunikationsstüchtiges lexikalisches Wissen ist aber schon im Vorschulalter ausgebildet. Es weist alle strukturellen und prozeduralen Eigenschaften auf, Gliederung in Einheiten, lexikalische Ebenen, generative Systeme und Abrufprozeduren für Produktion und Rezeption in der erforderlichen Geschwindigkeit. Die Voraussetzungen für den Wortschatzerwerb durch das Kind sind im Wesentlichen dieselben wie die für den Spracherwerb generell (s. die in Kap. 3.2 genannten biologischen, kognitiven und sozialen Umstände). Unter deren Einfluss und im Zusammenspiel mit deren Entwicklung verläuft der Wortschatzerwerb in zwei Hauptphasen.

Phasen des Wortschatzerwerbs: Da der Übergang von der frühen Phase zur darauf folgenden zeitlich mit der Entwicklung syntaktischen Wissens einhergeht, hat man die beiden Phasen als **vorsyntaktische** und **syntaktische** Phase des Wortschatzerwerbs bezeichnet. Erstere liegt ungefähr in der ersten Hälfte des zweiten Lebensjahres. Angesichts der erheblichen Variation im zeitlichen Ablauf des Spracherwerbs von Kind zu Kind erfasst man sie besser anhand der spezifischen kognitiven Voraussetzungen für ihren Beginn, ihre Verlaufsmerkmale und ihr Ergebnis.

Aktives vs. passives Wortwissen: Um mit dem **Aufbau lexikalischen Wissens** beginnen zu können, muss das Kind im Wesentlichen bestimmte Wahrnehmungs-, Gedächtnis- und Wiedergabeprozesse entwickelt haben. Entscheidend sind die Gedächtnisfunktionen. Das Kind muss ein in seiner Umgebung wiederkehrendes lautliches Gebilde so gespeichert haben, dass es dieses in wechselnden Ausprägungen erkennt, mit einer konstanten nichtlautlichen Vorstellung von seiner Verwendungsbedingung (Situation, Ereignistyp, Sache, Situationsmerkmalen) assoziiert und schließlich auch artikulieren kann. Soweit es Letzteres nicht kann, ist die lexikalische Einheit eben noch nicht aktiv, sondern nur passiv beherrscht. Alle Beobachtungen sprechen dafür, dass die passive Beherrschung der aktiven zeitlich deutlich vorausgeht.

Die vorsyntaktische Phase

Der **vorsyntaktische Wortschatzerwerb** verläuft langsam, jedenfalls verglichen mit dem der folgenden Periode. In dem halben Jahr seiner Dauer werden zwischen 30 und 50 Wörter erworben, im Wesentlichen etwa gleichmäßig zwei bis drei pro Woche. Zum Inhalt dessen, was vom Kind eigentlich als lexikalische Einheit anfangs gespeichert wird, ist zurzeit wenig bekannt. Folgt man den Produktionsdaten, dann stimmt die gespeicherte Lautform eines Wortes nicht von Anfang an mit der Zielform überein. Sie ist infolge der begrenzten Arbeitsweise des Lauterwerbs gewissermaßen reduziert. Unbetonte Silben sind nicht gespeichert, Konsonanten sind nach Zahl, Art und Ort abgewandelt und Vokale geöffnet, entrundet oder ersetzt: [Ova] < Eule >, [bɛlə] < Brille >, [aɪz] < eins >, [brao = nicht hoher, runder Hinterzungenvokal] < Frau >. Die Beispiele sind Elsen (1991, S. 58–59) entnommen.

Überdehnung und Unterdehnung: Die mit dem Lautgebilde assoziierte Vorstellung ist mit der Zielbedeutung nicht gleichzusetzen. Mit einem Namen kann sich mehr als die Vorstellung der Person verbinden, die ihn

trägt, z. B. eine typische Handlung, eine typische Situation; mit einem Wort wie *heiß* kann sich die Erfahrung des Wehtuns verbinden, das Verbot, heiße Ding zu berühren oder der Anblick einer offenen Feuerstelle. Man spricht hier allgemein von **Überdehnung**, einem Spezialfall von Übergeneralisierung; das Umgekehrte ist häufiger: die **Unterdehnung**; sie ist allerdings auch schwerer nachzuweisen. Alle dahin gehenden Annahmen stützen sich, auf Produktionsdaten. Sind allerdings die Phänomene gar nicht auf das zu Grunde liegende Wissen zurückzuführen, sondern eben auf Produktionsvorgänge, dann wäre nicht auszuschließen, dass die lexikalische Information schon deutlich früher zielsprachlich ist (vgl. Höhle/Weissenborn 1999; Weissenborn 2000). Hall (2009) präsentiert experimentelle Daten, die dafür sprechen, dass das Kind beim Wortschatzerwerb von Anfang an zwischen Eigennamen und Klassennamen unterscheiden kann.

Zum Begriff

Die **vorsyntaktische Phase** des Wortschatzerwerbs ist gekennzeichnet von einem relativ langsamen Erwerb von Wörtern, wobei die Aufmerksamkeit des Kindes auf die Wortbedeutung und weniger auf die morphosyntaktischen Eigenschaften der gelernten Wörter gerichtet ist. Sie erstreckt sich zeitlich bis zum Alter von ca. 18 Monaten.

Frühe Syntax

Syntaktische Kategorien: Inwieweit die frühen Lexikoneinheiten des Kindes neben der lautlichen und der semantischen auch mit syntaktischer Information verknüpft sind, ist allerdings weiterhin fraglich. Formale Hinweise wie flexionsmorphologische Markierungen gibt es in einigen Sprachen, in anderen nicht, ebenso wenig Kongruenzphänomene, die es in Einwortsätzen eben nicht geben kann. Das Fehlen äußerer Anzeichen schließt allerdings die Existenz syntaktischer Informationen im frühen Lexikon nicht aus. Wie oben erwähnt, gibt es für die Annahme des passiven lexikalischen Wissens auch keine sprachliche Evidenz, die Verstehensreaktionen des Kindes ausgenommen. Aber gerade diese Reaktionen, experimentell überprüft, weisen auf die Kenntnis syntaktischer Phänomene im frühen Lexikon hin. So haben Weissenborn und Mitarbeiter in *head-turn-preference*-Experimenten gezeigt, dass Kinder gegen Ende der ersten Hälfte des zweiten Jahres syntaktische Unterschiede in Sätzen erkennen, die durch Funktionswörter angezeigt sind, z. B. den Unterschied zwischen Haupt- und Nebensatz und den Zusammenhang mit dem Vorhandensein bzw. Fehlen von *dass* (vgl. Weissenborn/Höhle/Kiefer/Čavar 1998; und allgemein Höhle/Weissenborn 1999, S. 54 ff.). Man mag aufgrund von Produktionsdaten die Existenz von Wortartinformationen im lexikalischen Wissen des noch nicht zweijährigen Kindes bezweifeln. Man kann aber andererseits nicht bestreiten, dass eine Sensitivität für *dass* und seine Strukturbedingungen im Satz vorhanden ist.

Eine Annahme, die beiden Positionen gerecht wird, geht davon aus, dass syntaktische Informationen letztlich Bündel von einzelnen syntaktischen Merkmalen sind und dass es wahrscheinlich ist, dass diese Merk-

male vom Kind nicht bündelweise, sondern, wie lautliche und semantische auch, isoliert erkannt und gespeichert werden können. Ob man nun, wenn das Kind offenbar auf Distributionseigenschaften von *dass* reagiert, schon die Existenz einer Wortartinformation, z. B. Complementizer, postulieren kann oder nicht, ist eine andere Frage als die nach syntaktischer Information im frühen Lexikon. Dazu muss man wissen, dass *dass* vom Kind erst relativ spät (ab etwa dem dritten bis vierten Lebensjahr) aktiv verwendet wird. Produktionsdaten aus der frühen Lexikonphase liefern, wie gesagt, keine direkten Hinweise auf die Existenz dieses syntaktischen Wissens. Bloom (1993) zeigt allerdings, dass Einwortäußerungen des Kindes in Antworten auf Fragen der Eltern syntaktisch fast immer dem Fragekontext entsprechen, was jedoch auch durch die Annahme semantischer Typen anstelle von syntaktischen Wortklassen zu erklären ist. Nomenähnliche Wörter werden hauptsächlich mit Bezug auf Objekte und Lebewesen im aktuellen Aufmerksamkeitsbereich verwendet, verhältnismäßig unter Bezug auf Umstände, in die die Objekte/Lebewesen verwickelt sind.

Die syntaktische Phase

Während der **syntaktischen Phase des Wortschatzerwerbs**, also mit Beginn der zweiten Hälfte des zweiten Jahres, setzt sich eine Tendenz stark fort, die auch während der vorsyntaktischen Phase beobachtet werden kann, nämlich die Überrepräsentation nominaler gegenüber verbalen Einheiten im semantischen Sinn (vgl. Gentner 1982). In neueren Betrachtungen wird zwar eine deutlich differenziertere, im Prinzip aber nicht andere Faktenlage sichtbar (vgl. Kauschke 2007).

Feststellungen zur Überrepräsentation nominaler gegenüber verbaler Einheiten (nach Kauschke (2007):

- In der Rede von Erwachsenen zu Kindern überwiegen in vielen Sprachen die Verben.
- Unter den gleichen Bedingungen überwiegen innerhalb der Verbtypes die transitiven gegenüber den intransitiven; vgl. auch Childers/Tomasello/Hirsh-Pasek/Golinkoff (2006), die spezifisch die Rolle der Argumentstruktur untersuchen.
- Es gibt eine Interaktion zwischen Wortklasse und Aufgabe; die Verb-/Nomen-Differenz ist beim Benennen stärker ausgeprägt als beim Erkennen.
- Bei Kindern gibt es einen Erwerbsfortschritt; die Benennungsgenauigkeit nimmt von 2;6 bis 8 zunächst zu, flacht dann aber ab. Die Nomina werden besser verarbeitet als die Verben.
- Genauigkeit und Geschwindigkeit sind beim Benennen mit Nomina besser als beim Benennen mit Verben, bei Kindern wie bei Erwachsenen.
- Auch beim Verstehen gibt es eine Interaktion zwischen Aufgabe und Wortart. Die Differenz zwischen Verstehen/Produzieren ist für Verben größer als für Nomina bei durchgehender Nomenüberlegenheit.
- Es gibt eine Interaktion zwischen Sprache und Aufgabe; die Nomen-

Zur Vertiefung

überlegenheit ist beim Benennen im Türkischen schwächer als im Deutschen und etwa gleich dem Koreanischen, beim Verstehen aber etwa so groß wie im Deutschen und größer als im Koreanischen.

- Im typologischen Vergleich bestätigt sich die Annahme einer universalen Nomenüberlegenheit mit einer Reihung von Deutsch (am stärksten nominal) über Englisch, Koreanisch, Türkisch.
- Globale, Wernicke- und Broca-Aphasiker bewahren die Nomenüberlegenheit im Deutschen.
- Es gibt einen Krankheitstyp-Effekt; bei Agrammatikern ist der Rückgang des Verbanteils größer als der Rückgang des Nomenanteils; bei anomischen Aphasikern ist es umgekehrt.

Zum Begriff

In der **syntaktischen Phase** des Wortschatzerwerbs, die um das Alter von 18 Monaten beginnt und durch das Auftreten von Mehrwortsätzen gekennzeichnet ist, richtet sich die Aufmerksamkeit des Kindes verstärkt auf die syntaktischen Eigenschaften von Wörtern.

Zweiwortsätze: Der Beginn der zweiten Phase des Wortschatzerwerbs fällt zeitlich etwa mit dem 18. Monat zusammen, der Zeitpunkt, zu dem das Auftreten der ersten **Zweiwortsätze** zu beobachten ist. Unter Zweiwortsatz ist nicht ein syntaktisch und morphologisch wohlgeformter Satz wie *Blumen erblühten* zu verstehen, sondern die – allerdings systematische – Zusammenfügung von zwei lexikalischen Einheiten zu einer intonatorischen Gruppe und Bedeutungseinheit wie *Jan Bett*.

Der Wortschatzspurt: In der zweiten Phase beschleunigt sich der Erwerb des Wortschatzes so beträchtlich, dass sich dafür die Bezeichnung Wortschatzspurt (*vocabulary spurt*) eingebürgert hat (vgl. Rothweiler 1999, S. 16). Die Erwerbsrate steigt auf mehrere Wörter pro Tag und am Ende dieses halben Jahres hat sich der Umfang des aktiven lexikalischen Wissens etwa vervierfacht, am Ende des folgenden Halbjahres versechsfacht. Die Entwicklung der Nomendominanz und der Anteile der übrigen Klassen von Wörtern zeigen für das Englische Bates/Marchman/Thal/Fenson/Dale/Reznick/Reilly/Hartung (1994).

Zum Begriff

Wortschatzspurt bezeichnet eine Phase des Wortschatzerwerbs. Sie beginnt etwa im Alter von 18 Monaten. Kennzeichnend ist die starke Zunahme des Erwerbs von neuen Wörtern durch das Kind von rund 30 bis 50 Wörtern in den ersten 18 Monaten auf mehrere Wörter pro Tag.

Fast mapping: Das rasante Tempo des Wortschatzspurts ermöglicht sich das Kind offenbar durch Anwendung einer Aufnahme-strategie, die sich als »schnell aber unvollständig« etikettieren lässt – in der englischsprachigen Literatur als *fast mapping* bezeichnet. Sie wird bis zum Ende des Vor-

schulalters beibehalten und verkürzt den Aufnahmevorgang dadurch, dass eine (unvollständige) Lautinformation (z. B. eine bestimmte Lautform einer lexikalischen Einheit) mit einer ebenfalls unvollständigen Bedeutungsvorstellung assoziiert und gespeichert wird. Das geschieht im Wesentlichen aus Schlussfolgerungen aus Vergleichen und Kontrasten des neuen Wortes angesichts seiner Verwendung auf Dinge und Eigenschaften, die das Kind bis dahin mit einem semantisch weiteren Wort bezeichnet hat; Beispiel: *rot* (alt) *pink* (neu). Dass dieses Verfahren von Kindern nicht nur, nicht einmal bevorzugt beim Erwerb von Farbwörtern eingesetzt wird, zeigen u. a. Heibeck/Markman (1987). Eine aktuelle differenzierte Übersicht über den Forschungsstand zu dem Konzept vermitteln Kucker/McMurray/Samuelson (2015). Dass *fast mapping* sich auch neuropsychologisch nachweisen lässt, zeigen Merhav/Karni/Gilboa (2015).

Insgesamt ermöglicht das *fast mapping* dem Kind, viele neue Wörter aufzunehmen, ohne auf erschöpfende Information zu ihren Bedeutungen zu warten. Auch dem in diesem Alter noch begrenzten Aufnahmevermögen des Gedächtnisses kommt das Vorgehen natürlich entgegen. Der Wortschatz des Kindes erweitert sich außer durch die Aufnahme neuer Simplizia auch dadurch, dass das Kind neue Wörter bildet. Dies geschieht schon ab dem dritten Jahr, zunächst bevorzugt durch Zusammensetzung, später zusätzlich durch Derivation. Wie viele lexikalische Einheiten das Lexikon des Kindes am Ende der Vorschulzeit umfasst, ist verständlicherweise nur vage anzugeben. Aitchison (1997, S. 221) nennt bis zu 3000 Wörter als aktiven Bestand von Fünfjährigen und bis zu 14.000 als passiven von Sechsjährigen.

Im Wortschatzerwerb bezeichnet **fast mapping** die Anwendung einer speziellen Aufnahmestrategie, die darin besteht, teilweise unvollständige lautliche Informationen mit teilweise unvollständigen Bedeutungsvorstellungen zu assoziieren und im mentalen Lexikon zu speichern.

Zum Begriff

3.4.3 | Der Erwerb von Morphologie und Syntax

Die Entwicklungen im Aufbau morphologischen und syntaktischen Wissens werden hier in einem Abschnitt zusammen dargestellt und zwar am Beispiel des Deutschen. Auf sprachspezifische Entwicklungen wird hier nicht eingegangen. Natürlich ist der Erwerb von Flexion und Syntax immer der Erwerb von Ausdrucksmitteln in einer spezifischen Sprache, hier eben des Deutschen. Zur Information über Entwicklungen in anderen Sprachen konsultiere man daher unbedingt Slobin (1985, Vol. 1). Jenseits aller Einzelsprachlichkeit vermitteln die Daten aber auch eine Ahnung übereinzelsprachlicher Reihenfolgen, z. B. hinsichtlich der wachsenden Länge der Äußerungen, dem Auftreten von Wortarten und von elementaren syntaktischen Regularitäten wie Kongruenz und Wortstellung. Snyder (2007, Kap. 4) weist in diesem Zusammenhang auf eine Parallelität zwischen der syntaktischen und einer analogen Entwicklung im Wortschatz hin (s. Kap. 3.6).

Morphosyntax

Wörter pro Äußerung: Das auffallendste äußerliche Merkmal der Entwicklung syntaktischen Wissens durch das Kind ist, dass seine Äußerungen im Lauf der Zeit von einem Jahr bis 6 Jahre länger werden, indem sie zunehmend mehr Wörter enthalten. Die Länge der Äußerung wird durch Mittlung der Zahl der Wörter aus den Äußerungen des zu charakterisierenden Zeitintervalls bestimmt und als MLU (*mean length of utterance*) häufig zur Bezeichnung von Stadien des Spracherwerbs verwendet. MLU 2.5 bedeutet also, dass die Äußerungen in dem betrachteten Intervall durchschnittlich 2,5 Wörter lang sind.

Zum Begriff

Mean length of utterance (MLU) ist ein Index für die durchschnittliche Anzahl der Wörter pro Äußerung in einem vordefinierten Zeitintervall.

Wie fast alles im Spracherwerb entwickelt sich auch die Länge der Äußerungen relativ stetig in der Zeit. Mit der Zunahme in der Länge der Äußerung steht das Kind zugleich vor der Aufgabe, sprachlich zu erkennen zu geben, wie die Wörter strukturell miteinander zusammenhängen, und diese Aufgabe verlangt Kenntnisse unterschiedlicher syntaktischer Verknüpfungen. Ein weniger differenzierender, für die Abgrenzung von Erwerbsphasen (engl. *stages*), aber nützlicherer Index, ist der PLU-Index (*predominant length of utterance*).

Zur Vertiefung

Definitionen der PLU-Stufen (nach Legendre/Hagstrom/Vainikka/Todorova 2002)

Stufe 1: Stufe vorherrschender Einwortäußerungen

- Fast alle Äußerungen (90 %) sind vom Einwort-Satztyp.

Stufe 2: Zwischenstufe zwischen Einwort- und Zweitwortäußerungen

- Der Einwort-Satztyp ist immer noch sehr verbreitet (60 %–89 % aller Äußerungen sind von Einwort-Satztyp).

Stufe 3: Zweitwort-Stufe

- Mehr als 40 % der Äußerungen beinhalten mehr als ein Wort.
- Trotzdem tendieren Äußerungen dazu, sehr kurz zu sein; Einwort- und Zweiwortsätze überwiegen im Vergleich zu Mehrwortäußerungen.

Stufe 4: Stufe vorherrschender Mehrwortäußerungen

- Im Vergleich aller drei Satztypen überwiegen die Mehrwortäußerungen.

Definitionen der zweitrangigen PLU-Stufen

Stufe a: Maximal 10 % aller Äußerungen enthält ein Verb.

Stufe b: 11 %–60 % aller Äußerungen enthält ein Verb.

Stufe c: Mehr als 60 % aller Äußerungen enthält ein Verb.

Stationen beim Erwerb der Syntax und Morphologie

Einwortphase: Wie schon im Abschnitt über den Wortschatzerwerb ausgeführt, sind gegen Ende des ersten Jahres und über die Zeit des vorsyntaktischen Wortschatzerwerbs Einwortäußerungen des Kindes zu beobachten, zum Beispiel *wauwau, mieze, weg, auf, mehr, auch, da, nein* (Belege aus Stern/Stern 1920; Miller 1976).

Ausgedrückt werden mit diesen Äußerungen meist Wünsche und Feststellungen. Sprachlich ausgedrückt ist zumeist der Bezug auf Dinge oder Personen, Orte oder Zielorte, seltener Umstände von Aktionen und die Verneinung. Direkte Bezeichnungen von Vorgängen durch Verben sind selten. Sofern sie belegt sind, weist das Verb, soweit erkennbar, die Infinitivform auf, die Nomina ebenfalls eine nominativähnliche Singular-Grundform, Adjektive die prädikative, unflektierte Form, wie etwa *heiß*. Mehrwortäußerungen bilden in dieser Zeit die Ausnahme und scheinen vom Kind als feste zusammenhängende Formeln gespeichert und eingesetzt zu sein. Syntaxwissen im engeren Sinne lässt sich den Äußerungen des Kindes nicht entnehmen, weil eben keine Verknüpfungen stattfinden. Wie schon im Zusammenhang mit dem Lexikonerwerb ist auch hier ausdrücklich zu erwähnen, dass zwischen produziertem und vorhandenem Wissen unterschieden werden muss (vgl. Höhle/Weissenborn 1999).

Zweiwortphase: Auf die Einwortphase des Syntaxerwerbs folgt die Zweiwortphase, die etwa die zweite Hälfte des zweiten Lebensjahres dauert (1;6 bis 2). Gekennzeichnet ist diese Phase durch folgende Beobachtungen (Ka = Kaltenbacher, 1990; El = Elsen, 1999):

- Ausgedrückt werden weiterhin **Feststellungen**, dass etwas so oder so ist, zunehmend nun aber auch **Hinweise und Kommentare**, dass sich nun etwas geändert hat.
- Nomina sind zahlenmäßig immer noch dominant, es treten, wie vereinzelt schon in der Einwortphase, **formelhafte Pluralformen** auf, wie z. B. *schuhe putt*. (Ka: 1;10).
- Die nominale Gruppe weist nur **selten Artikel** auf, wenn, dann in einer unflektierten Grundform den definiten und zwar nur im Singular, wie z. B. *de uwe auch ...* [... hat eine Brille] (Ka: 1;8).
- Mills (1985) berichtet über **Adjektivbelege in attributiver Position**, *großes Loch*, die schon Ansätze von Genusflexion aufweisen.
- In Nomen-Verb-Äußerungen stehen – agentivisch wie patientivisch – **Nomen voran**: *mama esse*. [Aufforderung, Mama soll ...]. Ka: 1;7 oder *schuhe auszieh*. [Feststellung] (vgl. Kaltenbacher 1990, S. 80).
- Ebenfalls in dieser Phase erscheinen **Partizipialformen des Verbs**, allerdings ohne temporale Bedeutung. Fragen werden mit Fragewort, am häufigsten zunächst *was*, ausgedrückt, aber auch weiterhin intonatorisch, Verneinung mit dem holophrastischen *nein*, z. B., *nein Brille*. [zu einem Mann, der keine Brille trägt]. (El: 1;3).
- Die holophrastische Form wird noch in der Zweiwortphase durch Auftauchen von *nicht* abgelöst, z. B., *naIn bot niç*. <nein. Brot nicht>. [Sie will das Brot nicht] El: 1;6; (vgl. Elsen 1999, S. 87). Die syntaktischen Ausdrucksmittel des Kindes beschränken sich in der Zweiwortphase auf die Wahl der Wortarten der beteiligten Wörter und deren Stel-

Die letzte Phase

lung in der Äußerung. Eine rein auf Distribution gerichtete Analyse englischer Daten hat Braine (1963) vorgenommen und aus ihr abgeleitet, das Kind verfüge in dieser Zeit über zwei Hauptwortklassen (*open class words* und *pivot words*) und die entsprechenden Kombinationsregeln. Mit zunehmender Datendichte ist allerdings zweifelhaft geworden, ob die Pivotphase einen eigenen abgrenzbaren Abschnitt in der Syntaxentwicklung darstellt. Die mit den Mitteln von Zweiwortsätzen ausdrückbaren Sprechhandlungen und Bedeutungsstrukturen sind sehr vielfältig (vgl. Kaltenbacher 1990; Elsen 1999).

Die **Drei- und Mehrwortphase** folgt der Zweiwortphase und sie dauert etwa von 2 bis 4 Jahre. Beobachtungen zeigen (Tr = Tracy 1991):

- innerhalb der nominalen Gruppe treten **häufiger Artikelformen** auf (auch indefinite, nach Genus korrekt flektiert).
- In maskulinen NPs erscheint die **Nominativ- und Akkusativmarkierung, Dativ** ab dem dritten Jahr.
- **Personalpronomen** der ersten und zweiten Person Singular sind in der zweiten Hälfte des dritten Jahres belegt, darunter auch deklinierte.
- **Subjekt-Verb-Kongruenz** ist durch Person- und Numerusmarkierung am Verb ausgedrückt. Daneben stehen auch noch unflektierte Sätze wie: *julia eis essen* (Tr: 2;0).

Die **finite Verbform** nimmt die Zweitposition im Satz ein, eine Entwicklung, die im Zusammenhang mit der Theoriediskussion viel Beachtung gefunden hat (s. Kap. 3.5). Zum Beispiel: *da falln jetzt i blätter runter* (Tr 2;2); *das weint* (Tr 2.0).

Fintheit: Im dritten Jahr ist die Vergangenheitsflexion des Verbs zunehmend belegt, die Formen unregelmäßiger Verben werden allerdings vielfach in Übergeneralisierung der regelmäßigen Konjugation schwach gebildet, *gegeht, stehlten*. Belegt ist auch ein Präsens *lauft* als 3. Person Singular. Mit Auftreten des Frageworts in der Initialposition ist die Inversion belegt: Die Subjekts-NP steht hinter dem finiten Verbs, wie z. B. in *Was macht die Frau?* (El: 2;0).

Letzte Schritte und Stabilisierung: Neben Aktivsätzen treten gegen Ende dieser Phase Passivkonstruktionen auf; nach Mills (1985) fast durchweg ohne Nennung des Agens-Referenten. Es treten die ersten komplexen Sätze auf; nach Mills (1985) zunächst Relativsätze (vgl. allerdings das reichhaltige Repertoire von Konditionalsätzen in Elsen 1999, S. 145 ff.). In der Folgezeit stabilisieren sich die zuvor kritischen Formen, die unregelmäßigen Flexionsformen von Verben lösen die übergeneralisierten schwachen Formen ab, komplexe Verbformen mit Auxiliar – zunächst mehr *haben* als *sein* – und Modalverben treten auf. Der Formenreichtum der Artikelwörter nimmt zu (Possessiv-Artikel), ebenso das Repertoire anderer Funktionswörter, besonders der Präpositionen und der subordinierenden Konjunktionen. Die Wortstellung in komplexen Nebensätzen weist bis ins Alter von fünf Jahren Abweichungen auf, ebenso vor- und nachzeitige Tempusformen des Verbs. Bei der Beurteilung dieser Phänomene ist allerdings zu berücksichtigen, dass sie auch in der Erwachsenensprache belegt sind.

3.5 | Erklärungsansätze

Leitende Fragen: Theoretische Aussagen über zusammenhängende Phänomene in einem Teilbereich unserer Erfahrungswelt gehen über die bloße Beschreibung der Phänomene hinaus. Sie machen Behauptungen darüber, wie es sich **erklärt**, dass die Phänomene eben gerade so sind, wie sie sind, und nicht auf eine andere denkbare Weise.

Die besondere Aufmerksamkeit der Wissenschaftler haben **drei fundamentale Fragen** erregt:

- Was treibt den Säugling, sprachliches Wissen und die Verwendungsmechanismen zur sprachlichen Kommunikation zu erwerben?
- Was befähigt ausschließlich den Menschen unter den Lebewesen zum Erwerb einer natürlichen Sprache?
- Wie erklärt sich der grosso modo einheitliche Verlauf des Spracherwerbs?

Bei der Suche nach Antworten auf diese Fragen wird plausiblerweise an den spezifischen Bedingungen angesetzt, die die Entwicklung der Sprachbeherrschung des Menschen möglich machen, das heißt an den in Kapitel 3.2 genannten Voraussetzungen. Das ist insofern plausibel, als Erklärungen, die mit den sozialen, biologischen und kognitiven Erwerbsbedingungen nicht im Einklang stehen, eben im Widerspruch zu ihnen stehen und eine Theorie, die Tatsachen widerspricht, hat einen problematischen Geburtsfehler.

Voraussetzungen

Was heißt es aber nun, mit der Suche nach Erklärungen an den Voraussetzungen der zu erklärenden Phänomene anzusetzen? Inwiefern lässt sich aus der Kenntnis der sozialen Interaktion des Kindes, seiner biologischen Ausstattung und seiner kognitiven Kapazität vorhersagen, dass nur der Mensch Sprachbeherrschung entwickelt und dass die Entwicklung auf dem beschriebenen Weg verläuft? Diese Frage mutet vielleicht sonderbar an. Zweifellos, wird man einwenden, kann man aus der biologischen Voraussetzung des Menschen vorhersagen, dass ein Neugeborenes sich sein Leben nicht unter Wasser einrichten wird, sondern auf dem Land, weil es nicht die Organe entwickeln wird, die für eine maritime Lebensweise erforderlich sind. Dem ist nicht zu widersprechen, denn die Voraussetzung, aus der die Vorhersage abgeleitet ist, ist im Einzelnen gut bekannt. Und das ist ein Umstand, der für die Haltbarkeit der Vorhersage entscheidend ist. Die Frage ist also, wie genau uns die Voraussetzungen für die Ableitung einer Spracherwerbstheorie bekannt sind, die Voraussetzungen, die den Spracherwerb von Anbeginn ermöglichen und im weiteren Wachstum des Kindes prägen.

Soziale Ausstattung: Man stelle sich folgende Begebenheit vor: Ein Säugling im Alter von ca. 18 Monaten sitzt in einem Raum, z. B. auf dem Schoß seiner Mutter. Eine externe Person, sagen wir E1, betritt das Zimmer mit einem Buch in der Hand. Sie schaut sich um, nimmt Blickkontakt mit dem Baby auf, wendet sich dann aber ab, legt das Buch auf einen niedrigen Tisch im Blickfeld des Kindes und verlässt den Raum wieder. Kurz darauf tritt eine weitere Person, E2, herein. Sie hat eine Zeitung in der

Hand; die legt sie auf das Buch, das damit verdeckt ist. Dann tritt E1 wieder ein, zögert und schaut sich suchend um, unter anderem blickt sie wieder das Kind an. Das schaut zurück, dann zu dem Tisch und zeigt mit dem Finger Richtung Zeitung. E1 versteht, hebt die Zeitung an und findet das Buch.

Im Rahmen dieses sog. Verstecken-und-Finden-Paradigmas wurden von Wissenschaftlern in der Gruppe um Michael Tomasello in großer Zahl mit wechselnden Ausgestaltungen Experimente durchgeführt; das oben skizzierte Setting ist dem in Liszkowski/Carpenter/Striano/Tomasello (2006) beschriebenen sehr ähnlich. In einigen wurden Kinder unterschiedlichen Alters beobachtet, u. a. solche im Alter von 12 Monaten im Vergleich mit Eineinhalbjährigen. Dass Kinder im Säuglingsalter mit Zeigegesten kommunizieren, ist schon seit Langem bekannt und seit rund 40 Jahren experimentell untersucht worden (vgl. Bates/Camaioni/Volterra 1975; Lempers 1979). Im hier gegebenen Zusammenhang sind drei Befunde bedeutsam.

- Das kommunikative Verwenden von Zeigegesten des Kleinkindes unterscheidet sich wesentlich von dem von Menschenaffen generell, indem es Mittel **kooperativer Kommunikation** ist, das von Primaten hingegen egozentrisch; ausführlich und entwicklungsgeschichtlich wird dies beschrieben in Tomasello et al. (2005).
- Kinder verwenden Zeigegesten zu zwei verschiedenen Aktivitäten: zum **Auffordern** (*imperative*) und zum **Erklären** (*declarative*). Nach den Befunden von Camaioni/Perucchini/Bellagamba/Colonnesi (2004) tritt letztere deutlich nach der ersten Funktion auf; so auch Liebal/Behne/Carpenter/Tomasello (2009).
- Zum dritten zeigen beide Verhaltensweisen, dass das Kind ab dem Alter von ca. vierzehn Monaten die Fähigkeit hat, **kommunikativ** mit einem Erwachsenen zu **interagieren** und – im Alter von etwa 18 Monaten – die Aufmerksamkeit des Erwachsenen auf einen von beiden Beteiligten als relevant erachteten Sachverhalt zu lenken. Insbesondere in dieser Fähigkeit wird übereinstimmend in der Forschung der Vorläufer kooperativer sprachlicher Kommunikation gesehen.

Biologische Ausstattung: Welche körperlichen Eigenschaften und Fertigkeiten erklären die Möglichkeit des Erwerbs der Sprachfähigkeit des Menschen? Phylogenetisch ist zum Nachweis dieser Voraussetzungen sehr weit zurückzuschauen, nämlich etwa sechs Millionen Jahre. Das ist nach paläoanthropologischer Schätzung das Erdzeitalter, in dem die Bewohner der Region in und westlich von Äthiopien sich vom Vierbeiner zum Zweibeiner und damit zum aufrechten Gang hin entwickelt haben. Damit war die Voraussetzung für die Entwicklung des sog. Ansatzrohres der Hominiiden gegeben und damit die Fähigkeit einer Dosierung des Ausatemstroms, wie sie der Gattung Mensch eigen und für die Produktion einer gegliederten sprachlichen Äußerung erforderlich ist (vgl. Schrenk 2008). Damit verbunden ist die ontogenetische Entwicklung des Rachenraumes des Menschen im ersten Lebensjahr, wie oben beschrieben; d. h. die Öffnung der Mundhöhle durch Wölbung des Gaumens und die Absenkung des Kehlkopfes im Lauf des ersten Lebensjahres. So lässt sich vorhersa-

gen, dass hintere Vokale später als vordere erworben werden und der vordere Konsonantismus früher als der pharyngale (s. Kap. 3.4.1). Bei letzterem ist zudem erklärend die Tatsache relevant, dass die Bildung vorderer Konsonanten visueller Wahrnehmung eher zugänglich ist, als die der hinteren.

Kognitive Ausstattung: Die wissenschaftliche Diskussion darüber, welche angeborene kognitive Ausstattung der Sprachfähigkeit des Menschen zugrunde liegt, ist kontrovers und zwar seit etwa einem halben Jahrhundert. Wie erklärt es sich, dass kein anderes Lebewesen als der Mensch ein so reichhaltiges und vernetztes lexikalisches Wissen und dermaßen differenzierte strukturelle Regelmäßigkeiten erwerben kann – und das in einem sonst so unausgereiften Organismus und in so kurzer Zeit? Und ohne eine systematische explizite Unterweisung! Als sicher ist anzunehmen, dass das leistungsfähige Gedächtnis, die damit operierende Fähigkeit der Begriffsbildung und Strukturerkennung für die Entwicklung des sprachlichen Wissens wesentlich sind. Viele Einzelheiten, auch wesentliche, sind aber nur über die Beobachtung der Ergebnisse der kognitiven Aktivität zugänglich, das heißt durch Interpretation der Denk- und Sprachäußerungen des Kindes im Lauf des Spracherwerbs. Das sprachliche Verhalten des Kindes bildet also das Fenster, durch das wir einen Blick auf Einzelheiten der kognitiven Ausstattung werfen, die das Kind bei der Geburt eben für die Entwicklung desselben mitbringt.

Zwei verschiedene Perspektiven: In der neueren Geschichte der Spracherwerbsforschung, im deutschsprachigen Raum also etwa von Beginn des 20. Jahrhunderts (Stern/Stern 1909) bis heute, wurde den genannten Umständen und weiteren mehr ein verschieden hoher Erklärungswert zugemessen. Bei aller Vielfalt im Einzelnen, konvergieren die Modelle zu zwei im Ansatz verschiedenen Sichtweisen. Die eine geht von linguistischen Eigenschaften natürlicher Sprachen aus, die andere von den Herausforderungen, in einer gegebenen Situation mit der anzunehmenden kognitiven Fähigkeit des Kindes sprachlich Sachverhalte und Intentionen zu kommunizieren. In der Fachliteratur hat sich für die erstere die Bezeichnung »**Nativismus**«, für die zweitgenannte »**Sprachgebrauchsmodell**« (*usage based theory*) etabliert.

3.5.1 | Das Nativismus-Modell

Die Grundannahme der **nativistischen Sprachtheorie** besagt: Der Mensch ist genetisch mit einem Sprachorgan ausgestattet und darin unterscheidet er sich von allen anderen Lebewesen.

Was ist genetisch?

Die Kernbehauptung der nativistischen Sprachtheorie, der Mensch sei genetisch mit einem Sprachorgan ausgestattet, lässt natürlich sofort Fragen und Zweifel entstehen. Spezialliteratur über den Spracherwerb ist außer Chomsky (1980) besonders die umfassende Darstellung in Pinker (1994) und die einschlägigen Teile in dem *Handbook of Child Language* (Fletcher/MacWhinney 1995).

Das Sprachorgan: Was hat man unter dem oben postulierten Sprachorgan zu verstehen? Offensichtlich ist es kein chirurgisch identifizierbares,

Evidenz abgegrenztes Stück spezialisierten Gewebes mit einer einheitlichen, komplexen Funktion, eben der, die Sprachfähigkeit zu beherbergen. Man hat es sich vielmehr als ein genetisch verankertes und neurophysiologisch repräsentiertes Informationssystem vorzustellen, ein spezielles Wissenssystem. Es ist dem Bewusstsein nicht zugänglich, ebenso wenig wie die Fähigkeit, die dem Menschen das räumliche Sehen ermöglicht. Es ist **universal** in dem Sinne, dass es die Gliederungseigenschaften spezifiziert, die allen und genau den natürlichen Sprachen gemeinsam sind. Es ist **modular**; das heißt, dass es als Ganzes mit dem Denken oder dem Artikulieren interagiert. Es steht dem Kind von Anbeginn des Spracherwerbs an zur Verfügung, und es prägt im Zusammenspiel mit den sich entwickelnden Wahrnehmungs- und Denkfähigkeiten des Kindes den Verlauf des Spracherwerbs.

Die vier wichtigsten Argumente: Betrachten wir die Behauptungen dieses Modells etwas genauer, zunächst die Argumentation dafür, dass ein solches Modul überhaupt existiert. Direkte Evidenz in dem Sinne, dass im Zentralnervensystem ein abgegrenztes Teilsystem von neuronalen Zellen, z. B. in der Großhirnrinde lokal mit klinischen Verfahren zu bestimmen ist, liegt nicht vor. Die Annahme der Existenz des universalen Sprachprogramms von Geburt an stützt sich auf Schlussfolgerungen aus verschiedenen Beobachtungen, die, so die Argumentation, nicht anders als durch die genannte Annahme zu erklären sind. Es sind im Wesentlichen Spracherwerbsbeobachtungen und neuerdings experimentelle Befunde aus Verhaltensexperimenten mit Kleinkindern.

1. Unterdeterminiertheit der Grammatik: Dafür, dass der Spracherwerb von Anbeginn durch Strukturprinzipien geleitet ist, wird angeführt, dass in den **Äußerungen des Kindes** Formen nicht belegt sind, die aber auf Grund der Äußerungen, die das Kind hört, theoretisch erwartbar wären. Ein Beispiel stellt die Bildung von Verb-Erst-Fragen dar.

- (3-1) *Die Puppe liegt im Wagen.*
- (3-2) *Liegt die Puppe im Wagen?*
- (3-3) *Die Puppe, die kaputt ist, liegt im Wagen.*
- (3-4) *Liegt die Puppe, die kaputt ist, – im Wagen?*
- (3-5) **Ist die Puppe, die kaputt __, liegt im Wagen?*

Würde die Regel für die Bildung der Verb-Erst-Frage nach dem einfachen, linearen Muster gebildet, so dass das erste Verb nach der Nominalphrase in der Frage dieser voranzustellen ist, wären Sätze wie (3-5) zu erwarten. Sie sind aber in der Kindersprache nicht belegt. Das wird als Evidenz dafür angeführt, dass solche Sätze durch Strukturkenntnis des Kindes ausgeschlossen werden, die ihrerseits schon vor der Entwicklung des spezifischen einzelsprachlichen grammatischen Wissens vorhanden ist, in diesem Fall Wissen über die hierarchische Struktur einer Phrase. Die Voranstellung des ersten finiten Verbs ist also strukturgeleitet und wird angewendet auf das erste passende Segment nach der Subjektphrase.

2. Kreativitätsargument: Eine zweite Erwerbsbeobachtung ist, dass Kinder Sätze bilden können, die sie zuvor nicht gehört haben. Dieses Fak-

tum, so die Argumentation, spricht für ein Strukturwissen, dass diese Kreativität ermöglicht.

3. Defizienter Input: Ein weiteres Argument wird daraus abgeleitet, dass das Ergebnis des Spracherwerbs grammatisches, wiederum unbewusstes Sprachwissen ist, das den Menschen in die Lage versetzt, wohlgeformte Sätze von nicht wohlgeformten zu unterscheiden, z. B. (3–8) gegenüber (3–9).

- (3–6) *Wer kommt?*
- (3–7) *Wer, glaubt Hans, kommt?*
- (3–8) *Welcher Besuch kommt?*
- (3–9) **Welcher, glaubt Hans, Besuch kommt?*

Das ist deshalb erklärungsbedürftig, weil das Kind im Lauf des Spracherwerbs durchaus auch viele nicht wohlgeformte Sätze und abgebrochene Äußerungen hört.

4. Das Fehlen negativer Evidenz: Schließlich ein Argument ex negativo. Es wurde erwähnt, dass die pure lineare Form der Inputäußerungen erwarten ließe, dass das Kind daraus Muster von Äußerungen wie (3–5) ableiten würde. Sie sind aber in der Kindersprache nicht belegt. Nun könnte dieses Fehlen auch damit erklärt werden, dass dem Kind Hinweise auf abweichende Äußerungen gegeben werden, die den Erwerb dann in die Zielrichtung steuern. Nach dem Stand der Kenntnis ist dem aber nicht so. Und eben dieser Umstand des Fehlens negativer Evidenz aus der Sicht des Kindes stärkt die Annahme, dass es vor dem Erwerbsbeginn vorhandenes ›Wissen‹ geben muss, dem das Kind bei der Verarbeitung des Inputs zu spezifischem sprachlichen Wissen folgt.

Für die **Beurteilung der nativistischen Konzeption** sind zunehmend Befunde aus experimentellen Untersuchungen und vom Sprachverhalten geistig kranker Kinder verfügbar geworden. Sie gelten hauptsächlich den Fragen nach der **Modularität des sprachlichen Systems**, besonders in Abgrenzung von bzw. Interaktion mit dem allgemeinen Denkvermögen (vgl. Weinert 2000, bes. Abschnitt 4) und der Existenz universalen sprachspezifischen Wissens vor dem Erwerb (vgl. Höhle/Weissenborn 1999). Die die Modularität betreffenden Befunde stärken weder noch widerlegen sie unbestreitbar die Grundannahmen der nativistischen Konzeption (vgl. Weinert 2000, Kap. 5). Die **psychopathologischen Befunde** sprechen eher für die Unabhängigkeit der Sprachfähigkeit von der sonstigen Denkfähigkeit. Direkt auf spezifisches sprachliches Wissen gerichtete Experimente zur rezeptiven Sprachbeherrschung bestätigen allerdings wiederum, dass Kleinkinder sehr viel früher, als bisher auf Grund von Produktionsdaten angenommen, für Strukturunterschiede in sprachlichem Material sensitiv sind (vgl. Höhle/Weissenborn 1999, Kap. 2.3.4 und 2.3.5). Inwiefern das die nativistische Konzeption bestätigt, bleibt noch zu zeigen.

Das UG-Wissen des Kindes: Was ist, nach Annahme der nativistischen Theorie, der Inhalt des angeborenen sprachlichen Wissens? Wie jede Theorie ist auch diese – bei aller Kontinuität in den Grundannahmen – Veränderungen über die Zeit und Unterschieden infolge unterschiedlicher Sichtweisen einzelner Wissenschaftler ausgesetzt. Das liegt daran, dass

die Beobachtungsdaten aus dem Spracherwerb Deutungsspielräume zu lassen, und an dem Auftauchen neuer Beobachtungen. Von Varianten abgesehen, ist das angeborene ›Sprachorgan‹ grammatisches Wissen. Es enthält (unbewusste) Kenntnis über den Aufbau sprachlicher Ausdrücke, sog. **grammatische Prinzipien**.

Zur Vertiefung

Die universale Struktur jeder syntaktischen Phrase

Jeder sprachliche Ausdruck ist hierarchisch gegliedert. Das Grundschema der Gliederung ist die sog. X-Bar-Struktur.

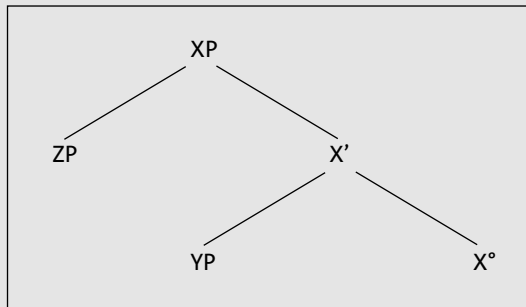


Abb. 3.2: X-Bar-Schema: Universale Gliederung der Phrasenstruktur (vereinfacht auf drei Projektionsebenen).

X° bezeichnet die Position des lexikalischen Elements, durch das Einheiten der XP-internen Struktur, z. B. die Zahl der YP-Positionen, festgelegt sind, und das die syntaktischen Eigenschaften der gesamten Phrase nach außen hin durch Projektion von syntaktischen Merkmalen auf XP-Ebene prägt. Ist X° ein Verb, dann ist XP eine Verbphrase und YP ist eine Nominalphrase, die Objekts-NP. Ist X° ein Nomen, dann ist XP eine NP usw.

Parameter: Nun sind bekanntlich nicht alle Sprachen einheitlich gebaut; dem trägt die Theorie dadurch Rechnung, dass angenommen wird, einige der Prinzipien seien **parametrisiert**. So unterscheiden sich Sprachen z. B. in der Reihenfolge von X° und YP, was durch Annahme einer Hilfsgröße »Kopfposition« im Strukturwissen theoretisch erfasst wird. Ein Parameter hat endlich viele Werte, der Kopfparameter z. B. zwei, »kopffinital« und »kopffinal«. Eine detaillierte Darstellung der derzeit anzunehmenden Prinzipien und Parameter liefern Stechow/Sternefeld (1988) und Chomsky/Lasnik (1993, Kap. 1).

Zusammenfassend: Das logische Problem des Kindes beim Spracherwerb besteht darin, die Parameterwerte ausfindig zu machen, die in seiner Umgebungssprache ausgeprägt sind. Erwerbslogisch stellt die **Parametrisierung** also so etwas dar, wie das strukturelle Bindeglied zwischen dem universalen sprachlichen Wissen und den spezifischen Strukturverhältnissen in der jeweiligen Umgebungssprache. Abgesehen davon, dass sich das Konzept in der typologischen Forschung zunehmend bestätigt, wurde seine Wirkung auch in materialreichen Studien des Spracherwerbs aufgezeigt (vgl. die Synchronie des Erwerbs von Doppel-Nomen-Zusammensetzungen und Verb-Partikel-Sätzen in der Kindersprache; Snyder 2007).

Zwei Positionen zum Erwerbsverlaufs: Wie erklärt schließlich die nativistische Theorie den beobachteten Erwerbsverlauf? Hierzu ist vorab etwas Grundsätzliches zu berücksichtigen. Es wird streng unterschieden zwischen dem sprachlichen Wissen des Kindes und dem Vorgang, die Inputdaten mit dem UG-Wissen in Verbindung zu bringen, was eine Reihe von Problemlösungen prozeduraler Art impliziert, z. B. das Segmentieren des Schallstroms in Laute, Silben und Wörter, das Zuordnen von Wortformen zu Begriffen, das Klassifizieren von Wörtern etc. Vor diesem Hintergrund kann nun entweder angenommen werden, dass das genetisch verankerte Wissen von Anbeginn in Gänze vorhanden ist (zur sog. **starken Kontinuitätsannahme** vgl. Pinker 1994) oder dass es – genetisch gesteuert – in den ersten Monaten und Jahren des Spracherwerbs wächst (zur **schwachen Kontinuitätsannahme** vgl. Borer/Wexler 1987). Auch neuere experimentelle Befunde stützen diese Annahme (vgl. Friederici 2005).

3.5.2 | Das Sprachgebrauchsmodell

Um das Wesentliche dieses Forschungsprogramms verständlich zu machen, ist es ratsam, zunächst die anfänglichen Grundannahmen vorzustellen. Es sind, wie in allen Spracherwerbstheorien, Annahmen über die spezifische Relevanz von Erwerbsvoraussetzungen; sie finden sich einführend gelistet in Tomasello (2009, Kapitel 2).

Grundannahmen: Was besagt das »Usage based-Modell« des Spracherwerbs? Zunächst einmal wird die Existenz von angeborenem universalem sprachbezogenen Wissen des Kindes kategorisch bestritten. Das Kind, so die Grundannahme, erwerbe die Sprachfähigkeit durch die aufmerksame, von Neugier und Wissensdrang getriebene Verwendung der Sprache mit den Erwachsenen seiner Brutpfleumgebung.

Kommunikativen Fertigkeiten: Es verfüge dazu über die folgenden kommunikativen Fertigkeiten:

- Die Fertigkeit, Aufmerksamkeit auf Gegenstände und Sachverhalte mit Kommunikationspartnern zu teilen.
- Die Fertigkeit, der Aufmerksamkeit und der Gestik von Personen zu folgen, die sich auf entfernte Gegenstände und Ereignisse außerhalb des Bereichs der unmittelbaren Interaktion befinden.
- Die Fertigkeit, selbst die Aufmerksamkeit anderer auf entfernte Objekte und Ereignisse zu lenken.
- Die Fertigkeit, kulturgeleitet die absichtsgeleiteten Handlungen anderer zu erlernen, einschließlich kommunikativer Aktivitäten und ihrer zugrundeliegenden Absichten.

Hier finden sich also die oben genannten Merkmale der »sozialen Ausstattung« des Säuglings wieder.

Kognitive Fähigkeiten: Des Weiteren notwendig und beteiligt an dem Gelingen des Spracherwerbs sind nach Tomasello die folgenden kognitiven Fähigkeiten:

- Die Fähigkeit, aus der Ähnlichkeit von wahrgenommenen Reizen Kategorien von einander ähnlichen Objekten und Ereignissen zu abstrahieren.

Spracherwerb
als Lernvorgang

- Die Fähigkeit, aus sich wiederholenden Mustern von Wahrnehmung und Aktion sensomotorische Schemata zu bilden.
- Die Fähigkeit, anhand von beobachteten Wahrnehmungs- und Verhaltenssequenzen häufigkeitsbasierte Verteilungen herauszufinden.
- Die Fähigkeit, aus einander ähnlichen Funktionen von einzelnen Bestandteilen komplexer Einheiten Analogien zwischen ihnen abzuleiten.

3.5.3 | Der kognitivistische Ansatz

Kennzeichnend für diese Theorie ist die Annahme, dass die Sprachfähigkeit und ihre Entwicklung auf der **Denkfähigkeit** des Menschen und deren Entwicklung beruhen. »Beruhen« heißt hier, dass die Sprachentwicklung die **Entwicklung der Intelligenz** voraussetzt und zwar so, dass die Entwicklung von sprachlichen Teilfähigkeiten durch die Entwicklung entsprechender Intelligenzleistungen bedingt und determiniert ist. Der Spracherwerb stellt demnach eine spezifische Denkaktivität des Kindes dar, die auf jeweils vorangehenden nicht-sprachlichen Intelligenzleistungen aufbaut.

Der besondere Nutzen der Sprache für das Denken ergibt sich aus ihrer **Repräsentationsfunktion**. Das sprachliche Symbol liefert die Voraussetzung, Vorstellungen im Geist darzustellen, zu kombinieren und frei von der aktuellen Situation und Anschauung damit geistig zu handeln. Forschungslogisch muss also zunächst herausgefunden werden, wie sich die Intelligenz/das Denken des Kindes entwickelt, von der Sensomotorik über das mentale Repräsentieren von Anschauungen, das Operieren mit diesen Repräsentationen bis hin zum abstrakten und formalen Denken z. B. das Erkennen von und Operieren mit logischen Relationen. Eben dieses Programm bestimmte die Arbeit von Jean Piaget, wie er selbst in einer knappen Autobiographie mitteilt (vgl. Piaget 1972).

Die kognitivistisch-konstruktivistische Konzeption der Sprachentwicklung des Kindes ist demnach grundsätzlich als in die **Entwicklung der Intelligenz des Kindes**, seiner Neugier und seiner sozialen Interaktionsfähigkeit eingebettet zu sehen. Zwar hat im Werk von Piaget die Beobachtung des Sprachverhaltens des Kindes am Anfang gestanden (vgl. Piaget 1923); sie war aber ebenso wie bei Preyer und den Sterns mehr eine Methode, die Entwicklung der kindlichen Psyche, genauer die Genese des Denkens beim Kind zu untersuchen. Diese weist nach Piaget **vier sukzessive Hauptstufen** auf: die sensomotorische Stufe, die Stufe des anschaulichen Denkens, die Stufe des konkret-operativen Denkens und – beim Erwachsenen schließlich – die Stufe des formal-operativen Denkens. Welche Beobachtungen würden diese Konzeption stützen? Man würde z. B. erwarten, dass der Verwendung von Sprache in der Interaktion ihre Verwendung in Vorgängen lauten Denkens in der Entwicklung vorangeht und dass diese Funktion des Sprechens auch prinzipiell erhalten bleibt. Man würde weiter erwarten, dass eine sprachliche Ausdruckseinheit erst dann aus dem Input aufgenommen wird, wenn ihr ein Konzept entspricht. Das muss natürlich nicht die Bedeutung in der Erwachsenensprache sein, aber jeden-

falls eine Vorstellungseinheit im Wissen des Kindes. Und so müsste es für alle Bestandteile des Sprachsystems sein, die phonologischen, morphologischen und syntaktischen Mittel; kurz gesagt, die kognitivistische Konzeption lässt einen sog. funktional gesteuerten Spracherwerb erwarten.

Evidenz

Lautes Denken: Die erstgenannte Erwartung sah Piaget in dem Phänomen des sog. Monologisierens des vier- bis siebenjährigen Kindes bestätigt. Die beim selbstorganisierten Spielen beobachteten Kinder einer Kindertagesstätte redeten vor sich hin, ihre Aktivitäten offenbar eher sprachlich begleitend als mitteilend, obwohl sich die Äußerungen nach Form und situativen Gegebenheiten nicht von kommunikativer Interaktion unterscheiden (vgl. Piaget 1973).

Objektpermanenz: Für die Erwartung eines konzeptgesteuerten Erwerbs sprachlicher Mittel sprechen Beobachtungen zur zeitlichen Reihenfolge von begrifflicher und sprachlicher Entwicklung. Von Geburt an bis etwa zum Ende des ersten Lebensjahres ist dem kindlichen Denken ein Objekt nur so lange präsent, wie es wahrgenommen wird. Erst zwischen 0;10 und 1;0 entwickelt sich die kognitive Fähigkeit, eine geistige Vorstellung eines Objekts zu bewahren, die sog. Objektpermanenz. Zeitlich mit ihr einher, genauer gesagt geringfügig nachzeitig, geht der Erwerb des ersten bedeutungshaltigen Wortes vonstatten. Sprachliche Mittel für Warum-Fragen sind zeitlich an die begriffliche Erkenntnis des Kausalzusammenhangs gekoppelt und zahlreiche Beobachtungen in Folgeuntersuchungen im Rahmen des kognitivistischen Paradigmas haben weitere Zusammenhänge zugunsten des funktionalistischen Modells erbracht.

Allerdings haben nicht alle Ergebnisse späterer Untersuchungen die ursprünglichen Annahmen bestätigt. Den generellen Zusammenhang zwischen kognitivem Niveau und sprachlicher Entwicklung haben Schaner-Wolles/Haider (1987) überprüft. Von rund 60 Kindern zwischen 5 und 9 Jahren wurde mit einer standardisierten Testbatterie die Entwicklung ihres operativen Denkens ermittelt. Parallel dazu wurde mit einer Satz-Bild-Matching-Aufgabe ihr Verstehen von Sätzen mit unterschiedlich komplexen anaphorischen Relationen gemessen. Die Ergebnisse zeigten einen signifikanten Zusammenhang zwischen dem Alter und der kognitiven Entwicklung, aber keinen durchgängigen Zusammenhang zwischen kognitiver und sprachlicher Entwicklung. Damit bestätigen sich Befunde früherer Experimente, besonders von Sinclair-de Zwart (1971).

Deutlicher positiv ist die Evidenz über den Zusammenhang zwischen der Struktur der Entwicklung der sensomotorischen Intelligenz und dem Erwerb semantischer Sprachmittel. So stehen nach Bloom (1973) und Szagun (2013) Stufen des Syntaxerwerbs mit Stufen der sensomotorischen Entwicklung in den ersten zwei Lebensjahren insofern in Analogie, als der syntaktischen Entwicklung die Entwicklung semantischer Konzepte, nämlich der Kasusrollen im Sinne von Fillmore (1968) zu Grunde liegen, welche ihrerseits analog zu den Stufen der Sensomotorikentwicklung abläuft.

Operationsprinzipien: Die Frage, wie das Kind in der ja nicht vorsegmentierten Folge von Schall die formalen Einheiten erkennt, denen sensomotorischen Bedeutungen zuzuordnen sind, eine Frage übrigens, die aus der Sicht jeder Theorie beantwortet werden muss, hat durch die sprachvergleichenden Erwerbsuntersuchungen von Slobin (1973) eine kogniti-

vistisch basierte Antwort gefunden. Die vergleichende Analyse von Erwerbsdaten aus vierzig Sprachen sowie die darauf bezogene Kategorisierung der Inputeigenschaften führte zur Annahme kognitiver Prinzipien, denen alle Kinder bei der Segmentierung, Klassifikation und beim Erkennen grammatischer Beziehungen wahrscheinlich gefolgt sind: sog. universale Operationsprinzipien.

Zur Illustration

Spracherwerbsdaten von ungarisch-serbokroatischen bilingualen Kindern (vgl. Slobin 1973)

Die Spracherwerbsdaten ungarisch-serbokroatisch bilingualer Kinder weisen auf, dass die Ausdrücke für die Bezeichnung von Ortsrelationen im Ungarischen früher gelernt werden als im Serbokroatischen. Zugleich ist aber klar, dass die Kinder die entsprechenden Konzepte schon haben müssen, auch wenn sie die sprachlichen Ausdrücke des Serbokroatischen nicht erworben haben. Sie kommunizieren sie auf anderen, lernersprachlichen Wegen, durch Wahl geeigneter Verben, durch Bezug auf kontextuelle Gegebenheiten o. A. Die Analyse der beteiligten Sprachen ergibt, dass die Ortsbeziehungen im Ungarischen einheitlich durch monomorphematische Postpositionen ausgedrückt werden, im Serbokroatischen durch Präpositionen, Nominalflexion oder beides in Kombination. Aus diesem und den Befunden aller anderen Daten ergibt sich eine universale Erwerbsbeobachtung: Postverbale und postnominale lokale Ausdrücke werden früher gelernt als präverbale und pränominal. (vgl. Slobin 1973, S. 187 ff.) leitet daraus die Existenz des Operationsprinzips ab: Richte deine Aufmerksamkeit auf das Ende des Wortes. Auf die gleiche Weise, abgeleitet aus universalen Erwerbsbeobachtungen, werden weitere Operationsprinzipien erschlossen (vgl. Slobin 1973, S. 205–206):

- Vermeide Ausnahmen.
- Der Gebrauch grammatischer Ausdrücke soll semantisch gerechtfertigt sein.

Die kognitivistische Spracherwerbsforschung weist eine große Zahl von Einzelergebnissen auf, die die semantische Basis des Formenerwerbs mehr oder weniger direkt belegen; Entwürfe eines kohärenten Modells des kindlichen Laut-, Wort- und Syntaxerwerbs wurden erst in jüngster Zeit durch Budwig (1995) vorgestellt. Als problemgeschichtliche Einführung in das Gebiet empfiehlt sich Weinert (2000).

3.6 | Der bilinguale Erstspracherwerb

Bilingualismus und Mehrsprachigkeit: Obwohl die meisten Menschen auf der Erde mehr als nur eine Sprache sprechen, sind die linguistische und die psychische Seite der individuellen Mehrsprachigkeit nicht so gut bekannt wie Verlauf und Ergebnis des monolingualen Erstspracherwerbs. So stehen bis heute weit auseinanderliegende Behauptungen über Verlauf und Ergebnis des mehrsprachigen Spracherwerbs nebeneinander. Dass

zugunsten aller, auch scheinbar einander widersprechender Behauptungen auf beobachtbare Fakten verwiesen werden kann, deutet einerseits darauf hin, dass das Ausmaß an Variation in der Entwicklung bilingualer Kompetenz groß und ihre Bedingungen eben noch wenig bekannt sind. Und so ist es auch. Andererseits hat die wissenschaftlich begründete Kenntnis in diesem Feld seit den Anfängen des 20. Jahrhunderts, wo auch etwa die Anfänge der Bilingualismusforschung liegen, deutlich zugenommen.

Im Rahmen dieses Einführungstextes werden wir uns auf Fragen und Antworten beschränken, die zu einem ersten zusammenhängenden Eindruck von den Phänomenen, Problemen und der derzeitigen Kenntnislage verhelfen.

Abgrenzungsproblem: Am Anfang stehen natürlich die Probleme. Was bezeichnet man als bilingualen Erstspracherwerb in Unterscheidung zu anderen Formen von Mehrsprachigkeit? In der englischsprachigen Literatur wird der Ausdruck ›*bilingual*‹ für die Existenz zweier Sprachen in einer Gesellschaft und für die Beherrschung zweier Sprachen durch eine Person verwendet. Ferner werden mit ›*bilingual*‹ alle Formen individueller Mehrsprachigkeit bezeichnet, also das Ergebnis des bilingualen Erstspracherwerbs, des kindlichen, jugendlichen und erwachsenen Zweitspracherwerbs und des Fremdsprachenlernens.

Im Deutschen wird ›*bilingual*‹ enger, nur für die individuelle Mehrsprachigkeit verwendet. **Bilingualismus** ist das Ergebnis des bilingualen Erstspracherwerbs, des gleichzeitigen Erwerbs zweier Sprachen mit gleich umfangreichen Kommunikationsanteilen von Geburt an. Die Bedeutung von ›mehrsprachig‹ im Deutschen entspricht hingegen etwa der des englischen Ausdrucks ›*bilingual*‹.

Bilingualer Erstspracherwerb ist der von den Erwerbsumständen geprägte, gleichzeitige Erwerb von Wissen und Fertigkeiten in zwei verschiedenen Sprachen.

Zum Begriff

Über den **Verlauf des bilingualen Erstspracherwerbs** lässt sich bei aller Variation doch ein Gesamteindruck gewinnen. Demnach verläuft der Bilingualismuserwerb rezeptiver und produktiver Fertigkeiten in der Reihenfolge ähnlich wie der jeweils vergleichbare monolinguale Spracherwerb. Unter der Bedingung etwa gleich umfangreichen Inputs beider Sprachen werden im Großen und Ganzen dieselben Schritte der Laut-, Wortschatz- und Syntaxentwicklung berichtet, wie oben für den einsprachigen Erstspracherwerb dargestellt.

Repräsentation sprachlichen Wissens bei Bilingualismus: Diese pointierte Charakterisierung legt die Frage nahe, ab welchem Zeitpunkt das Kind sein Wissen sprachenweise organisiert. Täte es das nicht, würde man ja Äußerungen erwarten, die aus einem Gemisch an lautlichen, lexikalischen und syntaktischen Mitteln beider Sprachen bestehen. Und solche Äußerungen sind auch belegt, z. B. in der materialreichen Studie von Deuchar/Quay (2001) über die englisch-spanisch bilinguale Tochter M. von Margaret Deuchar. In der Zweiwortphase produziert M. einerseits Äu-

ßerungen wie *more juice* [Wunsch: Mehr Saft!], *mas pasa* [mehr Trauben] und zur selben Zeit *more leche* [mehr (engl.) Milch (span.)] *more galleta* [mehr (engl.) Kuchen (span.)] und *cuchara floor* [Löffel (span.) Boden (engl.)] (vgl. Deuchar/Quay 2001, S. 80–81 sowie Anhang III).

Das Nebeneinander von Ausdrücken in beiden Sprachen in Äußerungen gegenüber dem selben Adressaten und die zeitliche Verteilung solcher sogenannter gemischter Äußerung wurde als Evidenz für eine dreistufige Entwicklung gesehen, eine erste Stufe, in der das lexikalische und das grammatische Wissen des Kindes nicht separat organisiert sind, eine zweite, in der das lexikalische Wissen sprachenweise getrennt, das grammatische aber noch nicht separiert ist und eine dritte, in der sich auch das grammatische Wissen in zwei mentale Grammatiken separiert (vgl. Volterra/Taeschner 1978; Taeschner 1983).

Erwerbsverlauf

Ein System oder zwei: Diese Ein-System-Hypothese (erste Stufe im Modell von Volterra/Taeschner 1978) ist allerdings mit der steigenden Zahl von Fallanalysen fraglich geworden. Zweifel erwecken u. a. die Beobachtungen zur rezeptiven Lautbeherrschung. Wie oben erwähnt, erkennt das Kind schon in den ersten Wochen den Unterschied zwischen der Sprache seiner Umgebung und anderen Sprachen, nicht aber spontan den zwischen zwei fremden Sprachen. Das ist bei Kindern, die in bilingualen Umgebungen aufwachsen, anders, selbst wenn sie selbst nicht bilingual sind. Eilers/Gavin/Kimbrough-Oller (1982) untersuchten die passive Lautdiskriminierungsfähigkeit englischsprachiger und spanischsprachiger Kinder zwischen 6 und 8 Monaten mit der *Visually Reinforced Infant Speech Discrimination*-Methode. Es zeigte sich, dass die spanischsprachigen, die zwar nicht verschiedensprachige Eltern hatten, aber Englisch in ihrer Familie hörten, die Unterschiede signifikant besser erkannten als die monolingual englischsprachigen Kinder. Auch die Lauteigenschaften einer ganz fremden Sprache (Tschechisch) wurden von den mehrsprachigkeitserfahrenen Kleinkindern besser identifiziert. Bosch/Sebastian-Galles (2001) konnten nachweisen, dass bereits 4 Monate alte Kinder aus Spanisch-Catalan-Elternhäusern Intonationsmuster beider Sprachen unterscheiden können (siehe auch Hamers/Blanc 2000).

An lexikalischer Evidenz sprachen für die Ein-System-Hypothese zunächst Wortlisten von zwei deutsch-italienischen Kindern, die keine Übersetzungsäquivalenten Paare enthielten. Dieser Umstand galt als Nachweis dafür, dass das Kind die beiden Sprachen im Input nicht als separate Systeme bearbeitet, sondern als eines und dem Prinzip folgt: ein Wort – eine Bedeutung. In Reanalysen wurden allerdings dann doch solche Paare auffindig gemacht (vgl. De Houwer 1995, S. 231–232).

Noch deutlichere Gegenevidenz präsentierte Quay (1993), die das lexikalische Repertoire und dessen Entwicklung bei der schon erwähnten Tochter von Deuchar analysierte. Die tabellarische Übersicht in Deuchar/Quay (2001, Anhang II) weist Übersetzungsäquivalente nach und zeigt, dass M. im zweiten Jahr (von 1;0 bis 1;10) mehr paarige als exklusiv einzelsprachliche Wörter beherrschte. Auch die Annahme fusionierter grammatischer Strukturen beider Sprachen zu einer Grammatik in der Stufe II der bilingualen Entwicklung (vgl. Volterra/Taeschner 1978) ist als problematisch nachgewiesen worden (vgl. De Houwer 1995, S. 233).

Positive Evidenz zugunsten einer frühen Trennung liefern Vergleiche der grammatischen Entwicklung bilingualer Kinder mit vergleichbaren monolingualen, wie sie Meisel (1989) unternimmt. Die Daten von zwei deutsch-französisch bilingualen Kindern im Alter von 1;0 bis 4;0 werden hinsichtlich der Entwicklung der Wortstellung und der Subjekt-Verb-Kongruenz mit solchen vergleichbarer monolingualer Kinder verglichen.

Frühe Trennung: Alle Vergleiche sprechen für eine frühe Trennung der grammatischen Systeme bei den bilingualen Kindern. Der Wortstellungsvergleich ergibt, dass die bilingualen Kinder bei der Annäherung an die Zielsprache gar weniger Umwege gehen und sich mit geringerer lernersprachlicher Variation auf die Erwachsenensprache zu bewegen. Insgesamt ist also anzunehmen, dass der bilinguale Erstspracherwerb tatsächlich ein doppelter Erstspracherwerb ist und nicht ein Beginn mit einem fusionierten System und der allmählichen Separation in zwei. Dagegen sprechen auch nicht Äußerungen mit Ausdrücken aus beiden Sprachen. Sie kommen auch bei erwachsenen Bilingualen vor und stellen sich – je nach Art und Verwendungssituation – als Ergebnisse von Code-Wechsel (*code switching*) oder Entlehnung heraus. Im Übrigen fallen sie mit einem Anteil von 2–4 % in den ersten vier Jahren nicht so sehr ins Gewicht, wie es in der Literatur den Eindruck machen mag (vgl. Tracy 2000).

Dennoch, die Frage ist nicht abschließend geklärt, selbst nicht hinsichtlich der phonologischen Entwicklung, die besser zu überschauen und daher genauer zu beobachten ist als die syntaktische (vgl. Tracy/Gawlitze-Maiwald 2000) oder gar die lexikalische. Es ist bekannt, dass Kinder prosodische Eigenschaften der Umgebungssprache sehr früh aufnehmen, also z. B. das trochäische Betonungsmuster deutscher Wörter: Betont – Unbetont (*Sessel, Lampe ...*). Das Französische weist hingegen ein jambisches Muster auf »unbetont – betont« (*maison, soleil*). Außerdem ist bekannt, dass Kinder in der Anfangsphase des Lexikerwerbs die lautliche Form der Wörter verkürzen, z. B. *Banane* > *nane*. Diese beiden Entwicklungsphänomene hat Paradis (2001) ausgenutzt, um die Separatheit der phonologischen Systeme von jeweils rund 20 knapp dreijährigen französisch-englisch Bilingualen und englisch-französisch Bilingualen zu ermitteln und mit Einsprachigen zu vergleichen.

Experiment zur Separatheit phonologischer Systeme bei Bilingualen

Experiment

In einem spielerisch eingebetteten Nachsprechttext haben die Kinder mehrsilbige Non-Wörter mit englischem bzw. französischem Akzentmuster nachgesprochen und dabei erwartungsgemäß verkürzt Paradis (2001). Die Ergebnisse der Bilingualen wurden zum einen für sich ausgewertet und zum zweiten mit denjenigen der Einsprachigen verglichen. Die erste Auswertung zeigt, dass die Bilingualen in ihren beiden Sprachen tatsächlich dem Reduktionsmuster der jeweiligen Sprache folgen, also ein Beleg für die Separatheit der beiden beherrschten phonologischen Systeme. Der Vergleich mit den jeweils Einsprachigen zeigte aber doch Unterschiede. Die Daten der französischen Bilingualen unterschieden sich im englischen Bereich von denen der monolingual englischen und zeigten einen französischen Einfluss.

Was heißt das? Manche sagen, die Frage nach Abhängigkeit oder Unabhängigkeit ist falsch gestellt. Vielmehr müsse man akzeptieren, dass es eben monolinguale Französischsprachige gibt und monolinguale Englischsprachige – und Bilinguale. Diese seien eben keine Kombination aus zwei Monolingualen sondern eine dritte Spezies.

Bilingualismus und kognitive Kontrolle

Verursacht Bilingualismus Kosten für den Spracherwerb? Die Sprache zu lernen, ist ein Vorgang, der die Aufmerksamkeit von Kind und Umgebung jahrelang beschäftigt – den größten Teil des Tages über. Was würde man also für den Fall erwarten, dass ein Kind von Geburt an nicht nur mit Äußerungen in einer, sondern in zwei oder mehr Sprachen konfrontiert wäre? Eine nicht unplausible Erwartung wäre, dass – gegeben dieselbe Menge an Zeit – von jeder Sprache eben nur ein Teil erworben werden kann, dass der vollständige Erwerb zweier Sprachen also etwa doppelt so lange dauert wie der Erwerb einer Sprache oder dass die Kapazität für den bilingualen Erstspracherwerb einem anderen Bereich der kognitiven Entwicklung des Kindes abgezogen wird, z. B. der Entwicklung des räumlichen Sehens oder des formalen Denkens. Kurzum, die plausible Annahme wäre, dass bilingualer Erstspracherwerb das Kind etwas kostet, seien es Einbußen in der Vollständigkeit, sei es Zeit, seien es Einbußen in anderen Domänen der kognitiven Fähigkeiten.

Heute kann man sicher sagen, dass die von Epstein im Jahr 1915 gemachte Behauptung über die Kosten des Bilingualismus für die Entwicklung der Intelligenz des Kindes so weder zutrifft, noch so begründet werden kann:

»Indes ahnen selbst diejenigen, die den erzieherischen Wert des Sprachstudiums bestreiten, nicht, dass dieser Unterricht mit unmittelbaren Beeinträchtigungen geistiger Fähigkeiten einhergehen könnte. Nun führt aber die jüngst gewonnene Gewissheit, dass mehrere Sprachen sich beim Menschen gegenseitig behindern, unausweichlich zu dem Schluss, dass Mehrsprachigkeit ein Störfaktor für das sprachliche Denken darstellt, und damit zu der Konsequenz, dass durch den Erwerb einer weiteren Sprache die Klarheit und Reinheit, bestenfalls nur die Geschwindigkeit in der Ausdrucksfähigkeit in den zuvor erworbenen Sprachen vermindert wird.« (Epstein 1915, S. 141, aus dem Französischen übersetzt von RD).

Eher hat sich die noch früher entwickelte Theorie – mit Differenzierungen – bestätigt, dass das Kind zwei Sprachen ohne kognitive Einbuße parallel lernen kann, wenn das sog. Partnerprinzip (Ein und dieselbe Sprache gegenüber dem Kind konstant aus dem Mund derselben Person) eingehalten ist; so Ronjat (1913). Anschaulich und enzyklopädisch eingängig vorgestellt wird das Thema u. a. in Baker/Prys Jones (1998, S. 62 ff.), in historischer Einordnung in Dietrich (2006). Ergänzend zeigen Befunde zur Fähigkeit der Aufmerksamkeitssteuerung, dass echte Bilinguale monolingualen Vergleichspersonen an Reaktionsgeschwindigkeit und Korrektheit in sprachlichen und in nichtsprachlichen Bereichen überlegen sind; vgl. Stafford (2011).

Vergleiche lexikalischer Repertoires von bilingualen und monolingualen Kindern zeigen übrigens, dass der Umfang des Wortschatzes pro Sprache bei bilingualen Kindern geringer ist, selbst in der sog. dominanten Sprache. Pearson/Fernandez/Oller (1993) verglichen die Wortschatzent-

wicklung von 25 englisch-spanischen Bilingualen mit derjenigen von 35 Einsprachigen (32 engl.; 3 span.) über die Zeit von 0;8 bis 2;6. Insgesamt erwarben die bilingualen Kinder ungefähr halb so viele Wörter wie die Einsprachigen, ein Unterschied, der in der weiteren Entwicklung, besonders in der Schule verschwindet. Die Zahl der Konzepte, die die Monolingualen ausdrücken können, gleicht etwa der, die Bilinguale mit dem Wortschatz ihrer beiden Sprachen beherrschen. Im Übrigen zeigt sich, dass die Bilingualen durchaus dem Prinzip ›ein Wort – eine Bedeutung‹ folgen, allerdings jeweils sprachenweise, wie eben das einsprachige Kind auch; die einschlägigen Beobachtungen stammen allerdings von Kindern im Vorschulalter (vgl. Hamers/Blanc 2000, S. 61).

3.7 | Zweitspracherwerb

Als **Zweitspracherwerb** (L2-Erwerb) wird der Erwerb sprachlichen Wissens und sprachlicher Fertigkeiten in einer weiteren Sprache **nach** Abschluss des Erstspracherwerbs bezeichnet.

Zum Begriff

Vorgeschichte: Systematisch bearbeitet wurde dieser Phänomenbereich bisher fast ausschließlich unter didaktisch-methodischen Zielsetzungen, also aus der Sicht der Unterrichtenden, nicht der Lernenden. Eine lange Tradition und eine entsprechend große Wirkung auf die einschlägige Forschung hat die Annahme, Verlauf und Ergebnis des Zweitspracherwerbs würden hauptsächlich davon geprägt, in welchen Bereichen die neue Sprache der schon gelernten ähnlich ist oder unähnlich, die sog. **Kontrastivitätsannahme**.

Neben methodisch-didaktischen Interessen galt Phänomenen des Zweitspracherwerbs auch die Aufmerksamkeit der Kultur- und Sprachkontaktforschung. Im Mittelpunkt stand hier die Entstehung von Zwischensprachen/Lernersprachen als Reaktion auf neu eingetretene Sprachkontaktsituationen durch Kolonialisierung und Migration. Die Arbeiten sind eher soziologisch, ethnologisch oder soziolinguistisch ausgerichtet. Gegenstand psycholinguistischer Forschung ist der L2-Erwerb merkwürdigerweise erst sehr spät geworden – auf dem amerikanischen Kontinent eher als auf dem europäischen. So nimmt das Thema in psycholinguistischen Gesamtdarstellungen keinen oder wenig Raum ein (Harley 2001, S. 135–138; Dijkstra/Kempen/Fuchs 1993, S. 101).

Die Lage hat sich seit den 1980er Jahren allerdings deutlich verändert. Die psycholinguistische Untersuchung von Ausgangsbedingungen, Erwerbsvorgängen und Ergebnissen hat sich rasch entwickelt sowohl hinsichtlich der Breite der Themen als auch der Vielfalt der angewandten Methoden. Die Entwicklung der Zweitspracherwerbsforschung im 20. Jahrhundert findet sich im Überblick dargestellt in (Dietrich 2006), das einschlägige Handbuch ist Doughty/Long (2003) und den methodologischen Stand dokumentieren Doughty/Long/Menn/Bernstein Ratner (2000) und Gullberg/Indefrey (2006). Demnach ist die kognitionswissenschaftliche Betrachtung besonders des ungesteuerten, in der alltäglichen Kommuni-

L2-Forschungs-
geschichte

kation und durch sie stattfindenden L2-Erwerbs gerade auch zur Grundlegung didaktisch-methodischer Untersuchungen produktiv und integrativ. Im Folgenden werden aus dieser Sicht Voraussetzungen, heuristische Konzepte, Verlaufsmuster und Erklärungsansätze präsentiert.

In Klein (1986, S. 33), dem ersten Entwurf einer kohärenten psycholinguistischen Rekonstruktion des L2-Erwerbs, findet sich am Anfang von Kapitel 2, dem analytischen Aufriss der Phänomene, zur Veranschaulichung der prototypischen Erwerbssituation des L2-Lerners die Schilderung eines exotischen Szenarios. Der Leser möge sich vorstellen, er habe den Absturz seines Flugzeugs über Neu-Guinea überlebt, werde von den Einheimischen freundlich aufgenommen und müsse sich angesichts der völligen Abgeschiedenheit des Volkes auf unbestimmte Zeit auf ein Leben in ihrer Umgebung einstellen. Dieser Fiktion sehr ähnlich müssen die neun britischen Offiziere und Matrosen ihre Lage empfunden haben, als sie nach der Meuterei auf der *Bounty* 1789 auf der Insel Pitcairn im Südpazifik an Land gegangen und auf Mitglieder der einheimischen Tahitianer-Bevölkerung getroffen sind. Das sprachliche Ergebnis ist bekannt, das pazifische Englisch-Kreol Pitcairnesisch.

Stand der
Forschung

Voraussetzungen: Fragen wir uns, wie auch bezüglich des L1-Erwerbs, was die Voraussetzungen des ungesteuerten L2-Erwerbs sind, welcher Art das erworbene sprachliche Wissen, welcher Art die spezifischen Probleme beim Aufbau der L2-Kompetenz, wie der Erwerb verläuft und wie variabel er ist. Möglichkeit, Verlauf und Ergebnis des L2-Erwerbs werden im Grunde von denselben Arten von Voraussetzungen bestimmt bzw. geprägt, wie beim L1-Erwerb: von biologischen, kognitiven und sozialen. Denn der Zweitsprachlerner soll, jedenfalls im Ergebnis, dasselbe leisten, wie das Neugeborene, nämlich eine sprachliche Kompetenz aufbauen. Anders als das Neugeborene ist der L2-Lerner allerdings mit einem entwickelten artikulatorischen Apparat und einem ausgereiften Zentralnervensystem ausgestattet. Damit entfallen Einschränkungen, die beim frühen L1-Erwerb wirken. Ebenfalls im Unterschied zum Kind verfügt der Mensch zu Beginn des L2-Erwerbs über lexikalisches und grammatisches Wissen sowie über Kenntnisse und Routinen der sprachlichen Kommunikation im Allgemeinen, also sprachliche Handlungskategorien wie Grüßen, Anweisen, Fragen etc., das Verstehen und Verwenden von konversationalen Implikaturen. Erwachsene verfügen des Weiteren über außersprachliches Wissen von Begriffen und Fakten ihrer Lebenswelt. Ob und in welcher Differenziertheit beim L2-Erwerb angeborenes universales sprachliches Wissen wirksam ist, ist hier ebenso ungeklärt wie für den Fall des L1-Erwerbs. Dazu unten mehr.

Bei aller Verschiedenheit in den Lernvoraussetzungen hat die Bezugnahme auf den Erstspracherwerb den Anstoß zu Arbeiten gegeben, durch die und in deren Tradition und aus deren Gegenpositionen wesentliche übereinstimmende Befunde hervorgegangen sind, die unser Verständnis von der Natur des zweitsprachlichen Wissens ausmachen. Am Anfang stand die lange Zeit unbezweifelte Annahme, der Zweitspracherwerb sei im Wesentlichen der Erwerb eines neuen sprachlichen Codes und die Einschleifung von Reiz-Reaktionsmustern, die den Ausdrucksmitteln dieses Codes entsprechen. Wo es im Französischen *lune* heißt, ist im Deutschen

Mond zu verwenden und französisch *la* ist deutsch *die*. Das lässt erwarten, dass der französische Deutschlerner wahrscheinlich zunächst *die Mond* statt *der Mond* sagen wird, ein **Interferenzfehler**, der durch gezieltes Training zu verhindern ist. Systematische Überprüfungen dieser Kontrastivitätshypothese sowie die Erkenntnis der L1-Forschung in den späten 1960er Jahren, dass auch der Erstspracherwerb kein Reiz-Reaktionslernen sondern der kognitiv betriebene Aufbau sprachlichen Wissens ist, standen am Anfang der heutigen Einblicke in das L2-Wissen (Dulay/Burt 1974).

Verlauf und Fossilierung: Das Ergebnis des L2-Erwerbs ist also eine unbewusste Kenntnis von lexikalischen Einheiten und Strukturregularitäten sowie die – ebenfalls unbewusste – Fähigkeit, dieses Wissen in der Kommunikation passiv und aktiv zu verwenden. Wie der Erstspracherwerb weist auch der Zweitspracherwerb Züge systematischer Entwicklung auf, mit dem bemerkenswerten Unterschied, dass der L2-Erwerb sich im Allgemeinen vor Erreichen der vollständigen zielsprachlichen Beherrschung deutlich verlangsamt oder endet. Man spricht von der Fossilierung des lernersprachlichen Systems auf einem mehr oder weniger zielsprachennahen Niveau.

Der folgende Ausschnitt einer kurzen Erzählung über einen Arztbesuch eines italienischen Deutschlerner veranschaulicht typische Merkmale einer relativ früh fossilisierten Lernervarietät (vgl. Dietrich 2003, S. 67).

Beispiel

Der Arztbesuch

Eine Woche krank
Komme Doktor
diese (zeigt auf sein Handgelenk) bissele kaputt
dottore verstehn ander
nix richtig
mache Creme
Creme viel kalde, viel kalde (d. i. caldo (ital.) = heiß)
Alles rot
fertig dies, müsse komme Doktor geh
bissele kaputt
besser Massage, nix Creme, Massage.

Die Erwerbssituation: Die Beobachtung, an der die Zweitspracherwerbsforschung ansetzt, ist, dass ein deutscher Muttersprachensprecher den Text versteht, ja dass man sogar relativ rasch lernen kann, etwa so zu reden. Die Sätze sind also Sätze einer Sprache, sie enthalten lexikalisches Material dieser Sprache, nach grammatischen Regeln zu Äußerungen zusammengefügt. Ziel der L2-Forschung ist nun, den Erwerb dieses Wissens zu beschreiben, die Faktoren ausfindig zu machen, die Verlauf und Erfolg des Erwerbs bestimmen, und zu zeigen, wie lernersprachliches Wissen über die Entwicklungsstadien hinweg in der Kommunikation verwendet wird. Die kognitiven Probleme, vor die sich der L2-Lerner gestellt sieht, sind, was die sprachliche Seite angeht, denen des Kindes beim L1-Erwerb nicht un-

ähnlich. In kommunikativen Interaktionen nimmt er Schallgebilde wahr, die auf nicht-sprachliche Gegebenheiten in der Situation und auf Vorstellungen von Sachverhalten bezogen sind. Die erste Aufgabe ist also, in den Schallgebilden Einheiten zu identifizieren und ihre Bedeutungen zu erkennen. Um sich selbst in der Zielsprache ausdrücken zu können, muss von diesen Einheiten dann produktiv Gebrauch gemacht werden. Sie müssen zu strukturierten und linearisierten Sätzen verbunden und auf die Gegebenheiten der Kommunikationssituation abgestimmt werden, also angepasst an das Zeitverhältnis zwischen Sprechzeit und der Zeit, über die gesprochen werden soll (Zeitreferenz), auf die Ortsverhältnisse (Raumreferenz) und auf die Personen und Dinge, über die gesprochen wird (Personen- und Objektreferenz). Schließlich muss, um das lernersprachliche Wissen in Richtung des zielsprachlichen Systems zu verändern, dieses immer wieder mit jenem verglichen werden. Eine detaillierte Rekonstruktion der hier nur aufgezählten Erwerbsaufgaben ist Klein (1986, Teil 2). Eine umfassende Gesamtdarstellung des Forschungsstandes ist Ellis (1994).

Zur Vertiefung

Die ESF-Studie

Wie Lerner bei der (unbewussten) Bewältigung dieser Aufgaben vorgehen, ist fast ausschließlich durch linguistische Analysen natürlicher oder quasi-natürlicher lernersprachlicher Daten aus verschiedenen Erwerbsabschnitten herausgefunden worden. Die umfassendste psycholinguistische Untersuchung stellt die longitudinale und sprachvergleichende Studie der European Science Foundation (E. S. F.) dar, die von einem Verbund von Projektgruppen in England, Frankreich, Holland, Schweden und Deutschland in den 1980er Jahren unternommen worden ist. Anlage und Hauptergebnisse der Untersuchung sind in Perdue (1984; 1993) dargestellt.

Erwerbsverlauf: In der Zusammensicht lassen die großen Feldstudien einige allgemeine Reihenfolgen in der Entwicklung der Lernaltersprache erkennen und eine systematische Variation. Generell belegt sind:

- Beim **Lauterwerb** offene Vokale vor geschlossenen; im Konsonantismus anfänglich Entstimmung des linken Silbenrandes und Reduktion am rechten Rand (vgl. Tropic 1983).
- Im **Lexikerwerb** werden nominale Einheiten früher aufgenommen als verbale (vgl. Dietrich 1989; Kauschke 2007). Einheiten mit lexikalischer Bedeutung werden deutlich früher erworben als Einheiten mit grammatikalischer Bedeutung. Generell werden lexikalische Einheiten früher erworben als grammatische Ausdrucksmittel; unflektierte Formen gehen flektierten im Erwerb voraus und der Erwerb kann auf einer Stufe vor dem Auftreten flektierter Formen fossilieren, wie etwa im obigen Beispiel illustriert. Eine detaillierte Fallstudie ist die von Pascual A (SP-25) in Rieck (1987, S. 101–158).
- In der **Entwicklung von Regularitäten des Äußerungsaufbaus** gehen pragmatische und semantische Aufbauprinzipien dem Erwerb syntaktischer Einschränkungen voran (vgl. Klein/Perdue 1992).

L2-Varietäten

Die Entwicklung insgesamt resultiert aus dem Zusammenspiel der Erwerbsschritte in allen sprachlichen Bereichen. So setzt z. B. der Erwerb der Endungsflexion im Deutschen voraus, dass das Wissen um die Struktur komplexer Konsonanten-Cluster am rechten Silben- und Wortrand vorhanden ist. Der Erwerb der Lexik und Syntax der Negation ist strukturell auf den Erwerb der Verbflexion bezogen (vgl. Dietrich/Grommes 1998; Becker 2005). Der Erwerb der bestimmten und unbestimmten Artikel im Deutschen bildet die Voraussetzung für die Ausnutzung der freien Wortstellung im Deutschen und der Erwerb von Zeitadverbien wie *zuvor*, *vorher*, *zugleich* ermöglicht mehr Unabhängigkeit von der zuvor stärker eingeschränkten Linearisierung der Äußerungen in narrativen Texten. Der Wissensaufbau verläuft akkumulativ. Vorhandenes zielsprachliches Wissen wird um neues ergänzt oder durch neues Wissen untergliedert.

Basic Variety: Der Zweitspracherwerb weist eine **universale Stufung** auf und innerhalb des Spielraums dieser Stufung eine ausgeprägte Variabilität. Die herauszuhebende Erwerbsstufe ist die der sog. **Basisvarietät**. Sie wird von fast allen Lernern erreicht, unter günstigen Erwerbsbedingungen nach drei Monaten, unter ungünstigen nach einem Dreivierteljahr. Das lexikalische Wissen umfasst produktiv 50 bis 200 Wörter, deutlich überhöhter Nominalanteil, Verben, Negation, wenige Adverbien, Adjektive, Präpositionen. Die Wortformen sind nicht flektiert; sie werden im Satz in einer unflektierten einheitlichen Grundform verwendet. Bei den Verben im Deutschen ist das überwiegend die Form des Infinitivs ohne das finale »n«, bei den Nomina die Nominativ-Singular-Form. Die Sätze sind kurz. Nomen + Verb oder Nomen + Verb + Nomen oder Nomen + Kopula + Adjektiv oder Nomen. Das agentivischere Nomen steht voran, Fokusinformation am Satzende. Die Erzählung vom Arztbesuch ist typisch für die Basisvarietät. Die Basisvarietät weist wenig Einfluss der Ausgangssprache auf. Die beschriebenen Merkmale sind für alle elf Ausgangs-/ Zielsprachenpaare belegt, die in dem erwähnten ESF-Projekt beobachtet worden sind. Der Basisvarietät voraus geht die Prä-Basisvarietät, eine Sprache, die fast ausschließlich nominales Material umfasst

Beispiel Prä-Basisvarietät

Beispiel

Die folgenden Äußerungen sind Beschreibungen einer Szene aus Harold Lloyds *In the City* durch denselben Lerner wie im Arztbesuch-Beispiel oben, allerdings ein halbes Jahr früher. Die Daten sind verfügbar in »The Language Archive« des Max-Planck-Instituts für Psycholinguistik (vgl. Wittenburg/Broeder/Offenga/Willems 2002).

Abschied

*Eine Mann. Harold Lloyd, Bahnhof.
Bahnhof Freundin de Mann.
Mutter Freundin, Vater Freundin.
Eine schwarze Frau Bahnhof.
etc.*

Post-Basisvarietät: Auf die Basisvarietät folgt die Stufe der Post-Basisvarietät. Den Übergang markiert der Erwerb von grammatischen Ausdrucksmitteln für die Finitheit. Diese Erweiterung des Satzaufbaus und löst entsprechend viele Erwerbsnotwendigkeiten aus. Im Deutschen entsteht eine ganz neue syntaktische Landschaft, das sog. Mittelfeld zwischen Finitum und Infinitum. Die Position der Negation muss neu organisiert werden, infinite Sätze können von finiten unterschieden werden, Tempus kann grammatisch ausgedrückt werden, was die Funktionalität der Zeitadverbien, die zuvor zur Zeitreferenz verwendet worden sind, verändert. So ist die Post-Basis-Stufe ihrerseits wiederum in mehrere strukturell unterscheidbare Schritte gegliedert. Im Bereich der Zeitreferenz z. B. eine anfängliche Phase, in der Präsens, Perfekt und Präteritum anaphorisch verwendet werden, gefolgt von der zielsprachlichen deiktischen Tempusbedeutung, gefolgt von aspektueller Unterscheidung zwischen Perfekt und Präteritum sowie dem Ausbau des Tempussystems im Futurbereich. Mit dem Übergang in die Phase der Post-Basisvarietät wirken sich Faktoren, die Verlauf und Tempo des Erwerbs beeinflussen, stärker aus und die Variabilität der Entwicklung nach Lernergruppen wächst.

Sprachkontrast: Der Einfluss der Ausgangssprache auf die Entwicklung des phonologischen Wissens wird mit dem Ausbau des Konsonantismus und besonders der Silbenränder deutlicher. Unterschiede in der lautlichen Differenziertheit flexionsmorphologischer Kategorien in verschiedenen Zielsprachen wirken sich als inputbedingter Erwerbsfaktor aus. So erwerben spanischsprachige Schwedischlerner die Verbflexion früher als spanischsprachige Französischlerner bei sonst gleichen Erwerbsbedingungen. Ursache ist die lautlich prominentere und differenziertere Endungsflexion des Schwedischen gegenüber den schwach differenzierten und insgesamt unauffälligen Verbflexionsmorphemen des gesprochenen Französisch (vgl. dazu die entsprechenden Kapitel in Perdue 1993, Vol. 2).

Weitere entwicklungsbestimmende Faktoren, die sich allerdings auf Tempo und Erfolg des Erwerbs auswirken und nicht auf Reihenfolge und Struktur der resultierenden Lerner Sprache, sind die Sozialfaktoren: Alter bei Beginn des L2-Erwerbs, Art und Häufigkeit des Kontakts mit der Zielsprache und – besonders im unterrichtlichen Kontext – das Ausmaß der Bekommenheit in der Klassenzimmersituation. Bemerkenswert ist, dass die Dauer des Sprachkontakts sich als Faktor deutlich schwächer auswirkt als Art und Frequenz. Starke Wirkung zeigen nach Befunden im Heidelberger Forschungsprojekt Pidgindeutsch (1977) die Häufigkeit des Sprachkontakts am Arbeitsplatz und in der Freizeit. Besonders für die Entwicklung der mündlichen, spontanen Sprachverwendungsfähigkeit wirkt sich der Faktor Sprachangst (*language anxiety*) aus (vgl. Young 1991).

Theorien des Zweitspracherwerbs

Psycholinguistische Theorien des Zweitspracherwerbs setzen an Beobachtungen der Frühphase (*initial state*, *basic variety*) an. Es werden hauptsächlich zwei Ansätze verfolgt. Sie unterscheiden sich in den Annahmen darüber, von welchem Wissen sich die Lernenden bei der Verarbeitung der

Inputdaten und bei der Konstruktion von Hypothesen über Einheiten und Aufbau der Zielsprache leiten lassen.

Nach der Grundannahme der **funktionalen L2-Erwerbstheorie** sucht der Lerner grundsätzlich Form-Funktionszusammenhänge in der Zielsprache zu erkennen und lässt sich dabei von seiner vom Erstspracherwerb her vorhandenen Kenntnis funktionaler Kategorien leiten. Typische funktionale Kategorien sind das Referieren auf Zeiten, auf Orte, auf Personen und Objekte, das Verneinen, das Unterscheiden von Topik- und Fokuginformation in der Äußerung. Anhand von Querschnittsdaten türkischer bzw. Longitudinaldaten italienischer und türkischer Deutschlerner zeigen Stutterheim (1986) und Dietrich/Klein/Noyau (1995), dass der Erwerb der Temporalitätsausdrücke eine Entwicklung zunehmender zunächst lexikalischer, dann grammatischer Differenzierung des zielsprachlichen Wissens von Form-Funktionszusammenhängen in der L2 ist; am Anfang stehen lexikalische Mittel zum Ausdruck einer zeitlichen Verankerung eines Ereignisses und lexikalische sowie diskursorganisatorische Ausdrucksformen für topologische Zeitrelationen und schließlich die elementare ›und-dann‹-Beziehung. Mittel zur Spezifikation von zeitlicher Distanz und anderer projektiver Zeitrelationen, u. a. Vorzeitigkeit, sowie zur aspektuellen Kennzeichnung folgen. Der Verlauf des Negationserwerbs italienischer Deutschlerner ist nach Becker (2005) wesentlich an der Differenzierung der Verneinungsmöglichkeiten für unterschiedliche Fokus-Hintergrund-Konstellationen orientiert und die Struktur und Entwicklung ist eng mit dem Erwerb der Finitheitsausdrücke verbunden. Nennenswerte Einflüsse des erstsprachlichen Formenwissens sind anfangs nicht zu beobachten. Eine Gesamtdarstellung des funktionalen Ansatzes mit Diskussion bietet Jordens (1997).

**L2-Erwerb
und referentielle
Funktionen**

Universalgrammatikbasierte Erklärungsansätze: Zwar auch auf über-einzelsprachliches Wissen, aber auf angeborenes Wissen von strukturellen grammatischen Prinzipien in Verbindung mit grammatischem Wissen über die Erstsprache wird der L2-Erwerb im **UG-basierten** Erklärungsansatz zurückgeführt. Unter dieser Perspektive wird der L2-Erwerb angesehen als ein unbewusster kognitiver Prozess, die zielsprachliche Grammatik aufzubauen durch Ausnutzen von strukturellen Einschränkungen auf Strukturmerkmale, die an den zielsprachlichen lexikalischen Einheiten erkannt worden sind. In der Frage, wie stark das L1-Grammatikwissen in den Prozess einwirkt, werden unterschiedlich weitgehende Behauptungen diskutiert; die konkurrierenden drei Varianten, die Full-Access-and/Full-Transfer-Hypothese, die Valueless-Features-Hypothese und die Lexical-Categories-Hypothese sind unter Bezug aufeinander dargestellt in Schwartz/Eubank (1996). Nach der FA/FT-Hypothese zieht der Lerner (unbewusst) das gesamte universalgrammatische und das spezifische Wissen über die Festlegung parametrisierter Prinzipien in seiner L1 für seine Hypothesenbildung über die zielsprachliche Grammatik heran. Nach der *valueless-features*-Hypothese wirkt dagegen nur das kategoriale Wissen, nicht hingegen die Merkmalsausprägung der grammatischen Kategorien in L1. Nach der *lexical-categories*-Hypothese schließlich wirkt lediglich die kategoriale Struktur der VP, keine funktionale Kategorie.

**Noch einmal:
Genetik**

3.8 | Gebärdenspracherwerb

Zum Begriff

Mit **Gebärdensprache** bezeichnet man eine natürliche, nicht laut-basierte Sprache, die sowohl über lexikalische als auch grammatisch funktionale Ausdruckseinheiten sowie über Strukturbildungsregeln verfügt. Sie wird von Gehörlosen zur Kommunikation verwendet.

Die Gebärdensprache ist die Sprache der Gehörlosen. Es ist eine natürliche Sprache wie die Lautsprache, und es gibt wie in dieser, einzelsprachliche nationale Gebärdensprachen, z. B. die amerikanische Gebärdensprache, die chinesische, die spanische, die der Maya-Indianer, die deutsche usw. Ihre elementaren lexikalischen Ausdruckseinheiten sind Konstellationen und Bewegungen der Finger, der Hände, der Arme, des Gesichts, des Kopfes und des Torsos im Raum, die motorisch produziert und visuell wahrgenommen werden. Es gibt lexikalisch einfache und komplexe, abgeleitete und zusammengesetzte Gebärden, solche mit sog. lexikalischem Inhalt und grammatisch-funktionale. Ihrer Kombinierbarkeit liegt ein lexikalisch-morphologisches System zu Grunde und die Syntax der Gebärdensprache ist wie in jeder natürlichen Sprache ein System von hierarchisch und lokal bestimmten Relationen, ausgedrückt, wie gesagt, durch eine Folge inhaltlicher und grammatischer Gebärden (vgl. die umfassende und gut lesbare Sammlung grammatischer Skizzen der deutschen Gebärdensprache in Leuninger/Happ 2005). Neben dem lexikalischen Repertoire an Gebärden verfügt die Gebärdensprache über ein sog. **Finger-Alphabet** zum Ausdruck von Buchstaben und Buchstabenclustern, z. B. *sch*. Das ist ein Set von Zeichen, die mit den Fingern einer Hand gebildet werden.

Das Ergebnis des Gebärdenspracherwerbs ist das mentale lexikalische Gebärdenwissen und die Gebärdensyntax verbunden mit dem prozeduralen Wissen und der Beherrschung ihrer aktiven motorischen und passiven visuellen Verwendungsmodalitäten.

Biologische, kognitive und soziale Voraussetzungen: Hinsichtlich der biologischen, kognitiven und sozialen Voraussetzungen für den Spracherwerb unterscheidet sich die Ausgangssituation des gehörlosen Kindes in allem, was durch seine Gehörlosigkeit bestimmt ist und darin, wie seine Umwelt auf die Gehörlosigkeit eingerichtet ist. Angesichts der motorischen Repräsentation und der Übermittlung über den visuellen Kanal kommt dem Vorhandensein normal entwickelter Gliedmaßen und Motorik sowie der Sehfähigkeit als biologischer Voraussetzung des Gebärdenspracherwerbs besondere Bedeutung zu. Auch ist natürlich die gebärdensprachliche Kommunikation darauf angewiesen, dass die Kommunikationspartner einander sehen, so wie die lautsprachliche Kommunikation auf den auditiven Kanal angewiesen ist.

Hinsichtlich der sonstigen biologischen und der kognitiven Voraussetzungen besteht zwischen dem Erwerb der **Gebärdensprache** und dem der **Lautsprache** kein Unterschied. Die biologischen und kognitiven Voraussetzungen der menschlichen Sprachfertigkeit schließen also nicht die Fähigkeit zur lautlichen Artikulation und zur auditiven Wahrnehmung ein;

sie sind vielmehr **modalitätenneutral**. Eine knappe Beschreibung der Phonologie, Morphologie, Syntax und Lexik der Gebärdensprache sowie eine Zusammenfassung des Forschungsstandes mit Nachweisen der Natürlichsprachlichkeit der Gebärdensprache gibt Leuninger (2000).

Umgebungsbedingungen: Ebenfalls eine besondere Rolle unter den Voraussetzungen kommt bei Gebärdenspracherwerb den Umgebungsbedingungen zu, in erster Linie dem Ausmaß der Gebärdensprachbeherrschung seitens der Eltern. Der zeitliche Verlauf und der Erfolg des Gebärdenspracherwerbs durch das Kind werden natürlich entscheidend davon beeinflusst, ob und von welchem Alter des Kindes an die Eltern die Gebärdensprache beherrschen; vgl. dazu den auffälligen Befund einer entsprechenden Erhebung von Mayberry/Lock/Kazmi (2002). Der für das gehörlose Kind in dieser Hinsicht günstigste Umstand, dass nämlich die Eltern infolge eigener Gehörlosigkeit vollkommen kompetente Gebärdensprachler sind, ist nach vorliegenden Daten für weniger als 10 % gehörloser Kinder gegeben (vgl. Leuninger 2000, Kap. 2.2).

Als weiterer wichtiger Umstand hat auf den Gebärdenspracherwerb Einfluss, wie Eltern, Erzieher und sonstige Bezugspersonen des Kindes zur Gebärdensprache als Kommunikationssystem eingestellt sind. Für den Erwerb förderlich ist die Einstellung, dass die Gebärdensprache ein eigenständiges sprachliches System ist, für den Gehörlosen so natürlich wie die Lautsprache für den Nicht-Gehörlosen. Hinderlich für den eigenen Umgang der Erwachsenen und damit hinderlich für die Natürlichkeit des gebärdensprachlichen Verhaltens und seiner Wirkung auf das Kind ist die Einstellung, die Gebärdensprache sei lediglich ein mehr oder weniger unzulängliches Repertoire von Gesten zur Begleitung des (stummen) Artikulierens und zur **Unterstützung des Lippenlesens**.

Evidenz für die Vollwertigkeit der Gebärdensprache als natürliche Sprache liefern nicht nur in zunehmender Zahl die linguistischen Analysen der Gebärdensprache, sondern auch die Ergebnisse der Gebärdenspracherwerbsforschung. Zwar sind die Forschungsmöglichkeiten infolge der vergleichsweise kleinen Zahl möglicher Fallstudien stark eingeschränkt; die aus solchen Studien vorgelegten, explorativ und experimentell gewonnenen Befunde deuten aber jetzt schon auf ein übereinstimmendes Gesamtergebnis hin. Demnach stimmt der Gebärdenspracherwerb mit dem Lautspracherwerb in allen phasentypischen Phänomenen, in der Abfolge und in der Erklärbarkeit überein mit Ausnahme der pränatalen Phase, in der das nicht-gehörlose Kind ab einer bestimmten Zeit über den Körperkontakt mit der Phonation der Mutter bekannt wird.

Belege für die Parallelität mit der lautsprachlichen Entwicklung treten beim sog. nativen gebärdensprachlernenden Kind um den sechsten Monat auf. Das Kind imitiert Hand- und Fingerzeichen der Eltern. Diese Gesten weisen Formmerkmale von Gebärden auf, stellen aber wie die Äußerungen beim Plappern und Babbeln noch keine Gebärden des zielsprachlichen lexikalischen Systems dar. Etwa am Ende des ersten Lebensjahres tritt das erste gebärdensprachliche Wort auf. Wie der Erwerb des Lautsystems weist auch der Gebärdenerwerb ein Nacheinander von Ausdrucksmitteln auf. Hand- und Finger-Anteile der Gebärde werden früher erworben als die Differenzierung des Gebärdenortes. In der Frühphase präferiert

Gebärden sind
SPRACHE

Gebärden als SPRACHE

das Kind die Höhe des Kopfes, die tiefer liegenden Areale des Gebärdenraumes werden später ausgenutzt. Die mit dem Daumen zu artikulierenden Bestandteile werden eher erworben als die mit Mittelfinger und kleinem Finger.

Wie im Lautspracherwerb treten im lexikalischen Repertoire am Anfang komplexe Ausdrücke auf, die vom Kind als feste formulaische Einheiten unanalysiert verwendet werden. Um das zweite Lebensjahr sind Kombinationen von zwei Gebärden zu beobachten und mit Beginn des Erwerbs grammatischer Elemente setzt wie beim Lautspracherwerb ein rasches Wachstum des Wortschatzes ein, der erwähnte Vokabelspurt.

Standardisiertheit: Wie steht es schließlich mit der **Einheitlichkeit des Gebårdenspracherwerbs**? Vom bilingualen Spracherwerb wissen wir, dass sein Ablauf und sein Ergebnis wesentlich von den Umgebungsbedingungen beeinflusst werden. So ist es auch mit dem Gebårdenspracherwerb, weil dort wie hier die Bedingungen stärker variabel sind als beim Lautspracherwerb. Je nach dem Stand der Gebårdensprachbeherrschung durch die Eltern setzt der Erwerb durch das Kind bei Geburt oder eben später ein. Der Zeitpunkt des Beginns wirkt sich auf die Vollständigkeit der Beherrschung entscheidend aus. Native Lerner haben größere Chancen, morphologische Ausdrucksmittel zu lernen als Kinder, deren Erwerb erst im Alter von vier bis sechs Jahren oder später beginnt (vgl. Newport 1990). Auf die Frage von Eltern und Erziehern, welches Kommunikationssystem einem gehörlosen Kind vermittelt werden soll, lässt sich aus den heute vorliegenden Ergebnissen der Forschung antworten, dass der Erwerb der Gebårdensprache zum frühestmöglichen Zeitpunkt und durch möglichst kompetente »Sprecher« der Gebårdensprache unterstützt werden sollte.

Natürliche Sprache oder ikonisches Repertoire bildhafter Gesten? Ein nach wie vor generelles Ziel der Gebårdensprachforschung ist, die grundsätzliche Einsicht zu verbreiten, dass die Gebårdensprache eine **natürliche Sprache** ist und nicht doch im Wesentlichen ein ikonisches Repertoire an bildhaften Gesten und ihre Verwendung eine differenzierte Pantomime. Der Erwerb funktionaler, grammatischer Ausdrücke lässt hierüber Rückschlüsse zu, u. a. der Erwerb von pronominalen Ausdrücken und von Klassifikatoren. Zeitlich ist der Erwerb syntaktischer Mittel vom zweiten Jahr an belegt. Personenreferenzielle Pronomina werden in der amerikanischen Gebårdensprache in Verbindung mit der Verbgebärde dadurch ausgedrückt, dass diese an dem Ort im Gebårdenraum ausgeführt wird, an der der Antezedens eingeführt worden ist. Unter der Annahme von Bildhaftigkeit der Gebärden würde man erwarten, dass Pronomina in Kontexten mit bildhaften Gebärden wie der für die Bedeutung von *geben* leichter gelernt werden als in nicht bildhaften Kontexten, die in der Gebårdensprache mehr als zwei Drittel ausmachen. Das ist nicht der Fall.

Auch die Reihenfolge beim Erwerb von Klassifikatoren stützt die Annahme der Nicht-Bildhaftigkeit der Gebärden. Klassifikatoren, die Klassen von Objekten, also nach Gestalt und Ort eher stabilen, anschaulichen Bedeutungen bezeichnen, wurden entgegen der Bildhaftigkeitserwartung später gelernt als die morphologisch einfacheren Klassifikatoren, die abstraktere Bedeutungen haben wie Klassen von Objektmanipulationen oder

Form- und Größenmerkmale von Objekten (vgl. Leuninger 2000 zu Einzelheiten der einschlägigen Untersuchungen; für einen informationsreichen Überblick vgl. Leuninger 2006).

4 Sprechen

- 4.1 Die Kommunikationsplanung
- 4.2 Die Konzeptualisierung
- 4.3 Die Formulierung
- 4.4 Die Artikulation
- 4.5 Selbstkontrolle und Selbstkorrektur

Sprechen aus der psycholinguistischen Perspektive: Das Sprechen selbst ist – wie auch das Verstehen von sprachlichen Äußerungen – ein Vorgang, dessen wir uns, während wir es tun, zum größten Teil nicht bewusst sind. Wir benutzen die Wörter, bilden Sätze und bewegen Zunge und Lippen unbewusst, automatisch, routiniert, so wie wir beim Aufrechtgehen oder Fahrradfahren die Balance halten.

Wie vieles andere in der Natur ist aber auch der Vorgang der Sprachverwendung ein verblüffend kompliziertes Geschehen, bewerkstelligt von einem so raffiniert gebauten und perfekt funktionierenden kognitiven und motorischen System, dass es bis heute nicht annähernd gelungen ist, es durch einen Automaten zu simulieren. Man kann sich das näherungsweise vor Augen führen, wenn man ein beliebiges Stück aus der Cockpitkommunikation am Anfang von Kapitel 1 sozusagen durch die Brille des Psycholinguisten betrachtet und sich die immer gleiche psycholinguistische Frage stellt, welche kognitiven Zustände und Prozesse dabei in jeder Millisekunde in Funktion sind.

Genau dies ist die Frage, die durch dieses Kapitel leitet, beschränkt auf den Vorgang des Sprechens, beginnend bei der Absicht, die einen zum Sprechen veranlasst bis hin zu den Schallwellen am Ende des Produktionsvorgangs. Es stellen sich also die Fragen: Wie ist das Sprachproduktionssystem aufgebaut und wie laufen die Prozesse des Systems bei der Äußerungsproduktion ab?

Sprachferne, sprachnahe und direkt sprachliche Aktivitäten: Die Feststellung, Sprechen sei ein komplexer Vorgang ist leicht getan und wird bereits durch einige Alltagsbeobachtungen unterstützt. Am Sprechen kann zum Beispiel vielerlei Störendes auffallen. Es kann stören, dass jemand zu viel oder wenig spricht, Gesagtes unzusammenhängend erscheint, Formulierungen gedrechselt wirken, das Redetempo auffallend schnell oder langsam, oder die Aussprache undeutlich ist. Außerdem fallen einem hin und wieder Fehler auf, die ein Sprecher, bemerkt oder unbemerkt, macht.

Dass zu viel sprechen stören kann, setzt die Existenz von Standards innerhalb einer Kommunikationsgemeinschaft voraus, nach denen entschieden wird, ob man in einer gegebenen Situation überhaupt sprechen soll, also eine Art Sprech-/Schweig-Entscheidung. Unzusammenhängende Rede weist auf misslungene gedankliche Planung hin, störende gedrechselte Formulierung auf Fehlgriffe in der aktuellen Wahl sprachlicher Mittel, Undeutlichkeit auf die Existenz von Maßstäben für die Artikulation beim Sprechen. Fehler beim Sprechen können ganze Wörter oder nur Teile von ihnen betreffen. Schon diese kursorische Betrachtung ergibt eine

Die Sprachproduktionsforschung betrifft alle Schritte von der Sprechabsicht bis zu den produzierten Schallwellen

Planung der Kommunikation

Trennung zwischen sprachfernen, sprachnahen und direkt sprachlichen Aktivitäten. Sie unterscheiden sich inhaltlich.

Sprachferne Aktivitäten beim Sprechen: Inhalt der sprachfernen Aktivitäten sind Bedingungen der Kommunikation. Soll ich unter den gegebenen Umständen überhaupt etwas sagen? In welche Richtung will ich durch mein Sprechen die Kommunikation lenken? An welchem Punkt der Kommunikation platziere ich einen Beitrag, an wen (im Fall mehrerer möglicher Adressaten) soll er gerichtet sein?

Sprachnahe Aktivitäten beim Sprechen: Das Ergebnis der kommunikationsstrukturierenden Aktivitäten ist – eine Entscheidung zum Sprechen vorausgesetzt – die gedankliche Skizze einer Kommunikationsabsicht. Sie umfasst zwei Komponenten. Zum einen ist in ihr im Großen und Ganzen die Sache festgelegt, über die sich der Sprecher äußern will, zum anderen, ob diese Sache berichtet, erzählt, begründet oder zum Vollzug eines Ereignisses eine Instruktion gegeben werden soll. In einem psycholinguistischen Gesamtmodell der Textproduktion wird das Ergebnis der Kommunikationsplanung in Reminiszenz an die antike Rhetorik die **Quaestio** genannt (vgl. Stutterheim 1997; 1994; Stutterheim/Klein 1989; Wintermantel 2003).

Makroplanung und Mikroplanung: Den Inhalt der sprachlichen Aktivitäten bilden die Gedächtnisinhalte, die zur Bewältigung der Quaestio aktiviert werden und Einzelheiten der Kommunikationssituation. Ein Redebeitrag enthält nicht alles, was einem Sprecher zu der thematisierten Sache im Gedächtnis ist; es muss also entschieden werden, welche Information für die Rede vorzusehen ist und welche nicht (**Informationsselektion**).

Planung des Inhalts

Rede vollzieht sich in der Zeit, eine Äußerung erfolgt nach der anderen. Somit muss die Gesamtinformation in äußerungsgröße Einzelsachverhalte zerlegt (**Segmentierung**) und diese in eine Reihenfolge gebracht werden (**Linearisierung**). Diese, den Gesamtsachverhalt gliedernden und ordnenden kognitiven Aktivitäten werden als Prozesse der Makroplanung bezeichnet, ein Terminus, der auf Butterworth (1980, S. 159) zurückgeht und von Levelt in seinem Gesamtentwurf des Sprechvorgangs (Levelt 1989) übernommen und ausgearbeitet ist.

Eine zweite Gruppe von sprachnahen Aktivitäten hat die Funktion, die mit einer einzelnen Äußerung verbundene Redeabsicht und den Äußerungsinhalt begrifflich zu strukturieren. Als Inhalt gilt die Sachverhaltsvorstellung, also die Personen und Objekte des Sachverhalts, ihre Eigenschaften und die Beziehung zwischen ihnen (wer tut was mit wem), die zeitliche und räumliche Einordnung und die modale Beziehung des Sachverhalts zur Diskurswelt, also seine Faktizität, Möglichkeit oder Notwendigkeit. Die Selektion dieser Teilinhalte für die Bestandteile einer Bedeutungsstruktur sowie die Abbildung nicht-begrifflicher Sachverhaltskomponenten auf in der Sprache ausdrückbare Begriffe wird als Mikroplanung bezeichnet (vgl. Butterworth 1980, S. 159; Levelt 1989).

Die präverbale Message: Der Output der Makro- und Mikroplanung ist eine begriffliche Struktur, die sogenannte präverbale Message. Makro- und Mikroplanungsprozesse, zusammengenommen auch als Konzeptualisierung der Äußerung bezeichnet, gelten als sprachnahe Prozesse, weil grammatische und lexikalische Gegebenheiten der Sprache, in der die Äußerung schließlich formuliert wird, bei der Verarbeitung eine gewisse Rolle zu spielen scheinen.

Monitoring: Als Sprecher kann man sich selbst offenbar korrigieren, wenn man das, was man gerade sagt oder sagen will, nicht als passend oder als sprachlich inkorrekt beurteilt. Es muss also ein System geben, das das Ergebnis der Äußerungsproduktion mit dem verglichen kann, was geäußert werden sollte. Die Funktion heißt Monitoring. Das Monitorsystem muss Zugang zu dem Geplanten haben und gleichzeitig zu der Bedeutungsstruktur, die tatsächlich produziert worden ist. Zugang zum Geplanten kann der Monitor nur haben, wenn er die ursprüngliche Message kennt; Zugang zu dem, was produziert worden ist, kann er nur haben, wenn ihm das Produkt des Sprechvorgangs, also die Äußerung, durch das eigene Verstehenssystem (d. i. das des Sprechers) über eine Verstehensschleife wieder zugänglich gemacht wird, so dass das Resultat der Äußerung mit der zu Grunde liegenden Message verglichen werden kann.

Direkt sprachliche Aktivitäten des Sprechens: Die direkt sprachlichen Aktivitäten des Sprachproduktionssystems sind alle, die die Informationen der Message in **sprachliche Formen** überführen und diese abstrakten Informationen letzten Endes physikalisch wahrnehmbar machen, also das **Artikulieren** bzw. Schreiben. Um eine begriffliche Struktur durch sprachliche Formen auszudrücken, muss das Produktionssystem natürlich Wissen über diese Formen heranziehen; es muss also Zugang zu dem in Kapitel 2 beschriebenen lexikalischen und grammatischen Wissen haben.

Formulierungsprozesse: Welcher Art sind nun die sprachlichen Formulierungsprozesse und in welchen Teilsystemen sind sie organisiert? Zunächst einmal ist klar, dass es Prozeduren geben muss, die für jeden begrifflichen Bestandteil der Message genau die lexikalische Einheit auffinden und verfügbar machen, die ihn ausdrückt. Da ein Satz aber auch mehr ist, als eine bloße Folge von ein paar Wörtern, muss es ferner Prozeduren geben, die seine syntaktische Form ausarbeiten und zwar so, dass diese der Gesamtbedeutung der Message entspricht. Das Ergebnis dieser **syntaktischen Enkodierung** ist eine syntaktische Struktur, repräsentiert in der Form von Bündeln syntaktischer Merkmale und hierarchischer Beziehungen zwischen ihnen. Die syntaktische Struktur muss den grammatischen Regeln entsprechen, die Merkmalbündel basieren auf den lexikalischen Informationen und dem, was durch syntaktische Prozesse damit geschieht. Eine syntaktische Struktur kann natürlich nicht ausgesprochen werden.

Es muss also zusätzlich ein Prozess stattfinden, der die phonologische Form und die daraus abzuleitende phonetische Kette erzeugt: das ist die Folge von zu artikulierenden Lauten und die Spezifikationen der Intonation. Die Angaben zur phonologischen Form eines Wortes müssen wiederum dem mentalen Lexikon entnommen werden, die Regularitäten der **phonologischen Enkodierung** dem phonologischen Wissen. Die phonetische Kette ist immer noch kognitiv repräsentiert, kann also weder gehört noch gelesen werden. Um die Äußerung für den Adressaten wahrnehmbar zu machen, muss es noch eine Aktivität geben, durch die, ausgehend von der phonetischen Repräsentation, die Artikulation initiiert und gesteuert wird. Inhaltlich ist dieses System offenbar wiederum von der syntaktischen und der phonologischen Kodierung verschieden. Aus dieser rein inhaltlichen Betrachtung würde sich die folgende Gliederung des Sprachproduktionssystems ergeben (s. Abb. 4.1).

Selbstkontrolle
und Selbst-
korrektur

Finden und
Kombinieren von
lautlich noch nicht
spezifizierten
Wörtern

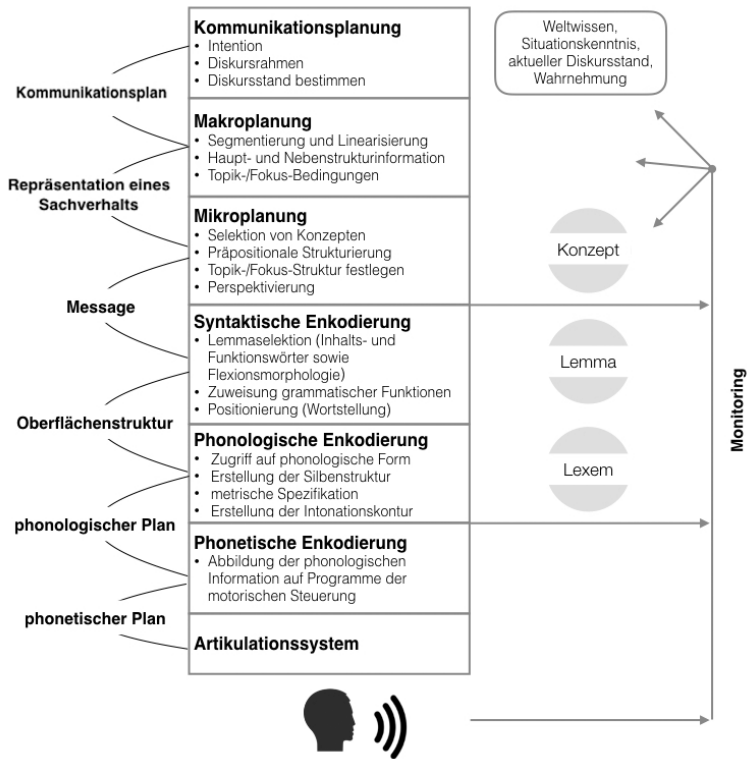


Abb. 4.1: Schema des Sprachproduktionssystems nach Verarbeitungsinhalten gegliedert

Informationsverarbeitungsprozesse in einem Informationsverarbeitungssystem: In den folgenden Abschnitten werden die oben summarisch dargestellten Aktivitäten der Sprachproduktion und ihrer Vorbereitung genauer dargestellt. Das Augenmerk wird darauf liegen, welche Prozesse in dem jeweils betrachteten Verarbeitungssystem stattfinden und wie sie zeitlich mit den anderen Prozessen ablaufen. Unter einem Prozess wird dabei, wenn es nicht anders angemerkt ist, ein **Informationsverarbeitungsprozess** verstanden. In allen Modellen der Sprachverarbeitung wird diese Fähigkeit des Menschen als Informationsverarbeitungssystem aufgefasst (vgl. Carroll 1999; Whitney 1998; Kess 1992; Garman 1990). Ein Informationsverarbeitungssystem wird dadurch charakterisiert, dass angegeben wird, welche Informationen (Inhalt und Form) es als Input akzeptiert, wie sie verändert werden (Prozeduren) und welche Informationen den Output bilden.

4.1 | Die Kommunikationsplanung

Worin besteht die menschliche **Fähigkeit zur sprachlichen Interaktion**? Nach Clark (1996) ist es die Fähigkeit, gemeinschaftlich mit einem Adressaten eine gemeinsame sprachliche Aktivität (*joint activity*) auszuüben

und dies in dem reziproken Bewusstsein eben dieses koordinierten Handelns. Die wesentlichen Besonderheiten des Kommunizierens sind also die Gemeinschaftlichkeit und das Bewusstsein der Kommunikationsteilnehmer.

Der Rahmen der Kommunikation: Das äußerlich beobachtbare Phänomen einer Kommunikation ist der Text des Diskurses. Er kann auf Tonträger, Videoband oder Papier aufgezeichnet und analysiert werden. Durch sog. Diskursanalysen (vgl. Gee 1999; Clark 1996; Hünig 1998) lassen sich nun Annahmen darüber gewinnen, welche Umstände in die Produktion eines Diskursbeitrags (*Turn*) eingegangen sind. Das ultimative Ziel der diskursanalytischen Forschung ist, regelhafte Zusammenhänge zwischen den Umständen kommunikativen Handelns und Formmerkmalen von Diskursen zu finden.

In diesem Zusammenhang kann z. B. festgestellt werden, dass die Form eines Diskurses durch Verfahrensregeln stark vorgegeben sein kann, wie bei der standesamtlichen Trauung. Sie kann jedoch auch offen sein. Die Kommunikationsteilnehmer können für sich selbst agieren wie Schauspieler einer Theateraufführung oder Nachrichtensprecher im Fernsehen, die als Vermittler tätig sind. Unterschieden werden kann auch, ob die Kommunikation eine mündliche oder schriftliche Form hat.

Die Gesamtheit der Umstände, die sich auf Diskursmerkmale auswirken, werden Diskursrahmen (*discourse setting*) genannt. Clark (1996, S. 8) unterscheidet hinsichtlich des Grades an Vorgeprägtheit des Kommunikationsgeschehens und der Rollen der Teilnehmer sieben Typen von Diskursrahmen und innerhalb von jedem wieder mündliche und schriftliche Formen. Die sieben Typen sind dem nachfolgenden Kasten zu entnehmen.

Einfluss der Rahmenbedingungen auf die inhaltliche Planung

Typen von Diskursrahmen (vgl. Clark 1996, S. 8)

Zur Illustration

Unterschieden wird je nach Institutionalisierung des Ablaufs und der Teilnehmerrollen

Typ	Beispiel mündlich	Beispiel schriftlich
Persönliche Kommunikation	A unterhält sich mit B.	A schreibt B einen Brief.
Nichtpersönliche Kommunikation	Dozentin A hält eine Vorlesung vor dem Auditorium B.	Reporter A schreibt einen Artikel für die Lesergruppe B.
Institutionelle Kommunikation	Verteidiger A verhört den Zeugen B vor Gericht.	Unternehmer A schreibt dem Kunden B einen Geschäftsbrief.
Präskriptive Kommunikation	Bräutigam A legt vor Zeugen ein ritualisiertes Versprechen gegenüber der Braut B ab.	A unterschreibt bei einem Notar ein offizielles Formular für B.
Fiktionale Kommunikation	A führt vor dem Publikum B ein Stück auf.	Autor A schreibt einen Roman für die Leserschaft B.
Mittelbare Kommunikation	C dolmetscht für B, was A zu B sagt.	Ghostwriter C schreibt für A ein Buch an die Leserschaft B.
Selbstgerichtete Kommunikation	A redet zu sich selbst über geplante Schritte.	A notiert für sich selbst geplante Schritte.

Worin bestehen nun die relevanten Umstände der verschiedenen Typen von Diskursrahmen? Diese leitet Clark von dem prototypischen Kommunikationsgeschehen, dem persönlichen Gespräch ab.

Rahmenbedingungen des persönlichen Gesprächs (vgl. Clark 1996, S. 9)

- Gleichzeitige räumliche Gegenwart der Teilnehmer (Ko-Präsenz)
- Blickkontaktmöglichkeit
- Hörkontaktmöglichkeit
- Unverzögerlichkeit; d. h. die Teilnehmer können ihre Aktivitäten ohne zeitliche Verzögerung wahrnehmen.
- Flüchtigkeit; d. h. die akustische Seite des Diskurses ›verweht‹ rasch.
- Spurlosigkeit; die Kommunikation hinterlässt keine Aufzeichnungen, kein Objekt.
- Gleichzeitigkeit; die Teilnehmer können unverzüglich und gleichzeitig sprechen und hören.

Die **Rahmenbedingungen** des persönlichen Gesprächs sind insofern elementar, als sich die anderen **Diskursrahmen** ex negativo bestimmen, also durch das Fehlen einer oder mehrerer davon. So ist dem Telefonanrufer bewusst, dass das Setting der Telefonkommunikation sich in zwei wesentlichen Umständen von dem des persönlichen Gesprächs unterscheidet: Die Gesprächsteilnehmer haben keinen Blickkontakt und sie befinden sich nicht am selben Ort. Der Sprecher hat also damit zu rechnen, dass visuelle Ausdrucksmittel wie Gestik und Mimik vom Adressaten nicht wahrgenommen werden und dass andere Bedingungen der lokalen Deixis gegeben sind. Der Anrufer identifiziert sich explizit; durch eine längere Gesprächspause entsteht Veranlassung zu prüfen, ob der Kontakt noch besteht (»Bist du noch da?«) etc. (vgl. Hopper 1992; Schegloff 1979).

Die erste kommunikative Planungsarbeit besteht also darin, dass der Sprecher die Besonderheiten des jeweils gegebenen Diskursrahmens zur Kenntnis nimmt und die Produktion seines Diskursbeitrags darauf abstimmen muss.

Absicht des
Sprechers

Die Bestimmung des aktuellen Kommunikationsziels: Die zweite wesentliche kommunikative Planungsarbeit ist die Bestimmung des aktuellen Kommunikationsziels. Ziele können mehr oder weniger abstrakt sein und entsprechend mehr oder weniger generelle Folgen für die Formulierung des Diskurses haben. Das Ziel, einem Adressaten den Weg vom Alexanderplatz zum Bahnhof Zoologischer Garten zu beschreiben, wirkt sich auf den Aufbau des gesamten Diskurses aus. Um es zu realisieren, ist eine **Vorstellung** des gesamten Weges erforderlich und eben die **globale Absicht**, dem Adressaten zu verstehen zu geben, wie er sich bewegen muss, um vom Ausgangspunkt zum Zielpunkt zu gelangen. Ersteres bildet den Inhalt des Diskursziels, letzteres die globale **Intention**.

Je nach Umfang des Inhalts ist für seine Vermittlung mehr als eine Äußerung erforderlich und da jede Äußerung für sich genommen zum Erreichen des globalen Ziels beiträgt, ist sie inhaltlich und intentional auf das globale Ziel in einer für den Adressaten erkennbaren Weise bezogen. In

einer Wegbeschreibung wird es also eine Reihe von Anweisungen geben, von da nach da (weiter-) zu gehen oder zu fahren, und Erläuterungen zu markanten Wegmarken wie Kreuzungen, Gebäuden, Ampeln etc.

Die äusserungsspezifischen Intentionen heißen **Äußerungsabsicht**; die bekannteste Theorie darüber ist die Sprechakththeorie, in der die intentionale Komponente der Äußerung illokutionärer Akt, kurz auch Illokution heißt (vgl. Searle 1969). Die Umsetzung des globalen Ziels in Illokutionen geschieht im Konzeptualisierungsprozess.

Kontrolle und Anpassung des Kommunikationsverlaufs: Die dritte kommunikative Planungsaktivität ist die laufende Koordination des Zusammenspiels zwischen Sprecher und Adressat. Zu ihr gehören eine große Zahl von Kontroll- und Anpassungsmaßnahmen (zu Einzelheiten vgl. Clark 1996, besonders Kap. 2–4). Der Sprecher muss die Informationen, die er auswählt, für den Adressaten verständlich machen, was in erster Linie erfordert, dass er diese mit schon vorhandenen Informationen im Wissen des Adressaten verknüpft. Um das zu gewährleisten, muss der Sprecher also Mutmaßungen über das von ihm und dem Adressaten geteilte Wissen anstellen und berücksichtigen; Anhaltspunkte dafür liefern ihm ggf. Kenntnisse über die Person des Adressaten (Beruf, Alter, Geschlecht, Wohnort etc.), Äußerungen des Adressaten und natürlich gemeinsame Erlebnisse sowie der vorangegangene Teil des aktuellen Diskurses. Ob die Kommunikation in diesem Sinne Verständigung erzielt, wird vom Sprecher laufend kontrolliert, und der Adressat ist sich dieser Notwendigkeit auch bewusst und liefert dem Sprecher Rückmeldung dazu (Kopfnicken, bestätigende verbale Reaktionen oder eben ein Stirnrunzeln), sogenannte **Feedbacksignale**.

Turn Taking: Charakteristisch für den prototypischen Fall von Kommunikation, das persönliche Gespräch, ist der **Wechsel von Sprecher- und Adressatenrolle**. Durch ihn wird besonders augenfällig, dass Kommunizieren ein gemeinschaftliches Handeln ist, denn beim Wechsel der Rede von Teilnehmer A zu B wirken beide auf eine subtile Weise zusammen.

Zunächst einmal müssen alle Beteiligten erkennen, dass überhaupt ein Punkt im Diskurs gekommen ist, an dem ein Wechsel geschehen kann. Ferner muss entschieden werden, ob ein Wechsel geschehen soll, und schließlich – bei mehr als zwei Beteiligten – muss klar sein, wer die Sprecherrolle übernimmt.

Der Sprecherwechsel ist ein Bestandteil der Kommunikationsdynamik und er ist in eine ganze Reihe von dynamischen Gliederungen eingebettet. Eine grundlegende ist, dass der Informationsfluss im Diskurs in zeitliche Abschnitte gegliedert ist, die im Wesentlichen inhaltlichen Blöcken entsprechen. Mit dem Begriff ›**Turn**‹ wird ein zusammenhängender Beitrag eines Sprechers bezeichnet. Dieser kann einen oder mehrere solcher Abschnitte, Phasen genannt, umfassen.

Turnwechsel findet dadurch statt, dass am Ende einer Phase von den Beteiligten ein **Turnübergabepunkt** erkannt wird und ein anderer als der Sprecher die Rede aufnimmt. Er kann dazu vom Sprecher veranlasst sein oder es von sich aus tun. Natürlich kann ein Gesprächsteilnehmer auch im laufenden Turn eines anderen den Turn übernehmen wollen. Dies erfordert dann den Einsatz spezifischer sprachlicher und/oder gestischer Signale (vgl. dazu im einzelnen Clark 1996, S. 320 ff.; Sacks/Schegloff/Jefferson

Einfluss von
Feedback auf die
weitere Planung

Wechsel von
Sprecher- und
Hörerrolle

son 1974). Für den, der nun das Wort hat, stellen sich damit drei Aufgaben. Er muss zu erkennen geben, dass er den Vorgängerbeitrag verstanden hat, an ihn anknüpfen und mit dem eigenen Beitrag den Gesprächsverlauf gestalten, und ggf. beenden. Wie in einem Text muss also auch im Gespräch Kohärenz hergestellt werden. Ein theoretischer Rahmen zur Analyse dieser Prozesse ist in Dietrich/Grommes (2003) vorgeschlagen.

4.2 | Die Konzeptualisierung

Zum Begriff

Als **Konzeptualisierung** bezeichnet man alle Sprachplanungsaktivitäten, deren Inhalt die Informationsselektion und -organisation betrifft.

Informationsselektion

Das Ergebnis der Kommunikationsplanung kann offenbar nicht direkt in Schallwellen umgesetzt werden. Wie in Abbildung 4.1 skizziert, müssen Prozesse angenommen werden, die das Ergebnis der kommunikativen Planung in Formate überführen, die durch die **Versprachlichungsvorgänge** bearbeitet werden können. Unser Modell des Konzeptualisierens gibt nur eine grobe Vorstellung von den Vorgängen selbst und sagt gar nichts über ihr Zusammenspiel in der Zeit. Die derzeitigen Kenntnisse darüber sind lückenhaft, und die Hypothesen gehen beträchtlich auseinander. In Einzelheiten am weitesten ausgearbeitet und im Zusammenhang modelliert sind die Konzeptionen von Levelt (1989), Herrmann/Grabowski (1994) und Stutterheim (1997), ferner Teiltheorien einer umfassenderen Konzeption von den Denkvorgängen insgesamt, der sog. **Mental-Model-Theorie** von Johnson-Laird (1983); vgl. auch Garnham (2001).

Auswahl zu versprachlichender Konzepte

Alle **Theorien des Konzeptualisierens** beruhen auf Beobachtungen darüber, wie die kommunikative Aufgabe mit Struktureigenschaften des produzierten Diskurses zusammenhängt. Natürlich kann dies genauer beobachtet werden, wenn die kommunikative Aufgabe unabhängig vom resultierenden Text bekannt ist oder sogar im Hinblick auf spezifische Analyseziele in der Datenerhebung kontrolliert und verändert werden kann. Diese Voraussetzung ist unter den Bedingungen eines persönlichen Gesprächs und generell unter normalen Kommunikationsumständen nicht gegeben. Daher werden, um trotzdem zu geeigneten Beobachtungen zu kommen, Sprecher in mehr oder weniger formalisierten Settings gezielt vor bestimmte Aufgaben gestellt und die resultierenden Texte systematisch auf aufgabenabhängige Variation in ihrem Aufbau hin analysiert. Das führt dazu, dass die Annahmen über Konzeptualisierungsvorgänge im Wesentlichen auf monologischen, mündlich produzierten Diskursen beruhen. Typische Aufgaben sind Erzählungen von eigenen Erlebnissen (Labov/Waletzky 1967), Nacherzählungen von Filminhalten (Chafe 1980; Herrmann/Grabowski 1994, S. 243 ff.), Beschreibungen von selbst beobachteten komplexen Ereignissen, Bastelanweisungen (Kohlmann 1996), Nacherzählen von Bildergeschichten (Stutterheim/Klein 1989).

Einfluss der Sprachverwendung auf die Informationsselektion: Welche Einzelheiten eines früher wahrgenommenen Geschehens ein Sprecher bei der Bearbeitung einer späteren kommunikativen Aufgabe wiedergibt und wie, wird schon durch die Einstellung der Aufmerksamkeit bei der Aufnahme des Geschehens in sein Wissen beeinflusst, bei der sog. **Kognizierung** (vgl. Stutterheim 1997, S. 135 ff.). Bedeutet dies auch, dass sich die Grammatik und die verfügbaren lexikalischen Mittel einer bestimmten Sprache auf die Sprachplanung auswirken? Für diese Annahme einer sogenannten **L-Perspektivierung** sprechen gut dokumentierte Beobachtungen von Textproduktionen durch Sprecher des Deutschen, Englischen, Französischen und anderer Sprachen. Sie zeigen einen systematischen Zusammenhang zwischen der Planung der zeitlichen und der räumlichen Konzeptualisierung einerseits und grammatischen Unterschieden der jeweiligen Sprachen andererseits (zum Konzept der L-Perspektivierung vgl. Stutterheim/Klein/Graumann/Kallmeyer 2002; zu exemplarischen Untersuchungen vgl. Stutterheim/Nüse 2003).

Des Weiteren deutet eine ganze Reihe von Blickbewegungsstudien darauf hin, dass sowohl das lexikalische als auch das grammatische Repertoire einer Sprache einen Einfluss auf die Aufmerksamkeitssteuerung bei der Verarbeitung visueller Stimuli hat, was in einen direkten Zusammenhang mit den sprachlichen Planungsaktivitäten gesetzt wird. Insbesondere im Bereich der Bewegungsereignisse konnte gezeigt werden, dass der Unterschied zwischen Sprachen, die Informationen über den räumlichen Pfad einer Bewegung im Verb selbst oder in einem verbbegleitenden Element ausdrücken, Blicke auf entsprechend relevante Bereiche des Stimulus steuert (vgl. Papafragou/Hulbert/Trueswell 2008; Trueswell/Papafragou 2010; Stutterheim et al. 2012; Flecken/Gerwien/Carroll/Stutterheim 2015).

Auslassen rekonstruierbarer Informationen: Nicht alles, was der Adressat verstehen soll, wird im Diskurs sprachlich ausgedrückt. In einer von einer Versuchsperson produzierten Bastelanweisung (Stutterheim 1997: 95) heißt es »*Man nimmt zuerst [...] einen grünen Bauklotz, tut die grüne Schraube durchn Loch [...]*«. Verstanden wird hier üblicherweise, dass die grüne Schraube durch ein Loch im grünen Bauklotz gesteckt werden soll; das ist nicht gesagt, wird aber aufgrund einer Schlussfolgerung verstanden. Diese kontextbezogene Folgerung wird in der Sprachverstehensforschung **Bridging** genannt (vgl. Singer 1993). Dass der Sprecher systematisch Information implizit lässt, lässt vermuten, dass er bei der Planung seines Diskurses auf *Bridging*-Prozesse des Adressaten baut. Das bedeutet, dass beim Selektionsvorgang als Teilprozess des Konzeptualisierens (a) über den Satz hinausgeschaut wird und (b) der jeweils aktuelle Diskursstand des Adressaten kalkuliert wird.

Implizite
Informationen

Informationsorganisation

Segmentierung: Es war bereits die Rede davon, dass die für die Verbalisierung ausgewählten Informationen portioniert werden müssen, damit sie auf einzelne Äußerungen abgebildet werden können. Durch diesen Prozess, die Segmentierung, entstehen Informationseinheiten, die in etwa einem Einzelsachverhalt entsprechen und sich prinzipiell auf genau einen

Portionierung
von Inhalten

Satz abbilden lassen. Jedes so entstandene Segment, wird als eine sog. **Message** repräsentiert, indem jede darin enthaltene Information je nach referenziellem Bereich, Modalität, Person/Objektbereich etc. in eine Message-Komponente umgeformt wird. Die Entscheidung, welche Information welche Komponente bildet und wie sie begrifflich ausbuchstabiert wird, wird von dem Diskursrahmen, dem aktuellen Diskursstand und von Bedingungen der Äußerungssituation beeinflusst.

Linearisierung: Ein Diskurs ist eine Folge von in etwa satzgroßen Informationspaketen. Ihrer Reihenfolge (lineare Ordnung) liegt ein **Linearisierungskriterium** zu Grunde. Das kann in der Natur des konzeptualisierten Ereignisses liegen oder eine Folge von Kapazitätsbeschränkungen im Arbeitsspeicher sein. Ein typisches Beispiel für die erste Kategorie ist die Erzählung. Die Äußerungen folgen einander in der chronologischen Reihenfolge der erzählten Ereignisse (**Prinzip der natürlichen Ordnung**). Sachverhalte, deren natürliche Eigenschaften kein hinreichendes Linearisierungskriterium abgeben, werden durch Projektion auf eine Zeitachse chronologisch gegliedert, beispielsweise Wegbeschreibungen durch eine vorgestellte Wanderung den Weg entlang (Herrmann/Schweizer 1998) oder durch prototypische Schemata, Körperschema für die Personenbeschreibung, Dorfschema u. a., für die eine Reihenfolge konventionalisiert ist: Personen werden – nach knapper Aussage über die Gesamtstatur – von oben nach unten beschrieben, Wohnungen durch die Vorstellung eines imaginären Gangs durch die Räume, Einzelräume mittels einer Blicktour durch den Raum linearisiert (vgl. Linde/Labov 1975; Ullmer-Ehrich 1982).

Levelt zeigt, dass Sprecher derartige Informationen nach einem globalen prozeduralen Prinzip linearisieren, dem **Prinzip des geringsten Aufwandes** (Levelt 1982, S. 211). Die Linearisierung bearbeitet offenbar auch Informationen auf Diskurs- und nicht nur auf Äußerungsebene, gehört also zur Makroplanung.

Referenzsemantische Gliederung: Das Resultat, über die Entscheidung des Was pro Komponente (Message), wird die **referentielle Besetzung genannt**. Die Informationen der Message sind also referenzsemantisch gegliedert. Es gibt immer eine Komponente, die einen Zustand, eine Aktion bezeichnet (*Es regnet*), meistens eine oder mehrere Objekte/Personen, die den Zustand/die Aktion realisieren (*Die See tobt.*), oft weitere Objekte/Personen, die in dem Zustand/der Aktion eine Rolle spielen (*Der Purser meldet dem Kapitän das O. K. für die Kabine.*). Diese Referenzen auf Personen/Objekte und Zustände/Aktionen (inkl. evtl. weitere Beteiligte) bilden den sog. **propositionalen Kern**. Hinzu kommen Informationen zur zeitlichen Einordnung des Sachverhaltskerns (*Dann haben die Piloten beide zu bremsen versucht.*) und zur räumlichen Einordnung des Sachverhalts (*Dann sind sie auf 3.800 Fuß gesunken.*). Am wenigsten bemerkt, weil am unauffälligsten sprachlich ausgedrückt, sind noch zwei weitere Informationskomponenten, die Modalitätskennzeichnung (*Brems mal mit*; Modalinformation im Imperativ: Notwendigkeit) und die Sprachhandlungsinformation (*Was machen wir jetzt?* hier: Interrogation), auch Satzmodusinformation genannt.

Perspektivierung: Die Entscheidungen über das Wie des Message-Aufbaus heißen Perspektivierung. Beispiele für unterschiedliche Perspektivie-

Reihenfolge von
einzelnen
Inhaltskompo-
nenten

rungen (und entsprechend unterschiedliche referentielle Besetzungen) sind die folgenden Sätze:

- (4-1) *Das Flugzeug wurde vom Seitenwind gedreht.*
- (4-2) *Der Seitenwind hat das Flugzeug gedreht.*
- (4-3) *Der Seitenwind bewirkte eine Drehung des Flugzeugs.*

Ein interessanter empirischer Befund ist, dass die Entscheidung über die Art und Weise der Perspektivierung in unterschiedlichen Sprachen unterschiedliche Präferenzen aufweisen kann (vgl. u. a. Talmy 1985; Slobin 1991; Stutterheim/Nüse 2003; Gullberg 2011; Fausey/Boroditsky 2011).

Verschiedene Perspektivierung in unterschiedlichen Sprachen?

Experiment

In einem Experiment von Fausey/Boroditsky (2011), wurden Versuchspersonen mit Muttersprache Spanisch bzw. Englisch Filme von zwei Typen von Ereignissen präsentiert mit der Aufforderung, sich die Ereignisse gut einzuprägen. In einer Gruppe der Filme waren von Akteuren absichtsgelenkte Ereignisfolgen dargestellt, in der anderen Gruppe eher unabsichtlich entstandener Ereignissequenzen. Anschließend wurde die Probanden eine sprachliche und eine Erinnerungsaufgabe gestellt. Sie sollten den Hergang der beobachteten Ereignisfolgen a) nacherzählen und b) die Ereignisse aus dem Gedächtnis frei aufzählen. In den Daten beider Aufgaben zeigte sich die Auswirkung der jeweiligen Muttersprache: In den Nacherzählungen der agentivischen Ereignisfolgen zeigte sich kein Spracheffekt, in den Nacherzählungen der akzidentiellen dagegen ein signifikant größerer Anteil an intendierten Szenenbeschreibungen bei den Englischsprachigen gegenüber den Spanischsprachigen. Dem entsprachen auch die Befunde der Erinnerungsaufgabe. Das bestätigt die Annahme eines sprachlichen Einflusses sowohl auf die Kognizierung als auch auf die Versprachlichung von Sachverhalten.

Kohärenz: Die Informationen in einem Diskurs sind vernetzt und das macht den Diskurs kohärent. Kohärenz ist eine wesentliche Eigenschaft des Diskurses. Sie ist das Resultat eines Ensembles von Verknüpfungen von Informationen über die Äußerungsgrenzen hinweg. Die wesentlichen Verknüpfungsarten sind: Informationserhalt (*Hier schicke ich Dir ..., Du kannst ...*), Neueinführung (*Als erstes haben wir ein paar alte Titel von Clapton aufgelegt. Dann fing Walter mit Elton John an.*), Wiederaufnahme (*... Clapton ist mir irgendwie näher.*), und schließlich eine Kombination von Erhalt- und Neueinführungsbezeichnung, die Verschiebung (*Du gehst über die Leipziger Straße, dann rechts, dann 100 m geradeaus, dann an der Ampel links.*), hier eine Verschiebung von Intervall zu Intervall. Der eigentlich aufregende Befund besteht nun wiederum darin, dass die Art und Weise, wie Kohärenz im Diskurs jeweils zustande gebracht wird, durch den Inhalt der kommunikativen Aufgabe und durch die globale kommunikative Absicht beeinflusst ist. Die Erzählaufgabe legt durch die chronologiebasierte Linearisierung eine Verschiebung der Zeit von Hauptstruktur-

Verknüpfung von
Einzelinhalten

zu Hauptstrukturäußerung nahe und einen durchgehenden Erhalt der Modalitätsinformation ›faktisch‹. Die Gesamtheit der Verknüpfungen vom Anfang des Diskurses bis zum Ende wird **referentielle Bewegung** genannt. Der Befund lässt sich also kurz formulieren: Die kommunikative Aufgabe beeinflusst die referentielle Bewegung und ist damit ein Regulativ der Kohärenzbildung. Die kohärenzstiftenden Planungsaktivitäten gehören also auch zur Makroplanung.

Topik-Fokus-Unterscheidung: Zusätzlich zu ihrer Gliederung in referentielle Komponenten ist die Information einer Message hinsichtlich der Topik/Fokus-Unterscheidung gegliedert. Die Topik-Fokus-Gliederung der Diskursinformation wird von der kommunikativen Aufgabe beeinflusst. Information, die in der kommunikativen Aufgabe gegeben ist, gehört zur Topikinformation, solche, die im Diskurs zu spezifizieren ist, zur Fokusinformation. Da nun aber die Topik-Fokus-Gliederung eine Eigenschaft der Message ist, muss angenommen werden, dass im Zuge der Konzeptualisierung aus der kommunikativen Aufgabe systematisch eine, sagen wir, lokale, äußerungsspezifische Aufgabe abgeleitet wird, die die Informationsverteilungsvorgaben der globalen kommunikativen Aufgabe sowie die im Diskursstand eingeführten Informationen berücksichtigt.

Zur Illustration **Beispiel Topik-Fokusgliederung**

Kommunikative Aufgabe: Wie muss man das Orangenbäumchen behandeln?

- 1 *Man stellt es an einen hellen, sonnigen Platz,*
- 2 *gibt ihm zweimal pro Woche etwas Wasser.*
- 3 *Das gießt man von oben auf die Erde.*

Die kommunikative Aufgabe gibt vor: die Modalitätsinformation (Notwendigkeit), das Objekt (*Orangenbäumchen*), eine generische Person (*man*) und ein allgemeines Konzept *behandeln*. Fokussiert sind die einzelnen Behandlungsmodalitäten und Maßnahmen. Im Diskursstand am Ende von (2) ist das Gießen mit Wasser als Fokusinformation eingeführt, und die Information steht durch diesen Schritt für eine lokale Äußerungsaufgabe zur Verfügung: Wie wird das Wasser verabreicht? *Das* in (3) erfüllt damit die Bedingung, Topikinformation zu sein.

Zum Begriff

Die **präverbale Message** ist ein gegliedertes Ensemble von inhaltlich verschiedenen Informationskomponenten. Sie ist das Resultat von Informationsselektions- sowie Informationsorganisationsprozessen.

Schritte der Konzeptualisierung im Überblick:

- Die aufgabenbezogene Information aus dem Langzeitgedächtnis oder aus der aktuellen Wahrnehmung bereitstellen (Informationserschließung);
- die unter den gegebenen Diskursbedingungen relevante Information bestimmen (Informationsselektion; erfordert Zugang zur kommunikativen Aufgabe, zu Annahmen über das Hörerwissen und ggf. zu konventionellen Gliederungsschemata);
- die Gesamtinformation in Einzelsachverhalte einteilen und dabei die Reichhaltigkeit der Antwort, die Granularität, einstellen (Segmentierung; erfordert ebenfalls Zugang zur kommunikativen Aufgabe, zu Annahmen über Hörerwissen und den jeweils aktuellen Diskursstand);
- Die Reihenfolge der Einzelsachverhalte im Diskurs festlegen (Linearisierung; erfordert auf jeden Fall Zugang zum Inhalt des komplexen Gesamtsachverhalts, also zum Gedächtnis und zum Diskursstand);
- je Sachverhalt kalkulieren, ob er Haupt- oder Nebenstrukturinformation darstellt (Haupt-/Nebenstrukturkennzeichnung; erfordert Zugang zur kommunikativen Aufgabe);
- die Informationen je Sachverhalt als Topik- oder Fokusinformation markieren (Topik/Fokusgliederung; erfordert Zugang zur kommunikativen Aufgabe);
- die Message vorbereiten, also alle Informationen zu referentiellen Komponenten zusammenbauen und jede Komponente auf den Kontext und die Situation einrichten, also Vorerwähntheit, sonstige Kohärenzbedingungen, zeitliche räumliche Deixis etc. (Perspektivierung);
- Message aufbauen, also alle Informationen auf Konzepte abbilden und in der Message-Struktur, d. h. in Relation zu anderen Komponenten (Propositionalisierung; erfordert Zugang zu der perspektivierten, z. T. vorbegrifflichen Information und zum begrifflichen Wissen des Sprechers).

4.3 | Die Formulierung

Als **Formulierung** bezeichnet man alle Prozesse, die eine begriffliche Struktur (Message) in ein artikulierbares Format umwandeln. Dazu zählt die Aktivierung und Selektion von Lemmata, der Aufbau einer syntaktischen Struktur (syntaktische Enkodierung), der Abruf und die Verarbeitung phonologischer Informationen (phonologische Enkodierung) sowie die Ausarbeitung eines phonetischen Plans (phonetische Enkodierung).

Zum Begriff

4.3.1 | Überblick der ablaufenden Prozesse

Mit der Frage, wie der Gedanke als sprachliche Äußerung hörbar oder lesbar wird, kommen wir zu den Prozessen, die eine begriffliche Struktur in Schallwellen bzw. eine schriftliche Repräsentation umformen. Als Ober-

begriff für alle diese Prozesse zusammengefasst wird in der aktuellen Forschung der Begriff »**Formulierung**« (engl. *formulation*) verwendet.

Die Formulierung umfasst:

- die Auswahl benötigter Lemmata
- Strukturbildungsprozesse, die als syntaktische Enkodierung bezeichnet werden
- die Aktivierung phonologischer Segmente und Eigenschaften
- die Vorbereitung eines »Programms« für die Steuerung der für die Artikulation notwendigen Muskelgruppen

Auswahl von
Wörtern

Lexikalischer Zugriff: Obwohl noch bei Weitem nicht alle Einzelheiten aufgedeckt sind, führen linguistische Erkenntnisse über den sprachlichen Aufbau der Äußerung in Verbindung mit Verhaltensbeobachtungen, Erwerbsprozessen und Störungsanalysen zu einem differenzierten Bild der Komponenten, die an der sprachlichen Enkodierung beteiligt sein müssen (s. Kap. 2). Demnach macht der Sprecher von seinem lexikalischen und grammatischen Wissen Gebrauch. Das lexikalische Wissen ermöglicht, zu einem Konzept der Message das Wort zu identifizieren, das das Konzept ausdrückt; der Prozess heißt **lexikalischer Zugriff** (*lexical access*). Die Gliederung der lexikalischen Informationen in Lemma und Lexem bringt es mit sich, dass der lexikalische Zugriff zwei Teilprozesse umfasst, den **Lemma-Zugriff** und den **Lexem-Zugriff**. Eine für die psycholinguistische Forschung wichtige Fragestellung betrifft die zeitliche Koordinierung dieses zweifachen Lexikonzugriffs, auf die weiter unten genauer eingegangen wird.

Abruf
syntaktischer
Informationen

Syntaktische Enkodierung: Das syntaktische Wissen ermöglicht es dem Sprecher, die Wörter und die grammatischen Bestandteile des Satzes (Tempus, Modus, Kasus, Numerus etc.) so miteinander zu verknüpfen und in eine Reihenfolge zu bringen, dass das ausgedrückt wird, was verstanden werden soll, also zum Beispiel, dass für einen gegebenen Sachverhalt die Rollenverteilung zwischen Handelndem und Handlungsempfänger korrekt wiedergegeben wird (*Oskar erschreckt Leo* und nicht *Leo erschreckt Oskar*). Der Prozess heißt syntaktische Enkodierung. Klar ist, dass die syntaktische Enkodierung sowohl auf der Ebene der Phrase als auch auf der Ebene des Satzes eine Rolle spielt. Zum Beispiel erfordert der Aufbau einer komplexen Nominalphrase, wie *das schöne alte Haus, das wir gestern beim Spazierengehen gesehen haben* spezifisches syntaktisches Wissen, welches alle zu verwendenden Wörter in die richtige Reihenfolge bringt, und zwar so, dass die semantischen Beziehungen zwischen den Wörtern korrekt abgebildet werden. Dass zum syntaktischen Wissen auch Wissen über den Aufbau einzelner Phrasen gehört, ist daran zu erkennen, dass übersetzungsäquivalente Phrasen in verschiedenen Sprachen verschieden strukturiert sind. Im Spanischen zum Beispiel erscheinen Adjektive in der Regel nach dem Nomen. Im Mandarin und Koreanischen stehen Relativsätze vor dem Nomen.

Funktionale Verabeitung: Im Rahmen psycholinguistischer Untersuchungen sind in Hinblick auf die syntaktische Enkodierung insbesondere zwei zusammenhängende aber dennoch zu unterscheidende Aspekte immer wieder diskutiert worden: die funktionale und die positionale Ver-

beitung. Die Reihenfolge der Wörter hängt in vielen Sprachen von der syntaktischen Funktion der einzelnen Elemente ab (Subjekt, direktes Objekt, indirektes Objekt, ...). Der Schritt der Zuweisung syntaktischer Funktionen wird funktionale Verarbeitung (*functional processing*) genannt. Für das Deutsche gilt zum Beispiel, dass der Handelnde (Agens) in einem deklarativen Aktivsatz die syntaktische Funktion ›Satzsubjekt‹ erhält und der Handlungsempfänger (Patiens) die syntaktische Funktion ›Satzobjekt‹. Im Passiv hingegen wird der Patiens mit der syntaktischen Funktion Subjekt verknüpft (vgl. Bock/Levelt 1994).

Positionale Verarbeitung: Um die Reihenfolge der Wörter in einem Satz festzulegen, genügt es aber in vielen Sprachen nicht, lediglich syntaktische Funktionen zuzuweisen. Es muss zusätzlich noch berechnet werden in welcher Position ein gegebenes Element auftaucht. Im Deutschen ist es ja durchaus möglich, das direkte Objekt durch Topikalisierung an den Satzanfang zu stellen, die Position also, in der man (in den meisten Fällen) das Satzsubjekt findet (*Oskar hat seine Hausaufgaben gemacht* vs. *Seine Hausaufgaben hat Oskar gemacht*). Die Ausarbeitung der linearen Anordnung von Phrasen wird als positionale Verarbeitung (*positional processing*) bezeichnet (vgl. Bock/Levelt 1994).

Wie genau lexikalischer Zugriff, funktionale und positionale Verarbeitung im Zusammenhang funktionieren wird weiter unten diskutiert. Entscheidend ist an dieser Stelle, dass am Ende des Prozesses der syntaktischen Enkodierung eine Repräsentation vorliegt, die **Oberflächenstruktur** (*surface structure*) genannt wird. In ihr sind die Lemmata sowie die Funktionsmorpheme in der Reihenfolge spezifiziert, in der sie im zu artikulierenden Satz bzw. Äußerungssegment erscheinen sollen. Die Oberflächenstruktur wird dann in den folgenden Prozessen phonologisch enkodiert, das heißt, so spezifiziert, dass am Ende eine Repräsentation vorliegt, die die Aktivierung der Artikulationsorgane auslösen kann.

Bei der **phonologischen Enkodierung** werden **phonologische Codes** aus dem mentalen Lexikon abgerufen, und zwar genau diejenigen, die benötigt werden, um die als Resultat der syntaktischen Enkodierung vorbereitete Oberflächenstruktur lautlich zu spezifizieren. Die phonologischen Codes sind direkt mit den in der Oberflächenstruktur enthaltenen Lemmata und funktionalen Elementen (Funktionsmorpheme) verknüpft. Des Weiteren werden während der phonologischen Enkodierung die abgerufenen lautlichen Informationen zu einer Sequenz von Silben aufgebaut, in der Metrik und Intonation bestimmt sind. Das Resultat des Gesamtprozesses lässt sich als **phonologischer Plan** bezeichnen.

Den Input der **phonetischen Enkodierung** bildet der phonologische Plan. Dieser wird nun im finalen Schritt vor der Artikulation in eine Repräsentation übersetzt, die man sich als Programm für die motorische Steuerung der an der Artikulation beteiligten Organe vorstellen kann. Für die Produktion jeder artikulierbaren Silbe einer Sprache wird genau eine artikulatorische Geste angenommen. Die artikulatorischen Gesten für die häufigsten Silben in einer Sprache können direkt aus einem speziellen Speicher abgerufen werden (mentales Silbarium). Seltene artikulatorische Gesten werden von einem speziellen Mechanismus ›on the fly‹ konstruiert.

Zuweisung syntaktischer Funktionen

Resultat: Struktur ohne Lautform

Vorbereitung der Artikulation

4.3.2 | Evidenz für die Teilsysteme der Sprachproduktion

Versprecher

In den vorherigen Abschnitten wurden die einzelnen Prozessebenen, die dort ablaufenden Prozesse und die einzelnen Informationen vorgestellt, die es zusammengenommen ermöglichen, Sprache zu produzieren. Doch woher weiß man, dass genau diese Prozessebenen angenommen werden müssen, um ein kohärentes Modell der Sprachproduktion zu konstruieren?

Einen Meilenstein auf dem Weg, die Systematik der Äußerungsproduktion aufzuklären, stellt Garretts (1980) materialreiche und mikroskopische Analyse von Versprechern dar. Ihm ist es gelungen, ein überzeugendes Modell vom Aufbau des Produktionssystems und vom Ineinandergreifen der Prozesse vorzuschlagen. Daran knüpften bis heute die meisten Untersuchungen der Sprachproduktionsforschung an, denn unabhängig davon welche größtenteils experimentell gewonnenen Daten in einem Gesamt- oder Teilmodell der Sprachproduktion erklärt werden sollen, jedes Modell muss immer auch in der Lage sein, die verschiedenen Typen von Versprechern zu erklären (vgl. Dell 1986; Bock/Levelt 1994; Levelt et al. 1999; Vigliocco/Hartsuiker 2002).

Versprecher geben
Aufschluss
über Abläufe im
Sprachproduktions-
system

Charakteristik von Versprechern: Versprecher passieren nicht sonderlich häufig. Da aber die Diskrepanz zwischen beabsichtigtem Ausdruck und dem – ungewollt – falsch produzierten Ausdruck jeweils eine Prozedur isoliert erkennbar macht, eben die misslungene, sind Versprecher für die Analyse der Sprachproduktionsvorgänge sehr aufschlussreich. Infolge dessen sind umfangreiche Sammlungen von authentischen Versprechern angelegt worden. Die großen Sammlungen enthalten zwischen 4000 und 8000 Versprechern, die meisten mit Kontext plus/minus Erläuterungen. Das hört sich mächtig an, stellt aber, wie (Berg 1988, S. 15) anmerkt, das Minimum dessen dar, was für eine systematische Untersuchung von Versprechern aller Typen erforderlich ist. Für das Deutsche am besten aufbereitet und zugänglich ist die Datenbank von Wiedenmann (1998). Die umfanglichste Übersicht über Versprechersammlungen ist Wiedenmann (1992), die lustigste Sammlung die von Leuninger (1993). Garrett verwendet das M.I.T.-Corpus mit damals ca. 5000 Versprechern und zur Kontrolle das UCLASEC (University of California at Los Angeles Speech Error Corpus; vgl. Fromkin 1973) mit ca. 3500.

Systematik von Versprechern und Logik der Methode: Um zu Schlussfolgerungen über die Produktionssystematik zu gelangen, hat Garrett (1980) ermittelt, welche Typen von Versprechern überhaupt auftreten. Dazu hat er die Versprecher analysiert, ihre sprachlichen Merkmale beschrieben und sie danach zu Klassen zusammengefasst.

Eine der wohl wichtigsten Erkenntnisse aus dieser Untersuchung war, dass es zwei unterschiedliche Typen von Vertauschungen (*exchange errors*) gibt. Anhand dieser beiden Typen, wollen wir kurz erläutern, welche Logik sich hinter Garretts Herangehensweise verbirgt.

- Bei **Lautvertauschungen** erscheinen Laute, die zu zwei verschiedenen Wörtern gehören, in der falschen Reihenfolge (*die Sorte von Tacher*, siehe unten).

- Bei **Wortvertauschungen** sind es ganze Wörter, die in nicht-intendierter Reihenfolge produziert werden, wie in *Die haut jedem die Nase vor der Tür zu*.

Interessanterweise treten Lautvertauschungen meistens innerhalb einer Phrase auf. Die betroffenen Lexeme stehen nah beieinander und gehören in der Regel nicht zur selben Wortklasse und sind miteinander auch semantisch nicht verwandt. Ferner sind sich die vertauschten Laute lautlich meist ähnlich. Bei Wortvertauschungen ist dies anders. Die vertauschten Elemente stehen meistens relativ weit auseinander, meistens gehören sie zu verschiedenen Phrasen. Phonologische Ähnlichkeit spielt keine Rolle; dagegen gehören die vertauschten Wörter meist derselben Wortklasse an.

Dass diese beiden Typen von Versprechern so unterschiedliche Merkmale aufweisen, so Garretts Argumentation, lässt annehmen, dass die jeweils vertauschten Elemente simultan verarbeitet wurden. Das legt nun wiederum nahe, dass Wort- und Lautvertauschungen in verschiedenen Schritten der Produktion zustande kommen. Bei Wortvertauschungen des hier vorgestellten Typs handelt es sich um Fehlleistungen beim Zuweisen syntaktischer Funktionen: *Nase* (statt *Tür*) erhält irrtümlich die Funktion des direkten Objekts und *Tür* – ebenfalls irrtümlich – die Funktion des Präpositionalobjekts. Bei Lautvertauschungen hingegen handelt es sich um eine Fehlleistung bei der Anordnung lautlicher Segmente. In Garretts Modell, das unten genauer vorgestellt wird, weisen Wortvertauschungen auf Prozesse der funktionalen Ebene und Lautvertauschungen auf solche der positionalen Ebene hin.

Typen von Versprechern

Zur Illustration

Garrett entdeckt in seinen Daten die folgenden Typen von Versprechern. (Beispiele zum besseren Verständnis, wo nicht anders vermerkt, aus Wiedenmann (1998))

Wortsubstitution semantisch ähnlicher Wörter

Beispiel: *Und die Alte ist die Tochter von der Jungen* (Ziel: ... *die Mutter von ...*)

Eigenschaften: Die Wortersetzung betrifft ein Wort (immer ein Inhaltswort). Anstelle des beabsichtigten Wortes wird ein anderes ausgewählt und schließlich auch ausgesprochen. Das Versprecherwort ist semantisch mit dem Zielwort verbunden und diese Verbindung scheint den Versprecher zu begünstigen.

Wortsubstitution phonologisch ähnlicher Wörter (Malapropismus)

Die Bezeichnung für diese Versprecher geht auf Fay/Cutler (1977) zurück. Mrs. Malaprop ist eine Figur in Sheridans ›The Rivals‹ von 1775, deren Vorliebe für schwierige Fremdwörter, komische Folgen zeitigt, weil sie sie nicht beherrscht und in Folge dessen oft verballhornt.

Beispiel: *Lass uns morgen mal fotografieren*. (Ziel: ... *morgen (früh) telefonieren*; unveröffentlichte Sammlung von JG)

Eigenschaften: Betrifft Inhaltswörter. Der Versprecher ist ein Wort des Lexikons; drückt jedoch inhaltlich nicht die geplante Bedeutung aus. Versprecher und beabsichtigtes Wort sind bedeutungsfremd aber lautlich ähnlich und wortartgleich, wodurch der Versprecher begünstigt wird.

Schlussfolgerung: Garrett sah semantisch und phonologisch relationierte Wortersetzungsfehler als Hinweis darauf, dass die Selektion eines Wortes in zwei verschiedenen, zeitlich aufeinanderfolgenden Schritten bzw. Phasen geschieht. Im ersten Schritt wird der Teil eines Wortes ausgewählt, der sich auf Bedeutung bezieht. In einem zweiten Schritt wird auf die Information zugegriffen, die die lautliche Seite betrifft (Lemma-Lexem-Gliederung).

Wortvertauschung

Beispiel: *Die haut jedem die Nase vor der Tür zu* (Ziel: ... *die Tür vor der Nase* ...)

Eigenschaften: Betroffen sind zwei ganze Wörter, fast ausschließlich Inhaltswörter. Zwischen den vertauschten Elementen stehen korrekt gereimte Wörter, hier *vor der*. Die Intonationsstruktur der Äußerung ist korrekt. Die vertauschten Wörter stimmen in der Wortart überein (hier zwei Nomina) und haben syntaktische Gemeinsamkeiten, hier Kopf einer NP in einer vom Verb abhängigen NP bzw. PP. Die Wörter gehören verschiedenen Phrasen an und können weit auseinander stehen, meist jedoch im selben Einfachsatz.

Schlussfolgerung: Wortvertauschungen entstehen dadurch, dass aus dem Lexikon ausgewählte Wörter (genauer Lemma-Informationen der Wörter) auf falsche Weise in eine durch das Sprachproduktionssystem aufgebaute abstrakte Struktur eingesetzt werden. Garrett erklärt dies durch einen Fehler bei der Verknüpfung von Lemmata mit syntaktischen Funktionen. Mit der Existenz solcher Versprecher begründet Garrett eine Ebene der funktionalen Verarbeitung im Sprachproduktionsprozess.

Morphemvertauschung (Stranding)

Beispiel: *Am Freiheitspark platzen* (Ziel: ... *platz parken*)

Eigenschaften: Betroffen sind zwei Morpheme, meist Stammmorpheme von Inhaltswörtern. Die betroffenen Wörter können nah beieinander stehen oder auch nicht und können verschiedenen Wortklassen angehören; typischerweise sind sie satzgliedverschieden. Flexionsmorpheme werden zurückgelassen, meistens akkommodiert wie in *beziehen den Rang ihrer Gehaltsklasse* (statt: *das Gehalt ihrer Rangklasse* (ohne Fugen -s)).

Schlussfolgerung: Dass die Komponenten bei *Strandings* weit auseinander stehen können, deutet darauf hin, dass direkte Nachbarschaft nicht erforderlich ist. Dass gleichzeitig die Flexionselemente ihre Position behalten, deutet darauf hin, dass doch für Teile der Kette die Reihenfolge schon fixiert war. Es muss also einen Vorgang geben, der zwischen dem Lemma-Zugriff und dem Aufbau der phonetischen Kette liegt. Garrett nennt diesen »construction of detailed phrasal environments« (S. 212) und nimmt an, dass diesem Prozess sog. »planning frames« zu Grunde liegen. Das sind syntaktisch gegliederte, phrasengroße Rahmen mit Flexionskategorien und -merkmalen, mit Lemma-Information und mit prosodischer Information aber ohne phonetische Formen in den lexikalischen Slots. Das *Stranding* passiert beim Einsetzen von Lexem-Information von Inhaltswörtern in den (falschen) phrasalen Planungsrahmen.

Phonemvertauschung (Spoonerismus)

Beispiel: *Eine Sorte von Tacher.* (Ziel: ... *Torte von Sacher*)

Eigenschaften: Betroffen sind zwei Phoneme. Die Intonationskontur ist korrekt, die beiden Phoneme haben gemeinsame Merkmale und Positionsmerkmale; hier konsonantisch, alveolar, gespannt (Sacher wird in der Mundart des Sprechers gespannt gesprochen); beide bilden den linken Silbenrand (Onset). Die vertauschten Elemente stehen meistens nah beieinander, die betroffenen Wörter sind syntaktisch meist unähnlich.

Schlussfolgerung: Der Versprecher passiert beim (falschen) Einsetzen lautlicher Segmente in für ihre Positionierung vorgesehene Planungsrahmen. Phonemvertauschungen gelten Garret als ein Argument für die Existenz einer Ebene der positionalen Verarbeitung im Sprachproduktionsprozess.

Lautreihenfolgeversprecher (Antizipation und Perseveration)

Meringer/Mayer (1895) nennen die beiden Versprechertypen Vorklang und Nachklang, was ihre Eigenschaften augenfälliger ausdrückt. Da Antizipationen die weitaus häufigeren sind, behandeln wir hier nur diese Klasse.

Beispiel: *Ich komme morgen Vormittag* (Ziel: ... *morgen Vormittag* ...)

Eigenschaften: Betroffen ist ein Einzellaut. Er wird ersetzt durch einen Laut, der später in der Lautkette folgt. Versprecher und antizipierter Laut nehmen dieselbe Silbenposition ein, meist den Silben-Onset, bei Vokalen die Silbenkernposition. Sie können weit auseinander stehen oder nah beieinander. Die betroffenen Wörter können syntaktisch ähnlich sein oder auch nicht. Häufig sind Funktionswörter und Flexionspräfixe betroffen, in Garretts Daten mehr als 10 Prozent.

Schlussfolgerung: Die Prozesse betreffen die lautliche Form, müssen also am Aufbau von Ketten oder Teilketten beteiligt sein. Da die betroffenen Teile, anders als bei Phonemvertauschungen, weit auseinander liegen können und Flexionsmorpheme betroffen sind, kann es sich nur um den Vorgang handeln, bei dem der Inhalt der Planungsrahmen nach Auffüllung mit aller Lexem-Information in die phonetische Endkette überführt wird.

Wortkontamination

Beispiel: *Das ist gul* (Ziel: *Das ist gut/cool*; unveröffentlichte Sammlung von JG)

Eigenschaften: Der Versprecher besteht aus der Verschmelzung zweier bedeutungsverwandter Wörter, die im Prinzip zwei Alternativen der auszudrückenden Bedeutung darstellen.

Schlussfolgerung: Garrett argumentiert, dass derartige Versprecher deutlich für ›Parallelismus‹ sprechen (Garrett, 1980, S. 211): Wörter, die gleichermaßen gut geeignet sind, den Inhalt der Message auszudrücken, werden selektiert und durchlaufen alle Stufen des gesamten Verarbeitungsprozesses. Die Fehlleistung des Sprachproduktionssystems besteht bei Wortkontaminationen in der rechtzeitigen Unterdrückung eines der Kandidaten. Das Problem entsteht früh, aber der Fehler entsteht spät.

Sprachproduktionsmodell basierend auf der Analyse von Versprechern:

Garret (1980) entwickelt, basierend auf der aufgedeckten Systematik der analysierten Versprecher, einige Hypothesen, die zusammengenommen ein sehr detailliertes Sprachproduktionsmodell ergeben (s. Abb. 4.2). Betrachten wir zunächst die Hypothesen.

1. Sprachproduktion ist ein kognitiver Prozess der regelgeleiteten Symbolverarbeitung, wobei Äußerungsinformation auf mehreren unterschiedlichen Ebenen in unterschiedlichen Symbolsprachen repräsentiert ist.

- Auf der Message-Ebene liegt die konzeptuelle Struktur in einer mentalen Symbolsprache vor.
- Auf der funktionalen Ebene (= Ebene der zu Grunde liegenden syntaktischen Struktur) liegt Äußerungsinformation in einer Variante einer formalen Algebra vor.
- Auf der Ebene der Planungsrahmen liegt Äußerungsinformation in der Form von linearisierten und syntaktisch gegliederten Slot-Schemata (engl. templates) vor.
- Auf der Ebene der finalen lautlichen Kette ist sie als Folge von Bündeln phonetischer und suprasegmentaler Merkmale repräsentiert.

2. Sprachproduktion vollzieht sich als Folge mehrerer lexikalischer Zugriffe und Umwandlungsschritte.

- Lexikalischer Zugriff geschieht zum ersten Mal, wahrscheinlich gesteuert von konzeptuellen Merkmalen, auf semantische Informationen im Lexikon und dabei erfolgt die Aktivierung von Lemmata. Dieser Vorgang ist aus den Versprecherdaten nicht positiv zu erschließen, sondern nur ex negativo. Alle anderen Erklärungen für das Zustandekommen von bedeutungsverwandten Wortsubstitutionen scheiden nämlich aus. Der zweite, eigentlich erste sicher anzunehmende Lexikonzugriff ist die Selektion der Lemmata.
- Der zweite (sichere) Zugriff ist der auf die Lexem-Informationen der Inhaltswörter zur detaillierten »Auffüllung« der phrasengroßen Planungsrahmen.
- Umwandlungsprozeduren betreffen die Erstellung und Bearbeitung der funktionalen Repräsentation, die Ableitung von Planungsrahmen und die Erzeugung der Lautkette der positionalen Repräsentation, also drei, von denen man die ersten beiden als syntaktische Enkodierung, den letzten als phonologische Enkodierung bezeichnen würde. Garrett rechnet allerdings auch die Erstellung der Planungsrahmen zur phonologischen Enkodierung, weil eben phonologische Information verarbeitet und pro Rahmen in eine lineare Ordnung gebracht wird.

3. Das System arbeitet seriell und diskret; zu Einzelheiten des Zeitablaufs lässt sich anhand der Versprecherdaten allerdings nichts sagen.

Garretts Gesamtmodell der Sprachproduktion: Aus den einzelnen Schlussfolgerungen, setzt man sie zu einem Bild zusammen, ergibt sich die in Abb. 4.2 abgebildete Vorstellung von der Systematik der Äußerungsproduktion:

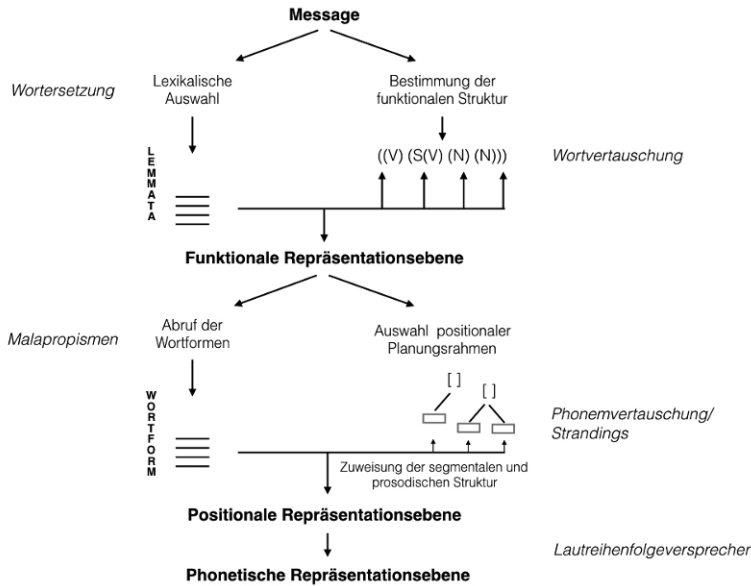


Abb. 4.2: Sprachproduktionsmodell von Garret (1980), Darstellung in Anlehnung an Vigliocco/Hartsuiker (2002)

Garretts Arbeit ist natürlich nicht die einzige Versprecheranalyse (für eine Übersicht vgl. Fromkin 1984; Marx 2000), und Versprecheranalysen sind nicht der einzige Weg zur Aufklärung von Sprachproduktionsvorgängen. Tatsächlich hat die Zahl experimenteller Untersuchungen seit Ende der 1980er Jahre stark zugenommen, auch ihr Beitrag zum derzeitigen Kenntnisstand, mit dem wir uns in den folgenden Abschnitten befassen.

Experimentelle Evidenz für Teilsysteme

Es liegen unzählige Experimente vor, die die durch die Analyse von Versprechern motivierte Differenzierung der Teilsysteme des Sprachproduktionssystems belegen. Eine der wohl meistbeachteten Studien kommt von Schriefers et al. (1990). Die Ergebnisse dieses Experiments verdeutlichen sehr klar, dass zumindest die Verarbeitung von Bedeutungs- und lautlichen Informationen von unterschiedlichen Systemen – und was unten noch mehr relevant wird – zu unterschiedlichen Zeitpunkten geschieht.

Evidenz aus dem Bild-Wort-Interferenz-Paradigma (Schriefers et al. 1990)

Experiment

Versuchspersonen haben die Aufgabe, Objekte, die auf Bildern dargestellt sind, zu benennen. Allerdings sind die Darstellungen der Objekte nicht der einzige Stimulus, der den Versuchspersonen präsentiert wird. Zusätzlich zur Darstellung eines Objekts hören Versuchspersonen auditiv präsentierte Distraktorwörter, die sie ausdrücklich während der Objektbenennung ignorieren sollen. Pro Versuchsdurchgang wurde genau ein zu benennendes Objekt gezeigt und genau ein Distraktor auditiv präsentiert. Die Distraktoren stammen aus unterschiedlichen Kategorien: Es wurden u. a. Distraktoren verwendet, die mit dem Zielwort (Name des zu benen-

nenden Objekts) semantisch relatiert sind (*Hund-Katze*), solche, die mit dem Zielwort phonologisch relatiert sind (*Hund-Hut*) und solche, die mit dem Zielwort weder semantisch noch phonologisch in einer Beziehung standen (*Hund-Tisch*).

Eine weitere Manipulation in diesem Experiment betraf den Zeitpunkt der Präsentation eines Distraktors relativ zum Zeitpunkt des Präsentationsbeginns des zu benennenden Objektes. Man spricht hier von *stimulus onset synchrony* (SOA). Alle Arten von Distraktoren wurden im Experiment entweder kurz vor (-150ms), gleichzeitig (0ms) oder kurz nach dem (+ 150ms) Objekt präsentiert, das die Versuchsteilnehmer benennen sollten.

Gemessen wurde die durchschnittliche Zeit, die zwischen der Präsentation des zu benennenden Objektes und dem Artikulationsbeginn vergeht, die Sprechbeginnzeit. Die Ergebnisse bestanden nun darin, dass semantisch relatierte Distraktoren die Benennzeiten im Vergleich zu nicht relatierten Distraktoren verlangsamten, allerdings nur dann, wenn sie früh präsentiert wurden. Phonologisch relatierte Distraktoren hingegen beschleunigten die Sprechbeginnzeiten im Vergleich zu neutralen, allerdings nicht, wenn sie früh präsentiert wurden. Dass beide Distraktorentypen einen unterschiedlichen Effekt hatten, war zu erwarten. Dies war schon aus vorherigen Studien bekannt. Der entscheidende Punkt ist, die zeitliche Versetzhtheit des Effekts: Der semantische Distraktoreffekt tritt nur bei früher Präsentation auf, der phonologische nur bei später. Zusammengefasst zeigen die Befunde, dass semantische Informationen vor phonologischen verarbeitet werden.

Evidenz durch den Tip-of-the-tongue-(TOT)-Zustand

Die meisten von uns haben schon einmal erlebt, dass man genau weiß, was man sagen will, einem das richtige Wort aber partout nicht einfallen will, obwohl man sicher weiß, dass man das Wort kennt und welche syntaktischen Eigenschaften es hat; nur der automatische Zugriff auf die Lautform ist gehemmt. Das Wort liegt einem, wie man sagt, »auf der Zunge«. Im Englischen wird dieser Zustand als *tip-of-the-tongue*-Zustand (TOT-Zustand) bezeichnet. Psycholinguistisch wird das Phänomen damit erklärt, dass zwar ein Lemma aktiviert werden kann, die Aktivierung der korrespondierenden Lexem-Information aber gehemmt ist (Butterworth 1989; 1992; Levelt 1989; Garrett 1984). Andere Autoren sprechen hier von partieller Lemma-Lexem-Aktivierung (Dell 1986; Meyer/Bock 1992). Die einschlägigen Arbeiten zeigen, dass man, wenn man einen TOT-Zustand erfährt, zwar nicht auf die vollständige Wortform zugreifen kann, bestimmte phonologische Informationen, wie die Anzahl der Silben oder der phonologische Onset des Wortes, kurz wortformrelatierte Informationen, dennoch überzufällig häufig angeben kann (Brown 1991; Brown/Mc Neill 1966). In anderen Arbeiten mit experimentell induziertem TOT-Zustand wurde gezeigt, dass zudem häufig auch syntaktische Informationen des Wortes, das man nicht hervorbringen kann, berichtet werden können, wie zum Beispiel das Genus (Vigliocco/Antonini/Garrett 1997). Zusammen-

Weitere Evidenz
für die Plausibilität
der Einteilung
des Gesamt-
prozesses in
Teilprozesse

genommen deuten die Befunde darauf hin, dass während der Wortselektion Informationen unterschiedlicher Qualität nicht gleichzeitig aktiviert werden sondern lautliche Informationen erst nach Bedeutungs- bzw. Strukturinformationen (vgl. Vigliocco/Hartsuiker 2002).

Evidenz aus der Aphasieforschung

Auch aus Befunden der Sprachstörungsforschung lässt sich die Annahme separater Teilsysteme im Sprachproduktionsprozess bestätigen. Die Performanz von Sprechern mit Anomie – einer Wortfindungsstörung – ist hier typischerweise aufschlussreich. Ähnlich wie Sprecher, die sich in einem TOT-Zustand befinden, scheinen Anomiepatienten Lemma-Informationen aktivieren zu können, wohingegen der Abruf von Lexem-Informationen gestört ist (Badecker/Miozzo/Zanuttini 1995; Goodglass/Kaplan/Weintraub/Ackerman 1976). Aufschlussreich ist in diesem Zusammenhang auch, dass die Krankheit bei diesen Patienten sich in unterschiedlichen Ausprägungen zeigen kann. Patient J. C. U. (Howard/Orchard-Lisle 1984) zum Beispiel half es, das intendierte Wort auszusprechen, wenn ihm phonologische Hinweise (Cues) auf die korrekte Form präsentiert wurden. Phonologische Cues, die auf einen semantisch relatierten Nachbarn verwiesen, evozierten hingegen Fehlleistungen. War das intendierte Wort z. B. »Baseball«, führte der Cue »beginnt mit einem R« zur (nicht-intendierten) Produktion des falschen Wortes »Racket«. Patient E. S. T. auf der anderen Seite (Kay/Ellis 1987) wurde durch »falsche« Hinweise nicht irritiert. J. C. U.s Performanz war auch in anderen semantischen Aufgaben (Verstehen und Produzieren von Sprache) eher schwach, während E. S. T. in dieser Hinsicht nicht auffällig war. Die Befunde von J. C. U. legen nahe, dass hier eine Störung vorliegt, die an der Schnittstelle von Konzeptualisierung und Lemma-Abruf zu verorten ist. Bei E. S. T. scheint die Störung eher die Schnittstelle zwischen Lemma- und Lexem-Information zu betreffen (vgl. Vigliocco/Hartsuiker 2002).

4.3.3 | Die Dynamik des Zusammenspiels der einzelnen Prozessebenen

In der bisherigen Darstellung ging es in erster Linie um die Identifikation der am Sprechen beteiligten Systeme bzw. Komponenten und um die innerhalb der Komponenten ablaufenden Einzelprozesse sowie die Informationen, die von diesen Prozessen verarbeitet werden. Nur wenig ist bisher dagegen zur Systematik des Zusammenspiels der erwähnten Komponenten gesagt worden. Davon wird nun im Folgenden die Rede sein.

Informationsverkapselung und Informationsfluss

Diskrete Arbeitsweise: Sowohl in Garretts (1980) als auch in Levels (1989) Sprachproduktionsmodell wurden die einzelnen Prozessebenen als spezialisierte **Module** aufgefasst, deren jedes einen spezifischen Input zulässt und einen spezifischen Output generiert. Der Input des Moduls

Diskret oder
kaskadierend?

›Syntaktische Enkodierung« akzeptiert als Input eine Message und erzeugt als Output die (lautlich noch nicht spezifizierte) Oberflächenstruktur. Das Modul ›Phonologische Enkodierung‹ erhält als Input diese Oberflächenstruktur und generiert als Output eine phonologische Repräsentation. Eine Kernbehauptung dieses Modells besagt, dass die Module **diskret** arbeiten: Informationen innerhalb des Moduls (M) werden erst vollständig verarbeitet, bevor das Ergebnis an das nächste Modul (M + 1) weitergegeben wird. Modul M + 1 hat keinen Zugriff auf die Informationen in Modul M. Sobald die Verarbeitung in Modul M abgeschlossen ist und das Ergebnis an M + 1 weitergegeben wurde, so die Behauptung, seien die Informationen dem Modul M nicht mehr zugänglich. Diese Besonderheit des modularen Modells wird als ›Informationsverkapselung‹ (*information encapsulation*) des Moduls bezeichnet. Für den zeitlichen Ablauf bedeutet dies zum Beispiel, dass zunächst genau ein Lemma ausgewählt wird, bevor die mit ihm lexikalisch zugeordneten Lexem-Informationen aktiviert werden. Dies bedeutet nicht, dass nicht anfangs mehrere Lemmata aktiviert sein können, also miteinander im Wettbewerb stehen, sondern, dass, bevor es an die phonologische Enkodierung geht, genau ein Lemma ausgewählt wird (s. Abb. 4.3).

Als Evidenz für diese diskret-serielle Arbeitsweise des Formulators gelten u. a. die Beobachtungen aus dem oben dargestellten Experiment von Schriefers et al. (1990). Insbesondere die Tatsache, dass bei der Objektbenennung die Sprechbeginnzeit von der Präsentation semantischer Distraktoren nur dann beeinflusst wird, wenn diese kurz vor dem zu benennenden Objekt (Stimulus) präsentiert werden. Dass dieser Effekt bei späterer Präsentation semantischer Distraktoren ausbleibt, so die Autoren, lege nahe, dass die semantische Verarbeitung des Zielwortes dann bereits abgeschlossen ist.

Kaskadierende Arbeitsweise: Als Gegenentwurf zu dieser Sichtweise gehen andere Psycholinguisten (vgl. Dell 1986; Peterson/Savoy 1998) davon aus, dass es zwar einzelne spezialisierte Subsysteme gibt, der Informationsfluss zwischen ihnen aber nicht diskret, sondern **kaskadierend** sei. Stehen beispielsweise während des Prozesses der Wortselektion mehrere Lemmata als potenzielle Kandidaten zur Verfügung (z. B. ein Extremfall; *Couch/Sofa*), aktiviert jedes dieser Lemmata noch bevor es endgültig ausgewählt ist bis zu einem gewissen Grad sofort und automatisch seine korrespondierenden Lexem-Informationen.

Informationsfluss

Unidirektional oder bidirektional? Eine weitere Frage ist, wie die Informationen vom Konzeptualisierer zum Artikulator gelangen, ob der Informationsfluss nur in eine Richtung geht (von der obersten in Richtung unterste Ebene) oder, ob es Feedback von den unteren Ebenen gibt, der Informationsfluss also auch in beide Richtungen laufen kann. Sowohl Garrett (1980) als auch Levelt (1989) und Levelt und Kollegen (1999) postulieren neben diskreter Arbeitsweise unidirektionalen Informationsfluss. In anderen Arbeiten (vgl. u. a. Dell 1986; Dell/O'Seaghdha 1992; Harley 1993) hingegen ist im Sprachproduktionsmodell Feedback von einer unteren zu einer höheren Ebene vorgesehen, der Informationsfluss wird also als bidirektional angesehen. Modelle dieser Art werden oft **interaktive Modelle** genannt (s. Abb. 4.3).

Nun ist es aber nicht so, dass jedes kaskadierende Modell auch zwangsläufig bidirektionalen Informationsfluss annehmen muss. Es gibt auch die Klasse der sogenannten kaskadierenden *feed-forward*-Modelle (vgl. Humphreys/Riddoch/Quinlan 1988).

Als eines der Hauptargumente der Vertreter interaktiver Modelle gilt der sogenannte **lexical bias-Effekt** (vgl. Baars/Motley/MacKay 1975). Dieser besteht darin, dass Lautvertauschungen überdurchschnittlich häufiger zu tatsächlich existierenden Wörtern führen als das statistisch zu erwarten wäre. Gut bestätigt ist zum Beispiel, dass das Wortpaar *darn-bore* (*verdammst-Langeweile*) ungefähr dreimal so häufig zu dem Versprecher *barn-door* (*Scheune-Tür*) führt als das Wortpaar *deal-back* (*verteilen-zurück*) zu *beal-dack* (Nonwort-Nonwort). Erklärt wird dies unter anderem von Dell (1986; 1988) damit, dass aktivierte phonologische Knoten zurück nach oben, also zur Lemma-Ebene feuern, wodurch unter Umständen andere als die intendierten Lemmata aktiviert werden und ihrerseits Aktivierung wieder nach unten zu den phonologischen Knoten abgeben. Da auf der Lemma-Ebene aber natürlich nur echte Wörter gespeichert sind, kann dieser Mechanismus prinzipiell den *lexical bias*-Effekt erklären. Abgesehen von diesem Effekt könnte ein Feedbackmechanismus auch sogenannte **mixed-errors** erklären. Das sind Versprecher, bei denen nicht eindeutig zu klären ist, ob sie bei der Lemma-Selektion bzw. -aktivierung oder bei der Aktivierung der lautlichen Information entstanden sind, z. B. *aufkommen* statt dem intendierten *auftreten* bzw. *vorkommen*. Durch Feedback, also Bottom-up-Aktivierung von der phonologischen Ebene, könnte eine bereits aktivierte Alternative (auf der Lemma-Ebene) mehr Aktivierung erhalten als eigentlich intendiert.

Zu beachten ist, dass beide theoretischen Überlegungen (diskret/kaskadierend und unidirektional/bidirektional) sowohl Einfluss darauf haben, wie der lexikalische Zugriff modelliert wird – also wie genau Lemma- und Lexem-Informationen aus dem mentalen Lexikon abgerufen werden – als auch darauf, wie man sich theoretisch den Aufbau von komplexen sprachlichen Strukturen vorstellt, also wie man von einer Message-Struktur zu einer syntaktischen Struktur gelangt und wie von dort zu einer phonologischen Repräsentation. Die theoretischen Annahmen betreffen demnach **Auswahl- und Konstruktionsprozesse** gleichermaßen (vgl. Vigliocco/Hartsuiker 2002).

Experimentelle Untersuchungen zum Informationsfluss während der Wortselektion

Die Frage, ob der lexikalische Zugriff als seriell-diskreter, kaskadierender *feed-forward* oder interaktiver Prozess zu beschreiben sei, wurde in der Forschung seit spätestens dem Beginn der 1990er Jahre kontrovers diskutiert (vgl. Jescheniak/Hahne/Schriefers 2003). Vertreter aus beiden Lagern präsentierten empirische Evidenz, die für oder gegen eine der beiden Positionen interpretiert wurde.

Vorhersagen der verschiedenen Konzeptionen: Um die im Folgenden zur Illustration präsentierten Studien besser nachvollziehen zu können, ist es hilfreich, einige Vorhersagen anzuführen, die aus den verschiedenen zur Disposition stehenden Modellen abzuleiten sind.

Argumente für
Interaktionismus

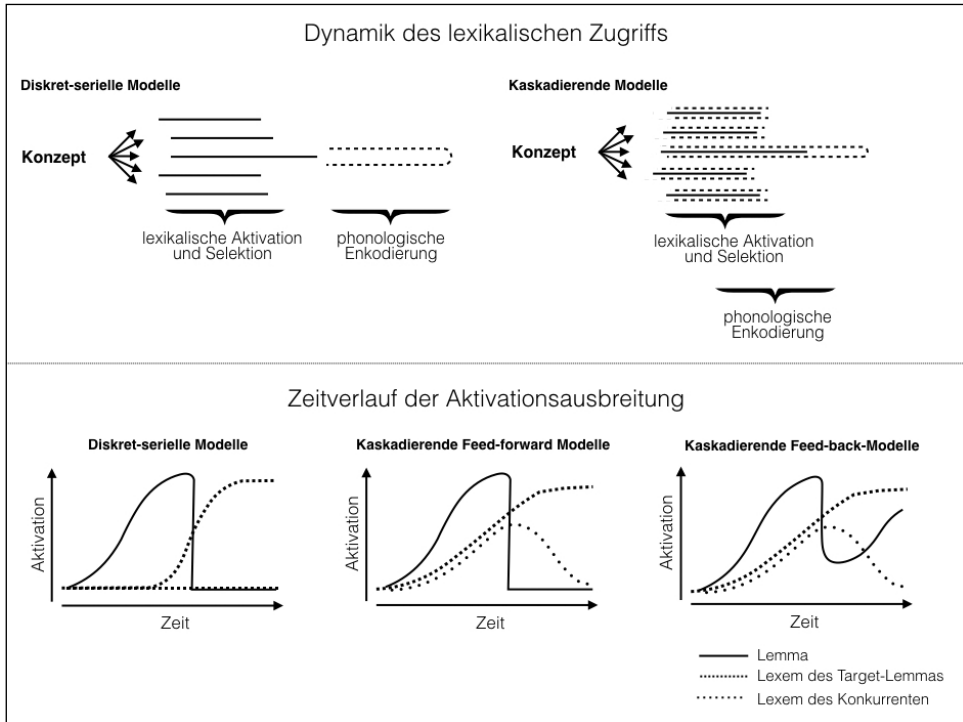


Abb. 4.3: Darstellung des zeitlichen Ablaufs des zweifachen lexikalischen Zugriffs und Aktivationskurven nach verschiedenen Theorien (vgl. Levelt et al. 1991)

Eine Vorhersage, die aus der diskret-seriellen Konzeption folgt, ist, dass semantische und phonologische Informationen zu keinem Zeitpunkt gleichzeitig aktiviert sein sollten und dass semantische Informationen stets vor phonologischen verarbeitet werden, denn phonologische Informationen werden erst dann aktiviert, wenn ein Lemma nicht nur aktiviert, sondern ausgewählt wurde. Modelle, die kaskadierenden Informationsfluss annehmen, sagen hingegen vorher, dass semantische und phonologische Informationen durchaus gleichzeitig Aktivierung erhalten, da die Arbeitsweise während der Enkodierung eben nicht diskret ist. Die Vorhersage folgt sowohl aus der Annahme der kaskadierenden Informationsausbreitung als auch aus der Annahme, dass es im System Feedback gibt.

Eine weitere Vorhersage der diskret-seriellen Konzeption lautet, dass die phonologischen Informationen ebenfalls aktivierter Lemmata keinerlei Priming-Effekt (Voraktivierung) auf eine Target-Lemma auslösen sollten, denn während der phonologischen Verarbeitung wird eben nur phonologische Information des vorher ausgewählten Lemmas prozessiert, aber nicht die vorher ebenfalls aktivierter aber nicht ausgewählter, konkurrierender Lemmata. Interaktive Modelle sagen genau einen solchen Priming-Effekt vorher, nämlich aufgrund der Annahme, dass Information auch von unten (phonologische Ebene) nach oben (Lemma-Ebene) fließen kann (s. Abb. 4.3).

Evidenz für seriell-diskreten Zugriff: Van Turennout/Hagoort/Brown (1997) berichten ein Experiment, in dem die Hirnaktivität auf ganz spezi-

fische Weise beim Produzieren einer Nominalphrasen als Reaktion in einer Objektbenennungsaufgabe gemessen wurde. Um einen Einblick in den Zeitablauf der Kodierung von Lemma- und Lexem-Informationen zu bekommen, sollten die Versuchspersonen in diesem Experiment zusätzlich zur Objektbenennungsaufgabe in der Hälfte aller Versuchsdurchgänge eine zweiteilige Kategorisierungsaufgabe lösen, wobei Versuchsdurchgänge mit und ohne zusätzliche Aufgabe zufällig alternierten. Der eine Teil der Kategorisierungsaufgabe betraf die Entscheidung, ob das gezeigte Objekt belebt oder unbelebt ist, also semantische Information. Der zweite Teil betraf die Entscheidung einer phonologischen Eigenschaft des zu produzierenden Nomens, mit dem das Objekt bezeichnet wurde (z. B. endet auf /s/ oder /n/ bzw. beginnt mit einem /k/ oder /s/). Die Reaktion der Versuchspersonen wurde durch beide Kategorisierungsaufgaben bestimmt. In Experiment 1 beispielsweise bestimmte die semantische Information (belebt/unbelebt), mit welcher Hand (rechts/links) eine Taste gedrückt werden sollte, wohingegen die phonologische Information bestimmte, ob der Tastendruck tatsächlich ausgeführt werden sollte oder nicht (**go/-no-go-Entscheidung**). In Experiment 2 wurde dies umgedreht. Hier bestimmte die phonologische Information, mit welcher Hand die Reaktion erfolgen sollte, und die semantische Information, ob der Tastendruck erfolgen soll oder nicht. In dieser Untersuchung wurde also eine sprachliche Aufgabe mit einer motorischen Aufgabe auf ganz spezifische Weise gekoppelt. Unabhängig davon, ob die Kategorisierungsaufgabe in einem Versuchsdurchgang auftrat oder nicht, sollten Versuchspersonen das gezeigte Objekt immer benennen. Von Interesse war nun unter welchen Bedingungen ein **laterales Bereitschaftspotential** (LRP) in den Go-/No-Go-Versuchsdurchgängen gemessen werden kann. Ein laterales Bereitschaftspotential ist ein neurophysiologisches Phänomen, das kurz vor einer willkürlichen Bewegung im Gehirn messbar ist (LRP, siehe Kasten zur Erklärung).

go/-no-go-
Entscheidung:
Ausführung des
Tastendrucks
hängt von
Kategorisierungsaufgabe ab

In Experiment 1 wurde ein LRP in den Go- und No-go-Trials festgestellt, was dafür spricht, dass die **semantische** Information des (für die Benennung notwendigerweise aktivierten) Wortes die Hand aktivierte, egal ob die phonologische Kategorisierungsaufgabe zur Ausführung der Entscheidung führte oder nicht. In Experiment 2 hingegen wurde kein LRP in den No-go-Trials festgestellt, sondern nur in den Go-Trials, was dafür spricht, dass die Vorbereitung der Reaktion (Tastendruck) nicht durch die phonologischen Informationen beeinflusst wurde. Die semantischen Informationen standen also schon zur Verfügung noch bevor die phonologischen Informationen zur Aktivierung der Reaktion führte. Tendenziell konnte in dieser Studie auch festgestellt werden, dass die gemessenen LRP in Experiment 1 zeitlich früher auftraten als in Experiment 2.

Zusammengefasst deuten die Befunde demnach darauf, dass semantische Informationen während der Sprachproduktion (automatisch) vor den phonologischen Informationen verarbeitet werden, und dass die Verarbeitung beider Informationstypen nicht überlappt. Die oben dargestellten Befunde aus der Studie von Schriefers et al. (1990) werden durch diese Studie also bestätigt.

Evidenz gegen bidirektionalen Informationsfluss: Levelt et al. (1991) untersuchten den lexikalischen Zugriff mit Hilfe eines Experiments, in dem eine Objektbenennungsaufgabe (Sprachproduktion) mit einer lexikalischen Entscheidungsaufgabe (Sprachrezeption) verbunden wurde. Die Hauptaufgabe der Versuchspersonen bestand darin, auf Bildern dargestellte Objekte zu benennen, jeweils eines pro Versuchsdurchgang. In einigen Versuchsdurchgängen wurde kurz nach Präsentation des zu benennenden Objektes ein lautlicher Reiz über Kopfhörer präsentiert und Versuchspersonen sollten per Tastendruck angeben, ob es sich hierbei um ein Wort oder Non-Wort handelt. In den kritischen Versuchsdurchgängen waren die auditiv präsentierten Testwörter semantisch oder phonologisch mit dem Wort verwandt, welches die Versuchspersonen verwendeten, um die Objektbenennungsaufgabe zu lösen. Als Vergleich fungierten weder semantisch noch phonologisch verwandte Testwörter. Wenn das visuell präsentierte Objekt beispielsweise ein *Schaf* war, dann war das semantisch verwandte auditiv präsentierte Testwort *Ziege* und das phonologisch verwandte *Schal*. Ein nicht-verwandtes Testwort wäre z. B. *Holz*.

Gleichzeitiges
Bewältigen
von zwei Aufgaben

Das Hauptinteresse der Autoren galt der Zeit, die Versuchspersonen für die lexikalische Entscheidungsaufgabe benötigten. Die grundsätzliche Idee hinter diesem experimentellen Design war, dass die Information, die für die Benennungsaufgabe aktiviert wird (Sprachproduktion), einen Einfluss auf die Performanz in der lexikalischen Entscheidungsaufgabe hat. Semantische Information des Target-Wortes (Objektbenennung) sollte sich auf die Reaktionszeit der semantisch verwandten Testwörter auswirken und phonologische Informationen des Target-Wortes auf phonologisch verwandten Testwörter. Manipuliert wurde in diesem Experiment außerdem der Zeitpunkt, zu dem das Testwort relativ zum zu benennenden Objekt präsentiert wurde. Für unsere Zwecke genügt es hier zwischen früh, mittel und spät zu unterscheiden.

In Hinblick auf die Ergebnisse ist hier nun relevant, dass die lexikalische Entscheidung für semantisch verwandte Testwörter nur dann von der Objektbenennungsaufgabe beeinflusst war, wenn diese früh präsentiert wurden. Bei späterer Präsentation, ergab sich kein Effekt. Semantische Information für die Objektbenennungsaufgabe wird demnach nur früh verarbeitet. Wenn die semantische Information des Wortes, das für die Objektbenennung verwendet wird, auch noch spät(-er) aktiviert ist, wie es von den interaktiven Modellen vorhergesagt wird (Feedback), dann hätte sich kein Unterschied zwischen früher, mittlerer und später Testwortpräsentation zeigen sollen. Ein Effekt für phonologisch verwandte Testwörter wurde übrigens bei später Präsentation festgestellt.

In zwei weiteren experimentellen Bedingungen wurde dann auch noch überprüft, ob mit dem Target-Lemma konkurrierende Lemmata jeweils ihre phonologischen Informationen aktivieren oder ob sich die Aktivierung phonologischer Informationen auf das Target-Wort beschränkt. Hierfür wurden nun als phonologisch verwandte Testwörter solche verwendet, die nicht mit dem Target sondern mit den Konkurrenten des Targets phonologisch verwandt sind. Was die Art der Konkurrenten betrifft, wurde zwischen **semantisch assoziierten** (*Schaf-Wolle*) und **kategoriegleichen** unterschieden (*Schaf-Ziege*). Wäre das zu benennende Objekt ein Schaf,

dann wäre im ersten Fall ein entsprechendes Testwort *Wolke*, denn *Wolke* ist mit *Wolle* phonologisch verwandt. Im zweiten Fall wäre *Zwiebel* ein phonologisch verwandtes Testwort, denn ein kategoriegleicher Konkurrent von *Schaf* ist *Ziege* und *Zwiebel* ist phonologisch mit *Ziege* verwandt. Es zeigten sich in Hinblick auf diese experimentellen Bedingungen jedoch keine Effekte, was die Autoren als Evidenz gegen die Annahme werteten, dass Lemmata, die aktiviert, aber nicht selektiert sind, ihrerseits automatisch phonologische Informationen aktivieren. Demzufolge halten die Autoren an einem seriell-diskreten Modell des lexikalischen Zugriffs fest, in dem Lemmata selektiert sein müssen, bevor der Abruf lautlicher Informationen initiiert wird.

Jescheniak et al. (2003) gingen in einer Studie, in der die Hirnaktivität unter verschiedenen Bedingungen gemessen wurde, noch einmal der Frage nach, ob mehrere gleichzeitig aktivierte Lemmata, nämlich solche, die durch semantische Assoziationen miteinander in Beziehung stehen (*category associates*), auch ihre jeweils korrespondierenden Lexem-Informationen aktivieren. Ein Grund dafür waren Bedenken, die an der von Levelt et al. (1991) verwendeten Methode geäußert wurden, der gemäß die Methode möglicherweise nicht sensibel genug sei, um die von den Verfechtern der kaskadierenden (Feedback-)Modelle vorhergesagten Effekte zu zeigen. Doch auch mit Hilfe der von Jescheniak et al. (2003) verwendeten neurophysiologischen Methode konnte kein Hinweis darauf gefunden werden, dass phonologische Information von um Selektion »kämpfenden« Lemmata während der Benennung von Objekten aktiviert wird.

EEG-Messung in der Sprachproduktionsforschung

Kognitive Vorgänge sind experimentell nicht nur anhand ihres zeitlichen Verlaufs zu messen. Wie alle Prozesse der Informationsverarbeitung werden auch sprachliche Aktivitäten neuronal, also in den Zellen des Gehirns, als **elektrochemische Zustandsveränderungen** in den Blutbahnen des Kortex realisiert. Und diese sind über Dioden auf der Kopfhaut messbar. Die Messdaten (in Mikrovoltsschwankungen) können graphisch dokumentiert werden. So entsteht das bekannte **Elektroenzephalogramm (EEG)**. Unterschiedliche kognitive Aktivitäten lösen unterschiedliche Energieschwankungsmuster, sog. Potentiale aus, die Bearbeitung lautlicher Phänomene andere als die Verarbeitung syntaktischer oder (wieder andere) semantischer Probleme.

Auch die Hirnaktivität bei der Vorbereitung motorischer Reaktionen weist spezifische Muster auf. Die Muster unterscheiden sich in der Polarität der Ströme und in der Latenz, in der sie – nach Auftreten des Problems – ihre maximale Amplitude erreichen. Ferner sind unterschiedliche Aktivitäten an unterschiedlichen Stellen des Gehirns mehr oder weniger genau zu messen.

Aus Ergebnissen einer großen Zahl von EEG-Untersuchungen sind einige spezifische Zusammenhänge zwischen externen, vom Menschen wahrgenommenen Ereignissen und den durch sie ausgelösten elektrophysiologischen zerebralen Vorgängen bekannt: in der psycholinguistischen Versteheforschung besonders ein negatives Potential, das über die gesamte Kopfhaut hin zu messen ist, stärker im hinteren als im vorderen Feld; es

Zur Vertiefung

erreicht seine maximale Ausprägung durchschnittlich 400 ms nach Ereignisbeginn (daher N400) und wird durch semantische Ungereimtheiten in einer Äußerung ausgelöst. In seiner Folge entdeckt wurde ein positives Potential mit einer durchschnittlichen Latenz von 600 ms (P600). Es wird durch das Auftreten syntaktischer Abweichungen im Satz ausgelöst und geht oft mit der N400 einher, was zu der Annahme geführt hat, dass es eigentlich einen kognitiven Reparaturmechanismus anzeigt (vgl. Kutas/Van Petten 1994).

Nicht auf sprachliche zerebrale Aktivitäten bezogen, aber für deren Untersuchung ausnutzbar ist ein **motorikbezogenes Potential**. Es ist negativ, links- und rechtszentral messbar, tritt einige hundert Millisekunden (etwa 200) vor einer Muskelaktivität von Fingern der linken oder der rechten Hand auf und endet abrupt mit Einsetzen der overtten Bewegung. Es fällt auf der kontralateralen Hirnseite etwas stärker aus als auf der ipsilateralen, also bei Bewegung rechts etwas stärker auf der linken Hirnseite und umgekehrt. Was macht es speziell für die Sprachproduktionsforschung so nützlich?

Overtte Sprachproduktion geht mit artikulatorischer Aktivität einher und motorischer Aktivität geht immer zerebrale Aktivität voraus. Diese vermischt sich mit Potentialen von höheren kognitiven Prozessen so, dass letztere an der Kopfhaut nicht mehr separat zu messen sind. Infolge dessen sind EEG-Messungen für die Untersuchung von Sprachproduktionsprozessen nicht direkt zu verwenden. Wenn es aber gelingt, präartikulatorische Sprachproduktionsaktivitäten, also Konzeptualisierungs-, syntaktische und phonologische Kodierungsprozesse oder Zugriff auf das lexikalische Wissen mit Handbewegungen zu koppeln, dann lässt sich durch Messung der lateralen Potentiale die Abfolge der kognitiven Prozesse natürlich genauer erfassen als durch die overtte Handbewegung selbst. Der Bewegung eines Fingers oder auch seiner Nicht-Bewegung ist nämlich nur spät oder auch gar nicht zu entnehmen, wann die overtte Reaktion ausgelöst wurde; eine Nicht-Reaktion ist sogar unbrauchbar mehrdeutig. Das Nicht-Bewegen eines Fingers auf einer Push-Button-Box lässt nämlich nicht erkennen, ob gar keine Bewegung initiiert war oder ob eine initiierte Bewegung gestoppt worden ist, bevor die Muskeln tätig werden konnten. Dazwischen können nämlich bis zu 1500 ms verstreichen. Und über die zerebralen Vorgänge in genau dieser Zeit gibt die Messung der lateralen Potentiale Auskunft.

Dazu nutzt man eine besondere Eigenschaft des lateralen Potentials. Sofern der Mensch darauf vorbereitet ist, auf ein bestimmtes Ereignis hin den Finger zu bewegen, baut sich das Potential nach Eintreten des Ereignisses und vor der overtten Reaktion eine Zeit lang auf und kann während dieser Zeit bis zu einem gewissen »Point of no return« noch abgebrochen werden. Der Organismus begibt sich quasi in Bewegungsbereitschaft, ein Vorgang, der sich in dem lateralen Bereitschaftspotential (LRP = *lateralized readiness potential*) äußert, auch wenn die Bewegung letztlich abgebrochen wird. Dieses Potential zu ermitteln, ist kompliziert, weil zum einen alle zerebralen Ereignisse ausgefiltert werden müssen, die nichts mit der Bewegungsvorbereitung zu tun haben; auch der ipsi- und kontralaterale Stärkeunterschied muss ausgeglichen werden und natürlich die Schwankungen von Einzelfall zu Einzelfall. Wir gehen auf die Details der Berechnung nicht ein. Sie sind gut verständlich dargestellt bei Coles (1989).

Lokalisation des zweifachen lexikalischen Zugriffs im Gehirn

Unter Anwendung der in Kapitel 2.5 erwähnten Subtraktionsmethode haben Indefrey/Levelt (2000) versucht, die in den Lexikonzugriff involvierten Funktionen im Kortex zu lokalisieren.

Hier noch einmal kurz die Logik des Verfahrens. Wenn die verschiedenen Aktivitäten der Sprachverarbeitung in verschiedenen zerebralen Arealen lokalisiert sind, dann sollte dies dazu führen, dass sich das Bild, das entsteht, wenn zwei Aktivitäten durchgeführt werden, z. B. ein Wort lesen und laut aussprechen, sich von dem Bild unterscheidet, das sich aus der Beobachtung von nur einer Aktivität ergibt, z. B. beim stillen Lesen. Des Weiteren sollten sich nach der Lokalisationstheorie die jeweils aktiven Areale überlappen. Durch Subtraktion des zweiten, des Lesen-Bildes vom Lesen-und-Sprechen-Bild sollte das Areal bestimmbar sein, an dem die Aktivitäten des Aussprechens lokalisiert sind. Gelingt es nun, sprachliche Aufgaben zu finden und experimentell zu operationalisieren, deren Aktivitäten sich paarweise so überlappen, dass man jede der interessierenden Aktivitäten durch irgendeine Subtraktion isolieren kann, so sollte das jeweils resultierende Areal dasjenige sein, das der isolierten Aktivität entspricht. Trägt man alle Vergleichsergebnisse zusammen in eine Darstellung der Oberfläche der linken Gehirnhemisphäre ein, ergibt sich die Topographie, die Abb. 4.4 zeigt (Zeiten von uns eingefügt; RD u. JG).

Wo im Gehirn
findet das
Sprechen statt?

Inkrementelle Arbeitsweise

Eine weitere Frage zur Dynamik der Arbeitsweise des Sprachproduktionssystems betrifft das *Timing* der einzelnen Prozesse. Stellen wir uns vor, ein Sprecher hat entschieden, einem Hörer zu schildern, wie sich ein bestimmter Sachverhalt zugetragen hat, z. B. ein Unfall. Nach den ersten

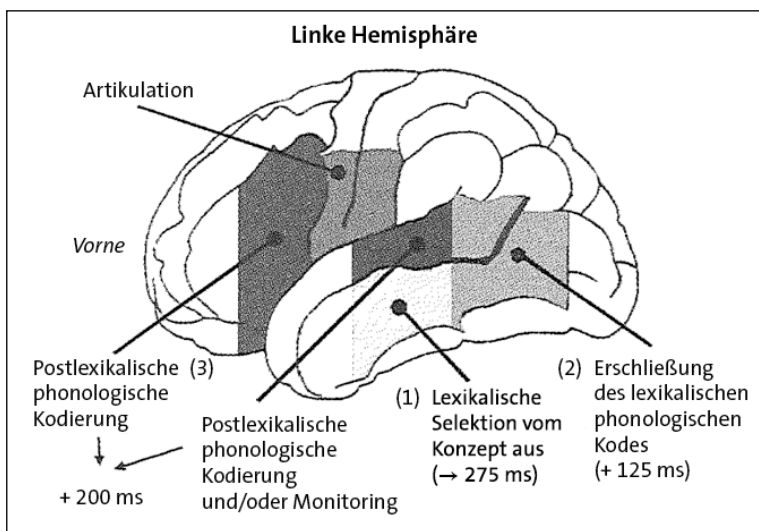
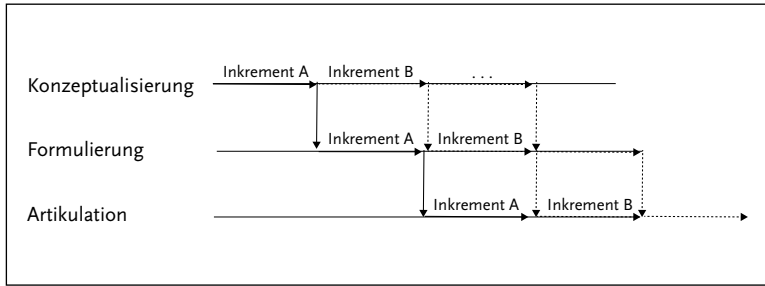


Abb. 4.4: Zusammenfassung mit einer ungefähren funktionalen Differenzierung der Gehirnareale, die an dem Informationsfluss beteiligt sind (vgl. Indefrey/Levelt 2000)

Abb. 4.5: Schematische Darstellung der seriell-parallelen Arbeitsweise bei der Äußerungsproduktion (vgl. Kempen/Hoenkamp 1987, S. 203; Levelt 1989)



Schritten im Prozess der Sprachproduktion, der Informationsselektion, der zumindest teilweise abgeschlossenen Segmentierung und Linearisierung, bereitet der Sprecher seine erste Äußerung vor. Hierfür ist es notwendig, die erste einer Folge von Messages aufzubauen. Die Frage lautet jetzt, wann relativ zum Beginn der Initialisierung des Message-Aufbaus, beginnt die Arbeit des Formulators, also die Wortselektion und die syntaktische Enkodierung? Wann beginnt die Arbeit auf der Ebene der phonologischen Enkodierung?

Es ist klar, dass zumindest ein kleines Stück der Message aufgebaut sein muss, bevor die nächsten Prozesskomponenten aktiv werden können. Doch »wieviel Message« benötigt der Formulator, um mit der Arbeit beginnen zu können? Eine andere Frage ist, was genau passiert auf der Ebene der Konzeptualisierung nachdem die erste Message bzw. ein erstes Teilstück davon fertig aufgebaut ist? Wartet die Konzeptualisierung mit dem Aufbau der nächsten Message bzw. des nächsten Teilstücks der Message, bis sie/es vollständig verbalisiert bzw. artikuliert wurde? Oder beginnt die Arbeit an der nächsten Message bereits, während an der aktuellen noch gearbeitet wird?

Letzteres würde bedeuten, dass an unterschiedlichen Segmenten der Äußerung auf unterschiedlichen Prozessebenen gleichzeitig gearbeitet wird. Für genau diese Arbeitsweise führten Kempen/Hoenkamp (1987) den Begriff der **inkrementellen Arbeitsweise** ein. Da ja immer mindestens ein kleines Stück (Inkrement) fertig sein muss, bevor es an die nächste Prozessebene weitergegeben werden kann, arbeitet das Sprachproduktionssystem in gewisser Weise seriell. Da aber gleichzeitig auf mehreren Ebenen gearbeitet wird, kann die Arbeitsweise ohne Weiteres auch als parallel bezeichnet werden. Der Begriff der Inkrementalität soll genau diese seriell-parallele Vorgehensweise erfassen (s. Abb. 4.5).

Experiment

Messung der Blickbewegung während einer Szenenbeschreibungsaufgabe

Griffin/Bock (2000) haben englischsprachigen Versuchspersonen die bildliche Darstellung eines Ereignisses mit zwei Beteiligten (beide belebt) am Bildschirm gezeigt. Die Ereignisse konnten entweder mit einem einfachen Aktivsatz oder mit einem Passivsatz beschrieben werden. Gemessen wurde, auf welche Bildregionen der Blick der Versuchspersonen bei der Produktion der Äußerung wie lange gerichtet war. Neben der Reihenfolge

der Blicke auf die beiden Beteiligten (Agens und Patiens) stand ebenfalls die zeitliche Relation zwischen den Blicken auf die Beteiligten und der Nennung der Beteiligten im Satz im Fokus des Interesses (*eye-voice-span*). Die Ergebnisse zeigen, dass die Handlungsbeteiligten nacheinander mit den Augen fixiert werden und dass ein Blick auf einen Beteiligten seiner Nennung etwa eine Sekunde vorausgeht. Außerdem zeigte sich, dass die Versuchspersonen nicht mit ihrem Blick auf einem Referenten verharren bis die Artikulation der korrespondierenden Nominalphrase abgeschlossen war, sondern dass die Artikulation des ersten mit der Vorbereitung (Formulierungsphase) der Artikulation des zweiten zeitlich überlappte. Da die Augenbewegung als ein direkter Indikator der Verteilung der Aufmerksamkeit des kognitiven Systems nachgewiesen ist, spricht dieses Ergebnis für die inkrementelle Arbeitsweise des Sprachproduktionssystems.

Mit **Inkrementalität** bezeichnet man eine Charakteristik des Sprachproduktionssystems, die darin besteht, dass auf den verschiedenen Prozessebenen gleichzeitig an verschiedenen Stückchen (Inkrementen) einer Äußerung gearbeitet wird.

Zum Begriff

Größe der Verarbeitungskremente: Wenn man die inkrementelle Arbeitsweise des Sprachproduktionssystems zu Grunde legt, stellt sich, wie gesagt, die Frage, wie groß die Verarbeitungseinheiten (*processing unit*) in den verschiedenen Teilsystemen sind. Die Frage lässt sich auch operational formulieren: Wie weit plant das Produktionssystem voraus? Die Beantwortung hängt offensichtlich davon ab, welches Teilsystem man betrachtet und wieviel sprachliches Material von einem Sprecher produziert werden soll.

Größe der Planungseinheit auf Message-Ebene

Griffin/Bock (2000) interessierten sich in ihrer oben bereits skizzierten Studie unter anderem dafür, ob Sprecher bei Szenenbeschreibungen die Message vollständig aufbauen, bevor die syntaktische Enkodierung beginnt. Um dies zu ermitteln, wurden die Blickbewegungen von Versuchspersonen unter vier verschiedenen experimentellen Bedingungen aufgezeichnet und analysiert. In allen Bedingungen wurden Versuchspersonen Bilder gezeigt, auf denen einfache Handlungen, jeweils mit einem Agens und einem Patiens, abgebildet waren. Vier Gruppen von Versuchspersonen erhielten jeweils eine unterschiedliche Aufgabe (es gab vier experimentelle Bedingungen): (1) Beschreibe das Ereignis sofort, (2) Beschreibe das Ereignis später, (3) Identifiziere den Patiens, (4) Inspiziere das Bild, um später verschiedene Angaben darüber zu machen.

Das Erfassen bzw. der Aufbau einer Ereignisstruktur stellt eine Notwendigkeit für die Beschreibung der Szenen dar (Bedingung 1 und 2). Ohne eine derartige Struktur könnten Sprecher keine syntaktischen Funk-

tionen wie Subjekt oder Objekt spezifizieren. Die Ereignisstruktur spielt also auf der Message-Ebene eine Rolle, weil die Message den Input für die syntaktische Enkodierung darstellt. Das Erfassen bzw. der Aufbau einer Ereignisstruktur stellt aber auch in der dritten Bedingung eine Notwendigkeit dar, denn nur so kann angegeben werden, wer der Patiens eines Ereignisses ist.

Die Analyse der Daten zeigte für die ersten drei Bedingungen, dass Versuchspersonen nach etwa 350ms signifikant stärker auf einen der beiden Referenten blickten (auf den Agens in Bedingung 1 und 2, auf den Patiens in Bedingung 3). In der vierten Bedingung zeigte sich eine ähnliche Präferenz nicht, jedenfalls nicht im selben Zeitfenster.

Die Autoren schlussfolgerten aus diesen Ergebnissen, dass der Message-Aufbau vollständig abgeschlossen ist (Referenten sind als Agens oder Patiens spezifiziert), bevor die Formulierungsphase beginnt und dass alle Prozesse vor dem Beginn der Formulierung nicht länger als 350ms in Anspruch nehmen.

Größe der Verarbeitungseinheit auf syntaktischer Ebene

Inkrement =
Verarbeitungse-
inheit

Auf die Frage, wie groß wohl ein Inkrement ist, das während der syntaktischen Enkodierung in einem Prozessdurchgang verarbeitet wird, kann auf der Grundlage aktueller Forschungsergebnisse nur die Antwort gegeben werden: Das hängt davon ab, vor welcher Aufgabe das Sprachproduktionssystem steht, genauer gesagt, wie komplex die sprachliche Äußerung ist, die produziert werden soll.

Führen wir uns noch einmal vor Augen, was bei der syntaktischen Kodierung geschieht: Die Leistung des syntaktischen Kodierens besteht darin, die Bedeutung der Message und die Informationsstruktur so auf eine Wortfolge abzubilden, dass zugleich den semantisch-pragmatischen Präzedenzeigenschaften jeder Komponente der Message und den syntaktischen Strukturbeschränkungen der Sprache Rechnung getragen ist. Das legt es nahe, dass eine syntaktische Verarbeitungseinheit gerade eine solche Teilstruktur ist, die intern nicht weiter nach semantisch-pragmatischen Präzedenzeigenschaften linearisiert ist, sondern nach syntaktischen. Anders gesagt: Eine syntaktische Verarbeitungseinheit ist das strukturelle Ensemble, das zwar als Ganzes semantisch und kontextuell in der Satzkette variabel positioniert wird, dessen Komponenten intern aber durch feste Verarbeitungsmuster positionsfest sind. Sollte sich herausstellen, dass zwei konzeptuelle Komponenten in einer Phrase ohne Rücksicht auf ihre individuellen Präzedenzeigenschaften (strukturiert und) linearisiert werden, dann würde man das als Hinweis auf das Wirken von syntaktisch autonomer Verarbeitung halten müssen.

Diesem Phänomen gilt eine Serie von Experimenten von Pechmann (1994). Sie hat die Produktion von Nominalphrasen zum Gegenstand. Diese weisen (im Deutschen) nämlich eine solche Merkwürdigkeit auf. Bildet ein Sprecher zur Bezeichnung eines Buches, das rot und relativ klein ist, eine attributiv erweiterte Nominalphrase, so ergibt sich in aller Regel: ... *das kleine rote Buch*. Bemerkenswert daran ist, dass die kognitive Zugänglichkeit von *rot* höher ist als die von *klein* und die von *Buch* höher

als die von *klein* aber geringer als die von *rot*. Ohne Einwirkung anderer als der rein semantischen Präzedenz sollte die Reihenfolge also sein: *das rote Buch kleine*. Die semantische Präzedenz lässt sich in separaten Reaktionszeitexperimenten (Farbe, Größe, Objekt) mit Bildbenennungsaufgaben ermitteln. Ebenso, und unabhängig davon, lässt sich in Ratingexperimenten und in Objektbezeichnungsexperimenten die kanonische Reihenfolge von Größen- und Farbattribut und Objektnamen ermitteln.

Pechmanns Befunde besagen unzweideutig, dass während der syntaktischen Enkodierung etwas produziert wird, das mit der semantischen Präzedenz nicht übereinstimmt, nämlich Ausdrücke der Form ... *das kleine rote Buch* ... bei Präzedenzverteilung wie in ... *das rote Buch kleine* ... Der Effekt der syntaktisch abgekapselten Verarbeitung ist offenbar sehr stark; er tritt auch bei der umgekehrten Prozedur auf, also beim syntaktischen Verstehen. Das erweist sich dadurch, dass Versuchspersonen für die Bewertung, ob eine Nominalphrase (*das gelbe kleine Buch*) mit einer zuvor auf dem Bildschirm präsentierten (*das kleine gelbe Buch* oder *das gelbe kleine Buch*) identisch ist oder nicht, immer gleich viel Zeit benötigen; die Beurteilungszeit wird von der Reihenfolgevariation nicht beeinflusst. Aus diesem Befund lässt sich schließen, dass die komplexe NP intern nicht inkrementell kodiert wird, dass also die NP für die Wortstellungsprozeduren eine geschlossene Verarbeitungseinheit des syntaktischen Kodierens ist.

Flexibilität inkrementeller syntaktischer Verarbeitung: Wagner/Jescheniak/Schriefers (2010) demonstrieren in einer Studie, in der Versuchspersonen Äußerungen mit mehreren Phrasen produzieren sollten (z. B. *Der Becher ist neben dem Messer*), dass die Vorausplanungseinheit auf syntaktischer Ebene keine strukturell fixierte Größe hat, sondern hier sowohl die Komplexität der einzelnen Phrasen (z. B. *Der kleine Becher ist neben dem langen Messer*) als auch die zur Verfügung stehenden kognitiven Ressourcen eine Rolle spielen. Versuchspersonen sahen in dieser Studie Bilder, auf denen jeweils zwei Objekte dargestellt waren, auf die sie mit einem einfachen Satz Bezug nehmen sollten. Zusätzlich wurde in jedem Versuchsdurchgang auditiv ein Distraktor präsentiert, der allerdings explizit ignoriert werden sollte. Die in den Experimenten verwendeten Distraktoren waren entweder mit dem ersten oder mit dem zweiten Objektnamen semantisch verwandt. Gemessen wurde die Sprechbeginnzeit. Die Logik ist die: Wenn ein mit dem zweiten Nomen semantisch verwandter Distraktor die Sprechbeginnzeit beeinflusst, dann kann dies als Evidenz dafür gelten, dass Versuchspersonen die Lemma-Informationen für beide Objektnamen vor dem Sprechbeginn abgerufen haben. Es zeigte sich nun, dass ein derartiger Effekt auftrat, und zwar dann, wenn Sprecher einfache Phrasen bilden sollten. War die Komplexität der Phrasen größer (*Der kleine Becher* vs. *Der Becher*) wurde der Effekt kleiner und wenn Versuchspersonen zusätzlich zu der Sprachproduktionsaufgabe noch eine nicht-sprachliche Entscheidungsaufgabe durchführten, verschwand er vollständig.

Ebenfalls zur Frage, wie groß die Planungseinheiten auf der Ebene der syntaktischen Enkodierung ist, sprechen Befunde aus einer Studie von vgl. Schriefers/Teruel/Meinshausen (1998), die weiter unten noch in einem anderen Zusammenhang vorgestellt wird.

Distraktor =
Ablenkword

4.3.4 | Die Erzeugung der Satzstruktur

Von der präverbalen Message zur Satzstruktur

Vor dem Hintergrund des bisher Gesagten scheint es mehrere Möglichkeiten zu geben, wie genau das Sprachproduktionssystem die syntaktische Enkodierung bewerkstelligt, also die Syntax eines Satzes generiert: Entweder erfolgt zuerst ein Message geleiteter Strukturaufbau, dem lexikalische Einfügung folgt oder es kommt zuerst zu einem Message geleiteten Lexikonzugriff und anschließend zum syntaktischen Strukturaufbau.

Wie immer man an die Klärung dieser Fragen herangeht, empirisch oder theoretisch, die Antwort muss der Tatsache Rechnung tragen, dass eine gelungene Äußerung drei Belangen genügt:

- Sie drückt den Inhalt aus, den sie ausdrücken soll, ist also semantisch treffend.
- Sie ist in Wortstellung und Intonationskontur kontextuell angepasst, d. h. sie drückt – linguistisch gesprochen – die diskursbedingte Informationsverteilung aus (Topik- vor Fokuginformation, erhaltene vor neuer Information etc.).
- Sie ist syntaktisch wohlgeformt, d. h. sie verletzt keine syntaktische Struktur- und Reihenfolgeeinschränkung.

Das sprachliche Kodierungssystem steht – so gesehen – vor der Aufgabe, die treffenden sprachlichen Mittel in der angemessenen Reihenfolge zusammenzustellen und für die Artikulation vorzubereiten. Das scheint in der spontanen Rede zügig vonstatten zu gehen, wenn man eine Output-Rate von rund sechs Silben pro Sekunde als zügig einstuft.

Strukturinformationen in der Message: Garrett (1980) schlägt vor, dass der Message während der syntaktischen Enkodierung Informationen entnommen werden, die zum Aufbau von syntaktischer Struktur gebraucht werden. Wenn die Message, wie Garrett annimmt, ein Ausdruck in einer logischen Symbolsprache ist, dann ist das Information über:

- den **logischen Typ der Konzepte**, also die Frage, ob es sich um ein ein-, zwei- oder dreistelliges Prädikat, eine Anapher, einen Skopusoperator, einen Konnektor, Modal- und andere Operatoren usw. handelt.
- die **semantische Struktur**; sie lässt erkennen, wie die Konzepte zueinander in Verbindung stehen. Nach Auffassung einiger Semantiker auch, welche Funktion gewisse Konzepte relativ zu andern erfüllen, also Akteur, Patiens oder Ort eines Ereignisses zu sein (thematische Rollen).
- **zeitliche Referenz und Aspekt**, also die Frage, über welches Intervall der Diskurswelt geredet wird, wie es zeitlich zum Äußerungszeitpunkt liegt und wie das Intervall des Message-Ereignisses relativ zum topikalisierten Intervall der Rede liegt.

In einem ersten Schritt wird aus den Informationen der Message eine abstrakte funktionale Struktur abgeleitet. In einem zweiten Schritt wird diese funktionale Struktur mit Hilfe sogenannter Planungsrahmen linearisiert. Die Spezifizierung welcher Referent zum Beispiel syntaktisches Subjekt oder welcher Referent direktes Objekt im zu produzierenden Satz wird, erfolgt also in einem separaten und früheren Schritt als die Festlegung der Reihenfolge, in der die syntaktischen Phrasen im Satz auftauchen.

Eine solche Arbeitsweise beim Strukturaufbau wird als **konzeptuell-getriebener** Strukturaufbau bezeichnet (vgl. Fisher/Gleitman/Gleitman 1991; Schriefers et al. 1998). Auf den Punkt gebracht lautet die Annahme also, dass die auf Message-Ebene aktivierten konzeptuellen Informationen genügen, um eine syntaktische Struktur aufzubauen, in der alle relationalen Informationen abgebildet sind. Daraus folgt, dass ein für den Aufbau der Satzsyntax wichtiges Element, nämlich das Verblemma, nicht notwendigerweise aus dem mentalen Lexikon ausgewählt sein muss, bevor mit dem Strukturaufbau begonnen wird. Zumindest teilweise wird dies durch Ergebnisse aus einer Studie von Schriefers et al. (1998) belegt.

Bild-Wort-Interferenz-Paradigma im Satzproduktionskontext (Schriefers et al. 1998)

Experiment

In diesem Szenenbeschreibungsexperiment sollten Versuchspersonen Ereignisse beschreiben, die auf Bildern dargestellt waren. Die Probanden hörten den Anfang eines Satzes (*Und auf dem nächsten Bild ... oder ... Auf dem nächsten Bild sieht man, wie ...*). Dann erscheint ein Bild von einem Ereignis mit einem oder zwei Referenten. Es soll so beschrieben werden, dass die beschreibende Äußerung syntaktisch zu dem Einleitungssatz passt, also einmal mit einer Subj-(Obj)-Verb-Äußerung, einmal mit Inversion, Verb-Subjekt-(Obj). In beiden Aufgaben wurde der lexikalische Zugriff auf Komponenten der Äußerung durch ein Ablenkword (*distraction item*) gestört, das zu geeigneter Zeit links und rechts neben der Ereignisdarstellung geschrieben erscheint. Es soll den Zugriff auf das Verblemma verzögern. Das Verfahren heißt **Bild-Wort-Interferenz-Paradigma**. Es dient dazu, den Zeitpunkt des Lexikonzugriffs zu ermitteln. Das gelingt dadurch, dass das Ablenkungswort während des Produktionsvorgangs zu unterschiedlichen Zeiten nach Auftauchen des Bildstimulus eingeblendet wird. Die Verschiebung heißt **Stimulus-Onset-Asynchrony** (SOA). Hat das Ablenkword nur zu einem SOA-Wert einen deutlichen Einfluss, so kann man daraus folgern, dass dieser in dem Intervall des Lexikonzugriffs liegt, was man damit zeitlich eingegrenzt hat. Als Wörter mit ablenkendem Effekt haben sich in früheren Objektbenennungsexperimenten und Produktionsexperimenten bedeutungsähnliche Wörter erwiesen (vgl. Schriefers et al. 1990; Meyer 1996).

Die Ergebnisse von Schriefers et al. (1998; hier nur Exp. 2) zeigen, dass bei Produktion eines S-V-O Satzes auf das Subjekts- und das Verblemma nicht gleichzeitig zugegriffen wird, sondern zeitlich gestaffelt, zuerst auf das Lemma des Subjektsnomens dann auf das Verblemma.

Strukturinformationen im Lemma: Nun weisen allerdings (Kempen/Hoenkamp 1987) darauf hin, dass die syntaktische Struktur in wesentlichen Einzelheiten von lexikalischer Information abhängt. So hätten die Sätze *John wanted to hit Peter* und *John knew he hit Peter* dieselbe Message-Struktur. Aus der Message des ersten darf aber nicht die (falsche) syntaktische Struktur **John wanted he hit Peter* abgeleitet werden. Die Information, die dies verhindern könnte, steht im Lemma von *want*. Da solche falschen Sätze

Belebte oder
kürzlich erwähnte
Entitäten
erscheinen
häufig früh in
einem Satz

tatsächlich nicht produziert werden, müsse angenommen werden, dass die lexikalische Information vor dem Aufbau der Satzstruktur zur Verfügung steht (lexikalisch-getriebener Strukturaufbau). Dieser Konzeption folgt prinzipiell das Sprachproduktionsmodell von Levelt (1989).

Der Einfluss der konzeptuellen Zugänglichkeit auf die sprachliche Struktur: Kathryn Bock hat in einer Serie von Experimenten herausgefunden, dass Eigenschaften von Referenten wie Belebtheit, Konkretheit vs. Abstraktheit u. a. sich auf ihre Positionierung in einer Äußerung auswirken. Sie erklärt das mit Unterschieden in der **konzeptuellen Zugänglichkeit** (vgl. Bock/Warren 1985). Die sog. Voranwirkung der Belebtheit ist vielfach überprüft und bestätigt worden. Einen Überblick über die wichtigsten Arbeiten enthalten van Nice/Dietrich (2003) und Jaeger/Norcliffe (2009). Konzeptuelle Zugänglichkeit beschränkt sich aber nicht auf Belebtheit. Es konnte auch nachgewiesen werden, dass informationsstrukturelle Spezifizierungen (Topik/Fokus) und Vorerwähntheit einen Einfluss auf die konzeptuelle Zugänglichkeit und damit die Positionierung eines Referenten im Satz hat (vgl. Prat-Sala/Brannigan 2000).

Nun könnte die Voranstellung von belebten Referenten am einfachsten damit erklärt werden, dass belebte Nominalphrasen in der positionalen Repräsentation eher vor diejenigen mit den nicht-belebten NPs platziert werden. Verschiedene experimentelle Befunde sprechen jedoch gegen eine solche Erklärung. So lässt sich der Belebtheitseffekt (belebte Referenten voran) zum Beispiel nur in Sätzen nachweisen (*The farmer bought the refrigerator*), aber nicht in koordinierten Phrasen (*the farmer and the refrigerator*). Dies führt zu der Annahme, dass die Belebtheit nicht per se für den Belebtheitseffekt verantwortlich ist, sondern, dass die Belebtheit in einem Zusammenhang mit der Ereignisstruktur, also mit den in der Message spezifizierten thematischen Rollen, wie Agens, Patiens oder Thema steht. Konzeptuelle Zugänglichkeit wirkt sich demnach nicht direkt auf die (lineare) Position eines Elements im Satz aus, sondern äußert sich vielmehr in der Zuweisung thematischer Rollen zu grammatischen Funktionen, die wiederum die relationale syntaktische Hierarchie in einem Satz reflektieren (vgl. Bock/Warren 1985).

Diese Beobachtung unterstützt Garretts (1980) Annahme, dass auf dem Weg von der Message zur syntaktischen Struktur zwei Prozesse stattfinden, nämlich eben die Verknüpfung thematischer Rollen mit syntaktischen Funktionen (z. B. verknüpfe Agens mit Subjekt) sowie die Linearisierung der Satzkonstituenten. Zu beachten ist, dass anders als Garrett (1980), Bock/Levelt (1994) in ihrer Beschreibung der Abläufe während der grammatischen Enkodierung die positionale Verarbeitung zu den syntaktischen Prozessen während des Sprachproduktionsprozesses zählen und nicht zu den phonologischen.

Wiederverwendung von Strukturinformationen: Noch ein anderes Phänomen darf nicht unerwähnt bleiben, wenn es um den Aufbau der Satzstruktur geht: **Strukturelles Priming**. Untersuchungen in unterschiedlichen Sprachen, unterschiedlicher konkreter syntaktischer Strukturen und mit unterschiedlichen Varianten des Priming-Paradigmas weisen darauf hin, dass Sprecher eine Tendenz zeigen, syntaktische Strukturen, die sie vor kurzem benutzt haben, um einen Sachverhalt zu verbalisi-

sieren, wiederverwenden, sozusagen »recyclen«. So konnten Bock/Loebell (1990) beispielsweise nachweisen, dass die Wahrscheinlichkeit größer ist, eine visuell dargebotene Szene mit einem Passivsatz zu beschreiben, wenn Versuchspersonen direkt vor ihrer Bildbeschreibung dazu veranlasst werden, einen Passivsatz laut vorzulesen (*The construction worker was hit by the bulldozer* → *The foreigner was confused by the blinking traffic light*). Man beachte, dass die Bedeutung des Prime-Satzes völlig verschieden von der Bedeutung des Target-Satzes ist.

Im selben Experiment wurde auch gezeigt, dass die Verwendung eines Passivsatzes auch durch partielles Priming hervorgerufen werden kann (hier durch Sätze mit lokativer Ergänzung). Sprecher produzierten mit einer höheren Wahrscheinlichkeit einen Passivsatz zur Beschreibung einer Szene, wenn sie kurz vorher einen Satz mit lokativer Ergänzung laut lasen (*The construction worker was digging by the bulldozer* → *The foreigner was confused by the blinking traffic light*). Zu beachten ist, dass ein Priming-Effekt immer relativ zu einer Kontrollbedingung evaluiert wird. Im vorliegenden Fall war das ein Aktivsatz (*The construction worker drove the bulldozer*). Derartige Priming-Effekte können weder direkt auf spezifische Ereignisstruktur, die in einer Message kodiert ist, zurückgeführt werden, noch auf Strukturinformationen des Verblemmas. Vielmehr scheint es so zu sein, dass sich das Sprachproduktionssystem bestimmter syntaktischer Rahmen (Frames) bedient, um die inhaltliche Information der Message auf eine Satzstruktur abzubilden.

Charakterisierung des Priming-Effekts: Zwei Punkte sind in diesem Zusammenhang noch wichtig. Beide treten besonders in Augenschein, wenn man Befunde aus Priming-Studien betrachtet, die sich auf die sogenannte Dativ-Alternation richten, zum Beispiel im Englischen, *Peter gave Mary the book* vs. *Peter gave the book to Mary*. In der *double-object*-Konstruktion (DO), dem ersten Satz, besteht auf der syntaktischen Ebene die Reihenfolge Subjekt, direktes Objekt, direktes Objekt. Auf der konzeptuellen Ebene lautet die Reihenfolge der konzeptuellen Rollen: Agens, Rezipient, Thema. In der *prepositional-object*-Konstruktion (PO) ist die Reihenfolge der Verbargumente: Subjekt, direktes Objekt, Präpositionalobjekt; auf der konzeptuellen Ebene lautet die Reihenfolge der konzeptuellen Rollen: Agens, Thema, Rezipient. Für beide Konstruktionen wurden balancierte Priming-Effekte nachgewiesen; PO-Sätze primen also PO-Sätze und DO-Sätze primen DO-Sätze gleichermaßen (Traxler et al. 2002).

Der erste Punkt betrifft die Frage, was eigentlich genau in Experimenten, in denen ein derartiger Effekt beobachtbar ist, geprimt wird. Ist es die syntaktische Funktionszuweisung der konzeptuellen Rollen (z. B. bilde Thema als direktes Objekt ab, bzw. bilde Thema als indirektes Objekt ab), oder ist es die lineare Anordnung der konzeptuellen Rollen? Ersteres wurde als »conceptual-to-function-mapping« bezeichnet und die zweite Möglichkeit als »conceptual-to-linear-mapping« (Cai et al. 2012).

Weil in vielen Sprache, so auch im Englischen, nicht zwischen der einen oder anderen Möglichkeit unterschieden werden kann, da ja die Reihenfolge der konzeptuellen Rollen immer mit einer bestimmten Funktionszuweisung einhergeht, untersuchten Cai et al. (2012) diese Frage im Mandarin Chinesischen. Mandarin stellt nämlich grammatische Struktu-

ren bereit, bei denen die lineare Abfolge von konzeptuellen Rollen nicht nur an genau eine Art der syntaktischen Funktionszuweisung gebunden ist. Die Tatsache, dass sich Mandarin-Sprecher sowohl dahingehend primen lassen, die Reihenfolge der konzeptuellen Rollen in der Satzstruktur beizubehalten als auch dazu die Funktionszuweisung vom Prime-Satz auf den Target-Satz zu übernehmen, und zwar unabhängig voneinander, deuteten die Autoren als Evidenz für einen einstufigen Prozess des Strukturbaus. In einem solchen Modell wird die Funktionszuweisung und die Berechnung der linearen Abfolge der Satzkonstituenten in einem Schritt erreicht. Anders als zum Beispiel Garrett (1980) oder Bock/Levelt (1994) gehen Traxler et al. (2002) und Cai/Pickering/Branigan (2012) davon aus, dass die syntaktische Funktionszuweisung und die lineare Positionierung von Konstituenten in einem Schritt erfolgt.

Der zweite Punkt betrifft den Umstand, dass strukturelles Priming stärker hervortritt, wenn in Prime- und Target-Satz dasselbe Verb benutzt wird (*lexical boost effect*). Wären syntaktische Rahmen tatsächlich in gewisser Weise unabhängig von einzelnen Verb-Lemmata, wie die Befunde von Bock/Loebell (1990) nahelegen, würde man diesen gut bestätigten Befund nicht erwarten.

Fazit: Informationen, die zum Aufbau der Syntax eines Satzes notwendig sind, stecken teilweise in der Message. Zu denken ist dabei vor allem an die Anzahl der Referenten aus der sich ja prinzipiell die Anzahl der syntaktischen Argumente des Verbs ergeben und damit gewisse Entscheidungen über die Markierung und Positionierung von Konstituenten gefällt werden können. Des Weiteren bestimmen Eigenschaften wie Belebtheit oder Diskursualienz, also nicht-syntaktische Informationen, die Position von Konstituenten innerhalb eines Satzes. Der Aufbau der Satzsyntax sollte sich also in gewissem Maße ohne initiale Verbselektion (Lemma-Zugriff) vollziehen können. Allerdings gibt es auch Informationen über den Strukturaufbau, die man eher mit dem Verb-Lemma assoziieren würde, die Konstruktion der Satzsyntax also der Wortselektion nachgestellt sein sollte. Zu denken ist hier vor allem an Subkategorisierungsinformationen (*kennen* z. B. selektiert ausschließlich ein direktes Objekt, *wissen* hingegen selektiert einen Objektsatz).

4.3.5 | Die Erzeugung der Lautkette

In diesem Abschnitt geht es nun um die Prozesse bei der Umwandlung der durch die syntaktische Enkodierung entstandenen Oberflächenstruktur in eine Repräsentation, die es ermöglicht, das Artikulationssystem in Gang zu setzen, also Schallwellen zu produzieren, die als Sprache verstanden werden. Hierfür werden zwei Schritte angenommen: die **phonologische Enkodierung** und die **phonetische Enkodierung**.

Phonologische Enkodierung

Die Ausgangsinformation der lautlichen Kodierung bildet, wie gesagt, die linear geordnete syntaktische Struktur der Äußerung mit Lemma-Information zu jedem Wort, flexionsmorphologischer Spezifikation und intonati-

onsrelevanter Information. Der erste Schritt der phonologischen Enkodierung ist der Abruf der Lexem-Informationen im mentalen Lexikon. Diese Informationen werden manchmal auch **phonologischer Code** genannt.

Ein **phonologischer Code** besteht aus einem geordneten Set phonemischer Segmente. Außerdem enthält er metrische Marker, also Informationen über die Betonung eines Wortes, falls diese vom Standard in einer Sprache abweicht (*Kaffee* vs. *Café*).

Zum Begriff

Resilbifizierung: Doch wie verhält es sich mit Informationen über Silben? Sind die zu einem Wort gehörenden Phoneme auch in Form von Silben gespeichert?

Je nach Kontext verändern Wörter bekanntlich ihre lautliche Form. Das Lexem des Wortes *Land* ist die Phonemfolge /lant/. Die Genitiv-Singular-Endung ist /-es/. Das ergibt die Folge /lant-es/; ausgesprochen als /landəs/ und hier ist der letzte Konsonant des Stammlexems offenbar stimmhaft. Auch die Betonung eines Wortes wird dem Kontext angepasst: *Afrika* gegenüber *afrikanisch*. Die Veränderungen können aber auch noch weitergehen, z. B. beim schnellen Sprechen in spontaner Rede. So ist es im Deutschen nicht unüblich, dass das Lexem /kɪpən/ in der Äußerung *Die Sonnenschirme kippen um* als [kɪpm] ausgesprochen wird, eine reguläre Folge von Tilgung des Schwa-Lauts der Endung in Verbindung mit Assimilation des [n] an das vorangehende bilaterale [p] (progressive Nasalassimilation). Generell spricht man bei diesem Phänomen von Resilbifizierung.

Slots and Fillers: Werden nun also phonemische Segmente organisiert in Silbenform hintereinander aufgereiht, dann ihre Struktur zerstört und das lautliche Material (verändert) in eine neue lautliche Form gegossen? Für diesen mehrphasigen Prozess, den Shattuck-Hufnagel (1979) in einer einflussreichen Arbeit als einen *Slots-and-Fillers*-Prozess modelliert hat, sprechen zahlreiche Beobachtungen. Für die eigenständige Rolle der Silbengliederung bei der phonologischen Kodierung spricht die Tatsache, dass Lautvertauschungen sozusagen silbentreu sind. Die vertauschten Einheiten nehmen systematisch die Silbenposition ihres Tauschpartners ein, zum Beispiel *Masenrärer* statt *Rasenmäher*, *C-Moss-Melle* statt *C-Moll-Messe*; die **Silbentreue** bewährt sich auch bei Lautverschiebungen: *Fischfußball* statt *Tischfußball*, *bei der Schlange halten* statt .. *Stange*. Shattuck-Hufnagel nahm deshalb an, im mentalen Lexikon seien die Silbenstruktur jedes Lemmas und die segmentale Kette vorhanden und im Zuge der phonologischen Kodierung würden die beiden getrennt verarbeitet. Es würde eine Folge von «leeren» Silbenstrukturen mit prosodischer Information erzeugt und in einem separaten Prozess Lautsegmente eingefüllt, wobei es zu Fehlbesetzungen der Slots kommen kann.

Nun reichen die erwähnten Assimilationsprozesse auch über Wortgrenzen hinweg wie in *Komm aber nicht zu spät*, was in fließender Rede als [kɔma:bɐ ...] gesprochen wird. Die beiden Wörter *komm* und *aber* sind lautlich durch [m] verbunden, das gleichzeitig den Auslaut von [kɔm] und den Anlaut von [ma: ...] bildet. Der Glottisverschlusslaut am Anfang von

Sind Phoneme in Silben organisiert?

/ʔa:bər/ ist getilgt. Das geschieht aber nicht in allen Kontexten, z. B. nicht in der Verbindung *Komm bald*, was nicht zu [kɔmalt] wird. Diese Beobachtung zwingt zu der Annahme, dass zwar zuerst die Kette der morphemspezifischen Lexeme gebildet wird, die Silbengliederung aber erst stattfindet, nachdem es zur Assimilation kam. Entsprechende Beispiele sind: *Ich finde täglich neue Tippfehler* wird zu [...fintɛ:gliç...] mit [t] als bisyllabischem Element gegenüber *Ich finde das nicht lustig*, gesprochen als [...fɪndas] mit bisyllabischem [d]. Wir müssen also annehmen, dass nach der Selektion der Lexeme zuerst eine Kette von Phonemen entsteht, über diesen gegebenenfalls Assimilationsprozesse stattfinden und erst spät die Silbengliederung stattfindet. Geht man nun davon aus, dass die Silbenstruktur Bestandteil der Lexem-Information ist, an ihr orientiert eine Folge von lexikalischen Silben erstellt würde und dann Assimilationen stattfinden, so müsste man mit einem weiteren Prozess rechnen, in dem die lexikalisch basierte Silbengliederung an die veränderte Lautkette angepasst wird.

Prosodifizierung: Ein alternativer und heutzutage von den meisten Forschern vertretener Ansatz ist, anzunehmen, dass die Silbenstruktur nicht als Teil der Lexem-Information abgerufen, sondern quasi »on the fly« aufgebaut wird. Dieser Vorschlag geht auf Levelt et al. (1999) zurück, die die Generierung der Silbenstruktur auf diese Weise »Prosodifizierung« nennen (vgl. Cholin/Schiller/Levelt 2004). Die Idee ist also, dass phonemische Segmente aktiviert und inkrementell zu aufeinanderfolgenden Silben kombiniert werden. Gleichzeitig wird den aufeinanderfolgenden Silben dabei die korrekte metrische Information zugewiesen, entweder so, wie es der Standard in einer Sprache verlangt oder entsprechend der Information, die als Teil des phonologischen Codes gespeichert sind (bei Abweichungen vom Standard). Die inkrementelle Prosodifizierung folgt auf der einen Seite universalen Silbifizierungsregeln und auf der anderen Seite sprachspezifischen Beschränkungen (phonotaktische Regeln).

Der zeitliche Ablauf der phonologischen Enkodierung: Der zeitliche Ablauf der phonologischen Kodierung wird von drei Gegebenheiten bestimmt: a) vom generellen Prinzip der inkrementellen Produktion, b) von der Reihenfolge, in der Lemmata aktiviert sind und c) von der Zeit, in der das Lexem aktiviert wird. Häufige Wörter werden schneller ausgesprochen und die im Lauf des Spracherwerbs gelernte Häufigkeitsinformation ist auf der Lexemebene gespeichert, wie durch RT-Experimente (Jescheniak/Levelt 1994) gezeigt worden ist. Die Idee, die diesem Experiment zu Grunde liegt, ist gleichermaßen pfiffig wie das Ergebnis überzeugend ist.

Experiment

Experiment zur Untersuchung der Lexem-Frequenz-Theorie

Das Experiment von Jescheniak/Levelt (1994) war dazu angelegt, zwischen zwei Theorien zu entscheiden, derjenigen, die Frequenzinformation sei eine Eigenschaft des Lemmas und derjenigen, sie sei eine Lexemeigenschaft. Versuchspersonen sollten zu einem englischen Stimuluswort das niederländische Zielwort so schnell wie möglich nennen. Das Material war so ausgesucht, dass darunter Items waren, die im Niederländischen mit homophonen Wortformen zu übersetzen waren, übertragen

ins Deutsche also *goal* und *gate*, die beide als *Tor* zu übersetzen sind. Die Homophone wurden so ausgewählt, dass es welche gab, deren eines Lemma häufig und deren anderes selten ist. Als Kontrolle dienten häufige und seltene Nicht-Homophone. Wäre die Frequenz eine Lemmaeigenschaft, sollten die Reaktionszeiten für hochfrequente Homophone denen für hochfrequente Nicht-Homophone ähneln, und die für niederfrequente Homophone denen für niederfrequente Nicht-Homophone. Ist die Frequenzeigenschaft aber eine Lexemeigenschaft, so sollten die niederfrequenten Homophone deutlich schneller genannt werden als die niederfrequenten Nicht-Homophone. Das Ergebnis spricht für die **Lexem-Frequenz-Theorie**.

Größe der Einheiten während der phonologischen Enkodierung: Es bleiben, wie immer bei der Betrachtung von Informationsverarbeitungsvorgängen, zwei Fragen: Was sind die Einheiten der Kodierung und in welcher Reihenfolge werden sie bearbeitet? Die theoretische Antwort auf die erste Frage ist immer dieselbe. Die Einheit eines Informationsverarbeitungsprozesses ist gerade das Stück Information, das der Prozess als Input für seine Arbeit braucht und das seinen automatischen Start verursacht. Den Input der flexionsmorphologischen Ausbuchstabierung muss also das ›Paket‹ an morphologischen Merkmalen bilden, das die Aktivierung der zugehörigen phonologischen Form ermöglicht und veranlasst. Das Steigerungsmorphem ›Komparativ‹ zum Adverb *schnell* zum Beispiel alleine genügt für den Zugriff auf das Lexem <-er> und seine Segmente. Anders verhält es sich mit der Kodierung der 2. Person, Singular, Präsens, Indikativ, Aktiv zum Verbstamm *geb*. Die resultierende Wortform [gipst] kann phonologisch nur eindeutig bestimmt werden, wenn alle fünf genannten Informationen vorliegen und zusätzlich die Information, dass *geb* unregelmäßig flektiert und zwar genau so, dass durch den sog. e-i-Wechsel der Stammvokal sich in der 2. und 3. Person Singular und im Imperativ Singular ändert.

Phonetische Enkodierung

Der Output der phonologischen Enkodierung ist also das phonologische Wort, das für seine metrischen, syllabischen und segmentalen Eigenschaften spezifiziert ist. Im letzten Schritt vor der Artikulation wird nun die phonologische Repräsentation dessen, was in Schallwellen umgesetzt werden soll, in eine phonetische Repräsentation übertragen. Eine derartige Repräsentation stellt man sich als Reihe von Anweisungen an das motorische System vor, also das System, das alle an der Artikulation beteiligten Muskelbewegungen steuert.

Einige Autoren sprechen in diesem Zusammenhang von einzelnen hochspezialisierten Programmen, die für die Hervorbringung genau einer phonetischen Silbe verantwortlich sind. Diese werden artikulatorische bzw. syllabische Gesten genannt (*syllabic gestures*). Die Aneinanderreihung solcher syllabischen Gesten innerhalb des phonetischen Plans ermöglicht es also einem Sprecher, aufeinanderfolgende Silben zu erzeugen,

Initialisierung
der für die Laut-
produktion
notwendigen
Muskelgruppen

die genau das abbilden, was durch den phonologischen Plan vorgegeben ist.

Sprechsilbenlexikon: Eine Annahme ist nun, dass Sprecher Zugriff auf einen speziellen Speicher haben, in dem die häufigsten Sprechsilben in einer Sprache gespeichert sind. Schiller/Meyer/Baayen/Levelt (1996) konnten durch Berechnungen zeigen, dass Sprecher des Englischen in etwa 85 % ihrer normalen Unterhaltungen mit nicht mehr als 500 verschiedenen Silben (500 von 12000 möglichen) führen. Für weniger häufige Sprechsilben stehen Mechanismen zur Verfügung, die die notwendigen artikulatorischen Gesten ›on the fly‹ berechnen.

Um die Existenz eines solchen Sprechsilbenlexikons zu untersuchen, ließen Levelt/Wheeldon (1994) Versuchspersonen Wörter mit hoch- und niederfrequenten Silben produzieren. Gemessen wurde die Sprechbeginnzeit. Wie vermutet, benötigten die Versuchspersonen weniger Zeit, um mit dem Sprechen zu beginnen, wenn ein Wort aus hochfrequenten Silben bestand. Entscheidend war: Die Wortfrequenz wurde in dieser Studie konstant gehalten. Würde das artikulatorische Programm für die Produktion einer Silbe jedes Mal neu berechnet werden, sollte die Silbenfrequenz keinen Einfluss auf die Performanz eines Sprechers haben. Zwei Wörter mit derselben Wortfrequenz aber unterschiedlichen Frequenzwerten für die einzelnen Silben, aus denen sie bestehen, sollten gleich schnell produziert werden. Die Befunde aus dieser Studie sprechen also für die Plausibilität der Annahme eines Sprechsilbenlexikons.

Intonatorische Enkodierung: Wenn das Ergebnis der finalen phonetischen Kodierung, also die Kette der Sprechsilben, eine präartikulatorische Repräsentation der Äußerung sein soll, dann sollte sie noch zwei Eigenschaften haben, die bisher nicht erwähnt worden sind. Sie sollte auch die Intonationskontur der Äußerung enthalten und sie sollte der motorischen Steuerung nah sein. Die Prozesse der wiederum komplizierten intonatorischen Kodierung werden hier ausgelassen, ebenso die Vorgänge der motorischen Steuerung selbst (vgl. dazu Herrmann/Grabowski 2003).

Zum Begriff

Als **phonetische Enkodierung** bezeichnet man den Prozess der motorischen Kodierung, d. h. die Programmierung artikulatorischer Gesten auf Grundlage der aktivierten Lexeminformationen und Regeln ihrer Kombinatorik.

4.3.6 | Das Zusammenspiel von syntaktischer und phonologischer Kodierung

Einige Aspekte des Zusammenspiels von syntaktischer und phonologischer Kodierung wurden bereits oben thematisiert. Im nun folgenden Abschnitt soll es vor allem um die **zeitliche Koordinierung** beider Prozesse im Rahmen von Mehrwortäußerungen gehen. Die Fragen, die sich diesbezüglich stellen, betreffen vor allem die inkrementelle Verzahnung beider Prozesse und damit einhergehend die Frage, was die Größe der Inkremente auf den jeweiligen Ebenen ist. Prinzipiell möglich erscheinen zwei

Alternativen. Entweder sind die Verarbeitungseinheiten gleich groß oder sie sind auf der Ebene der syntaktischen Enkodierung größer als auf der Ebene der phonologischen Enkodierung. Es könnte ja durchaus sein, dass zunächst mehrere Lemmata (z. B. einer Phrase) aktiviert werden, während der phonologischen Enkodierung die korrespondierenden phonologischen Codes aber sukzessive abgerufen werden. Man kann also die Frage stellen, ob und wenn ja, wie weit die syntaktische Enkodierung der phonologischen Enkodierung voraus ist.

Empirische Befunde: Direkt zu dieser Frage führte Meyer (1996) eine Objektbenennungsaufgabe durch, in der niederländische Versuchspersonen Paare von Objekten nennen sollten, die auf Bildern dargestellt waren. In einer experimentellen Bedingung sollten sie dafür die Nominalphrasen mit einer Konjunktion verbinden (Übersetzung aus dem Niederländischen: »der Pfeil und die Tasche«), in einer anderen Bedingung sollten sie einen Satz bilden (»Der Pfeil ist neben der Tasche«). In jedem Versuchsdurchgang wurde den Versuchspersonen zusätzlichen zu dem Bild, auf dem die zwei Objekte abgebildet waren, auditiv ein Distraktor präsentiert, der entweder bezüglich der Form oder Bedeutung mit dem Nomen für das erste oder zweite Objekt related oder hinsichtlich beider unrelatiert war (Bild-Wort-Interferenz, wie oben). Einen ganz ähnlichen Versuchsaufbau verwendeten auch Wagner et al. (2010) in ihrer Studie, die weiter oben schon vorgestellt wurde.

Gemessen hat Meyer (1996) die Sprechbeginnzeit. Das Ergebnis dieser Untersuchung bestand darin, dass sowohl für Sätze als auch für Phrasen längere Sprechbeginnzeiten gemessen wurden, wenn der Distraktor semantisch mit dem ersten und zweiten Nomen related war und, dass sie signifikant kürzer ausfielen, wenn der Distraktor mit dem ersten Nomen phonologisch related war. Es wurde jedoch kein Beschleunigungseffekt festgestellt, wenn der Distraktor mit dem zweiten Nomen phonologisch related war (alle Befunde natürlich immer im Vergleich zur neutralen Bedingung, also nicht-related Distraktoren). Insgesamt deuten diese Befunde darauf hin, dass vor Sprechbeginn die Lemmata für beide Nomen aktiviert wurden, aber nur für das erste auch die Lexem-Informationen. Sprecher scheinen also auf der Ebene der syntaktischen Enkodierung größere Einheiten zu verarbeiten als auf der Ebene der phonologischen Enkodierung.

Inkrementale auf der
phonologischen
Ebene

4.4 | Die Artikulation

Am Ende aller oben beschriebenen Sprachproduktionsprozesse steht die koordinierte Aktivierung von Muskelgruppen im Verlauf der Zeit, die den Luftstrom in spezifischer Weise an verschiedenen Orten zwischen Kehlkopf und Lippen eines Sprechers modulieren. Dadurch entstehen unter Beteiligung verschiedener Artikulationsorgane an verschiedenen Artikulationsorten Laute, die sich zusätzlich je nach Artikulationsweise unterscheiden (vgl. Goldrick/Blumstein 2006; McMillan/Corley 2010). Die Artikulation unterliegt bis zu einem gewissen Grad kognitiver Kontrolle, was unter anderem daran deutlich wird, dass Sprecher ihr Sprechtempo variie-

ren und die spezifische Art und Weise der Artikulation bewusst moduliert werden kann.

4.5 | Selbstkontrolle und Selbstkorrektur

Alle Produktionsschritte, die gelingen können, können auch misslingen und so finden sich in den Versprecherdaten entsprechend viele verschiedene Typen von »Fehlproduktionen«. Im Zuge einer Wegbeschreibung kann der Sprecher feststellen, dass er einen wichtigen Hinweis auf den einzuschlagenden Weg nicht genannt hat (Makroplanung), von einer Kreuzung statt von einer Gabelung gesprochen hat, *links* gemeint, aber *rechts* gesagt hat usw. Wie viele dieser Fehler vom Sprecher selbst entdeckt werden, ist schwer festzustellen, sicher nicht alle. Aber die Tatsache, dass der Sprecher sich selbst unterbricht und korrigiert, zeigt, dass das Ergebnis der Produktion auf Korrektheit geprüft wird. Die englischen Termini für diese Aktivitäten sind **self-monitoring** und **self-repair**. Wie bei allen kognitiven Aktivitäten interessiert auch bei diesen Phänomenen ihre Systematik, also welche Schritte involviert sind und ihre Dynamik, also wie sie in der Zeit ablaufen.

Experimentell
evozierte
Versprecher

Empirische Untersuchung: Selbstkontrolle und -korrektur sind nicht einfach zu beobachten. Brauchbare Evidenz stellen oftmals nur overte Sprecherreaktionen auf Versprecher dar (Korrektur oder Übersehen), und Versprecher sind, wie wir wissen, recht selten. Wird ein Versprecher nicht korrigiert, so ist das noch kein Hinweis darauf, dass er nicht bemerkt worden ist.

Dass overte Selbstkorrekturen aber nicht die einzige verlässliche Quelle für die Arbeit des Monitoring-Systems sein müssen, zeigen Motley/Camden/Baars (1982) in einer bemerkenswerten Studie, in der Versprecher experimentell evoziert wurden. Das Hervorrufen eines Versprechers (hier Phonemvertauschung) wird dadurch bewerkstelligt, dass Versuchspersonen eine Reihe von Wortpaaren leise lesen und ab und zu gebeten werden, bestimmte ebenfalls gelesene Wortpaare laut auszusprechen. Der Einfluss der zuvor gelesenen Stimuli führt häufig zu einem Versprecher. In der hier beschriebenen Studie, ergab ein erfolgreich evozierter Versprecher ein Tabu-Wort (z. B. *Tool-Kits*, *Hit-Shed*) und in einer anderen Bedingung ein neutrales Wort. Von Interesse war, wie häufig das Monitoring-System den Versprecher entdeckte und eine vollständige (Versprecher wurde nicht artikuliert) oder partielle Korrektur (*Kool-Kits*) vor der Artikulation ausgelöst werden konnte. Zusätzlich wurde bei den Versuchspersonen der Hautwiderstand gemessen, ein Indikator der die Aktivität der Schweißdrüsen in der Haut misst und die emotionale Erregtheit reflektieren kann.

Bei partiellen Tabu-Wort-Versprechern war die gemessene Reaktion größer als bei partiellen Nicht-Tabu-Wort-Versprechern, was dafür spricht, dass der vollständige Versprecher vor der Artikulation von den Versuchspersonen bemerkt wurde.

In jüngerer Zeit wurden auch Anstrengungen unternommen, Monitoring mit Hilfe des Eye-Tracking zu untersuchen. In einem Experiment von Kurucz/Gerwien (2016) sollten deutsche und chinesische Versuchsperso-

nen ein visuell präsentierte Kalenderdatum mit einem zuvor berechneten Kalenderdatum vergleichen (*Du hast einen Termin in 12 Tagen. Heute ist der 10. Juni 2016. An welchem Datum ist dein Termin?*). Das berechnete Datum durften die Versuchspersonen jedoch vor der Überprüfungsaufgabe nicht aussprechen. Es zeigte sich, dass die Blickbewegungen die Reihenfolge der einzelnen Elemente des Kalenderdatums abhängig vom kanonischen Datumsformat innerhalb der Sprechergruppen (deutsch: Tag, Monat, Jahr; chinesisch: Jahr, Monat, Tag) aber unabhängig vom visuell präsentierten Format (Tag, Monat, Jahr oder Jahr, Monat, Tag) widerspiegeln.

Was wird kontrolliert? Die Selbstkorrekturen lassen erkennen, dass Sprecher alle Arten von Fehlern, die im Output auftreten können, auch selbst entdecken können, das heißt ohne vom Gesprächspartner darauf aufmerksam gemacht zu werden.

Unterschiede in der Verzögerung der Selbstkorrektur (Korrekturlatenz) und natürlich in Inhalt und Form lassen mehrere große **Gruppen von Selbstkorrekturen** und entsprechend von Kontrollvorgängen erkennen:

- **Die inhaltliche Kontrolle:** Der Sprecher kontrolliert, ob der Sachverhalt, den er geäußert hat, tatsächlich am gegebenen Punkt in der Rede ausgewählt werden sollte. Beispiel (aus einer Bastelanweisung): [...] *legt dazwischen die/ (/ = Selbstunterbrechung) man nimmt den grünen und den gelben Würfel, legt dazwischen [...]* (Kohlmann 1996, S. 183).
- **Die Kontrolle der Ausdrucksweise:** Der Sprecher kontrolliert, ob das Gesagte eindeutig, präzise genug und mit dem Vorherigen in verstehbarer Weise kohärent war. Beispiel: [...] *dann nehme ich den andern Baukl/grünen Bauklotz* (Kohlmann 1996, S. 182).
- **Kontrolle der sprachlichen Korrektheit:** Der Sprecher hat geplant, was gesagt werden soll, und die Message ist erstellt. Bei der Korrektheitskontrolle liegt das Augenmerk auf der Leistung des Formulators, also dem Lexikonzugriff, der syntaktischen und der phonologischen Kodierung.
- **Lexikalische Kontrolle:** Beispiel (aus einer Anweisungsaufgabe): [...] *weiter nach Rot/Entschuldigung, weiter nach Schwarz* (Levelt 1983, S. 54, aus dem Niederländischen übers. von RD und JG).
- **Syntaktische Kontrolle:** Beispiel: *Ich find das nich sieht/das sieht nich so gut aus.* (Wiedenmann 1998, S. 205; Nr. 89/01/08–20).

Wird alles gleichmäßig kontrolliert? Wenn die Selbstkontrollaktivitäten mit gleicher Aufmerksamkeit auf alle **Aspekte der Äußerungsproduktion** gerichtet wären, würde man erwarten, dass alle Typen von Versprechern etwa gleich häufig entdeckt werden. Ob das so ist, ist allerdings wieder nicht sicher, da nicht bekannt ist, ob nicht-korrigierte Versprecher überhaupt entdeckt worden sind. Immerhin verteilen sich die Selbstkorrekturen tatsächlich nicht gleichmäßig. Die Verteilung der 309 Selbstkorrekturen in den Objektbeschreibungen und Bastelanweisungen von Kohlmann (1996) weisen z. B. einen Einfluss eines Faktors mit langer Wirkung und auf sehr hoher Ebene auf (Konzeptualisierung).

Neben der Ungleichverteilung hinsichtlich der Qualität der entdeckten

Versprecher werden Fehler bei der sprachlichen Planung oder Enkodierung auch in Abhängigkeit der zur Verfügung stehenden kognitiven Ressourcen unterschiedlich gut entdeckt. So steigt die Rate an Selbstkorrekturen je näher man beim Sprechen einer Phrasengrenze kommt. Levelt (1983) erklärt dies damit, dass Sprecher ihre limitierte kognitive Aufmerksamkeit (*selective attention*) während des Sprechvorgangs vor allem für die Konzeptualisierung einsetzen. Nähert man sich einer Phrasengrenze, steht kurzfristig mehr Aufmerksamkeit zur Verfügung, da der Konzeptualisierer mit seiner Arbeit bereits fertig ist, während der Formulator noch mit der Enkodierung beschäftigt ist (und Prozesse der Formulierung automatisch, also ohne Aufmerksamkeit ablaufen).

Experiment

Selbstkorrekturen in einer Beschreibungs-/Anweisungsaufgabe (Kohlmann 1996)

In einer komplexen Beschreibungs-/Anweisungsaufgabe sollten Versuchspersonen entweder a) ein Gegenüber anweisen, eine Konstruktion aus Holzbausteinen zusammenzubauen, in einer anderen Bedingung b) dem Gegenüber das fertige Produkt aus dem Gedächtnis beschreiben. Variiert wurde außerdem die Kognizierung, das heißt also der Prozess der Aufnahme der später wiederzugebenden Informationen. In einer Bedingung wurde der Versuchsperson vorab das fertige Produkt gezeigt, in einer anderen Bedingung der Aufbauvorgang. Es gab also vier Bedingungsgruppen, zwei homogene (Aufbau gesehen, Aufbau anweisen; Produkt gesehen, Produkt beschreiben) und zwei inhomogene (Aufbau gesehen, Produkt beschreiben; Produkt gesehen, Aufbau anweisen). Gezählt wurde u. a. die Zahl der Selbstkorrekturen in den Texten pro Gruppe.

Die meisten Selbstkorrekturen finden sich in den Texten der Produktkognizierungsgruppen und zwar unabhängig von der Art der sprachlichen Aufgabe: 179 Selbstkorrekturen insgesamt; das ist eine pro 3,5 Äußerungen; demgegenüber: 130 Selbstkorrekturen (= eine pro 4,9 Äußerungen) in den Aufbau-Kognizierungstexten. Die Kognizierung des Aufbaus führt also zu einem mentalen Modell des Sachverhalts, bei dessen sprachlicher Wiedergabe in Beschreibungen Anweisungen der Monitor weniger aktiv ist.

Die Darstellung sagt allerdings nichts über den Anteil nicht-korrigierter Versprecher. So ist also nicht zu entscheiden, ob der Monitor unterschiedlich genau kontrolliert hat oder immer gleich genau bei variierender Leistung des Produktionssystems. Die Ergebnisse zeigen ferner mehr Selbstkorrekturen in den Fokus- als in Topikteilen der Äußerung (Kohlmann 1996, S. 79 ff.)

Alle diese Korrekturen betreffen die Arbeit hoher Planungsprozesse, also die inhaltliche Kontrolle und die Kontrolle der Ausdrucksweise, nicht Korrekturen von Versprechern. Sie machen in Levelts Daten, die den Kohlmann-Daten von der Aufgabenstellung her ähnlich sind, nur 1 % (*D-repairs*, »D« für »different plan«) bzw. 30 % (*A-repairs*, »A« für »appropriateness«) aus (Levelt 1983, S. 51). Die übrigen 70 % sind Korrekturen von Versprechern, darunter die meisten lexikalische Versprecher (38 %), wenige syntaktische und lautliche (je 1 % bzw. 2 %).

Wie wird die Kontrolle bewerkstelligt? Es gibt zwei Gruppen von Modellen, mit denen versucht wird, Monitoring während der Sprachproduktion zu erklären (vgl. Postma 2000; Hartsuiker/Kolk 2001; Özdemir/Roelofs/Levelt 2007). Diese unterscheiden sich hauptsächlich darin, ob der Prozess von einem dem Sprachproduktionssystem internen Mechanismus oder einem externen System bewerkstelligt wird, auf wie vielen Prozessebenen die Zwischenschritte des Produktionsprozesses überwacht werden (können) und, was mit beidem zusammenhängt, ob es ein zentrales Monitoring-System oder mehrere dezentrale Instanzen gibt.

Die **perceptual-loop-Theorie** (Levelt 1983; 1989) geht von einem dem Produktionssystem verschiedenen, externen System aus. Dieses nimmt laufend das Ergebnis der Produktionsprozesse auf, analysiert es Stück für Stück und vergleicht das Analyseergebnis mit dem ursprünglichen, das heißt mit dem geplanten Produktions-Input. Bei Abweichungen zwischen dem Ergebnis der Analyse und dem ursprünglichen Input, wird unter Umständen die Produktion gestoppt und der Fehler korrigiert.

Zugriff des Monitors: Levelt (vgl. u. a. 1983; 1989) nimmt an, dass das Produktionsergebnis nur an drei Stellen im Produktionsablauf kontrolliert wird:

- Monitoring via **conceptual loop** erfüllt die Aufgabe, die Angemessenheit der aufgebauten Message in Hinblick auf die kommunikative Absicht des Sprechers zu überprüfen. Fehler, die mit der Struktur der Message oder der Bedeutung der ausgewählten Wörter zusammenhängen, können hier entdeckt werden.
- Monitoring via **internal loop** betrifft die Ebene der präartikulatorischen lautlichen Repräsentation, genauer, die silbengegliederten phonologischen Segmentketten.
- Monitoring via **external loop** richtet sich auf die overte, hörbare Rede (vgl. Wheeldon/Levelt 1995; Levelt et al. 1999).

Das System, das die Analyse durchführt, ist das Sprachverstehenssystem. Dasselbe System, das dem Menschen in der Hörerrolle ermöglicht, die Rede anderer zu verstehen, übernimmt beim Sprechen die Aufgabe des Sich-selbst-Zuhörens. Der Vergleich zwischen dem Analyseergebnis und dem intendierten Input wird essentiell vom Konzeptualisierer vorgenommen. Im Rahmen der Perceptual-Loop-Theorie ist Monitoring ein Prozess, der Aufmerksamkeit benötigt und zumindest teilweise bewusst abläuft. Diese Charakteristik macht es notwendig, dass Monitoring kognitive Ressourcen beansprucht.

Der **produktionsbasierten Monitoring-Theorie** zufolge (vgl. Laver 1973; Schlenck/Huber/Willmes 1987) findet Monitoring nicht über eine zentrale Instanz und damit auch unabhängig vom Sprachverstehenssystem statt. Vielmehr werden Zwischenergebnisse des Enkodierungsprozesses an mehreren Stellen im Produktionssystem, nämlich direkt dort, wo sie entstehen, überprüft. Monitoring erfolgt gemäß der produktionsbasierten Theorie essentiell durch Feedback zwischen den einzelnen Prozessstufen. Ein Unterschied zur Perceptual Loop Theorie besteht darin, dass Monitoring nicht nur an drei Stellen im Produktionssystem stattfindet, sondern prinzipiell überall dort, wo ein Informationstyp in einen anderen

umgewandelt wird. Die produktionsbasierte Monitoring-Theorie wird meistens von solchen Autoren angenommen, die den Gesamtprozess der Sprachproduktion als Netzwerkmodell erfassen (MacKay 1988; 1992; Berg 1988). Wie oben gesagt, wird das Produktionssystem als ein nach Ebenen gegliedertes, mehrdimensionales Netz gesehen und der Produktionsvorgang als Aktivationsfluss, ausgelöst von Knoten auf den hohen Ebenen, zu Knoten der präartikulatorischen Ebene. Die Kontrolle und ggf. Selbstkorrektur kommt dadurch zustande, dass Aktivierung nicht nur top-down fließt, sondern indem nach minimaler Verzögerung von jedem nach unten feuernenden Knoten auch wieder Aktivierung zurück nach oben fließt. Es gibt also kein externes System. Die Kontrolle findet intern statt.

5 Sprachverstehen

- 5.1 Die Schritte des auditiven Sprachverstehens im Überblick
- 5.2 Lexikalische Erkennung
- 5.3 Die Analyse der Satzstruktur
- 5.4 Verstehen der Satzbedeutung
- 5.5 Lesen
- 5.6 Sprachverstehen – ein separates Teilsystem?

Geistige Aktivitäten mit dem Ziel, sprachliche Äußerungen der Umgebung zu verstehen, finden beim Menschen von Beginn seines Lebens an statt. Das Ergebnis ist je nach Lebensalter verschieden. Wird zunächst lediglich der Unterschied zwischen sprachlichem und nicht-sprachlichem Geräusch erkannt, differenziert sich das Verstehen schon nach kurzer Zeit hinsichtlich spezifischer prosodischer Muster und einzelner Merkmale eines lautlichen Segmentes über weitere differenzierende Schritte bis zum inferentiellen Verstehen, dem sog. Interpretieren einer Äußerung beim kompetenten Sprecher/Hörer. Der herausragende Meilenstein auf dem Weg des Spracherwerbs ist offenbar der Übergang vom Umgang mit lautlichen Einheiten zur Beherrschung der Bedeutungsbeziehung zwischen lautlichem Ausdruck und einer begrifflichen Bedeutung.

Zur Aufnahme von Lauten mit Signalfunktion sind auch Tiere ausgerüstet, zur Verbindung von lautlichen Ausdrücken mit lexikalischen Konzepten hingegen nur der Mensch. Diese Verbindung wird beim Sprachverstehen immer wieder neu aktiviert und die Prozesse, die zur Worterkennung führen, sind von den höheren Prozessen der Satzerkennung so verschieden, dass man in erster Näherung sagen kann, das Sprachverstehen umfasst zwei große Schritte: **Die lexikalische Erkennung** und das **Verstehen der Äußerung**. Die beiden Schritte finden auch dann statt, wenn die Äußerung gerade nur ein Wort lang ist, wie *Danke!* und *Genau!* und andere, die sich beispielsweise in der Cockpit-Kommunikation am Anfang von Kapitel 1 finden.

Äußerungsverstehen ist im Grunde ein Konstruktionsprozess: Es ist der Aufbau einer **Sachverhaltensvorstellung** auf dem Weg über das Erkennen von Wortbedeutungen, syntaktischen Strukturhinweisen und die Heranziehung von außersprachlichen, begleitenden Wahrnehmungen, von Kontextwissen und Weltwissen.

Was Sie in diesem Augenblick – dieses Buch vor Augen – erleben, ist natürlich auch eine Modalität des Sprachverstehens, das Lesen. Auf den ersten Blick erscheint es nur insofern als einer eigenen Untersuchung wert, als die Inputsignale des Verstehensvorgangs eben Buchstaben und nicht Lautketten sind. Doch es ergeben sich auch hier einige interessante Fragen: Angeboren oder nicht? Beherrschen nur Menschen die Fähigkeit? Wie ist sie neuronal repräsentiert? Ist das Lesen ein lautloses Hören? Die visuelle Wahrnehmung ist um ein Vielfaches schneller als die auditive. Warum ist das Lesen eines Textes nicht entsprechend schneller als das Hö-

Was bedeutet es,
Sprache zu
verstehen?

ren einer Rede? Könnten Kinder das Lesen auch vor dem Hören erwerben oder nicht? Ist das Hören auch gestört, wenn das Lesen und das Schreiben gestört sind? Wie lange erinnert man sich daran, ob man eine sprachliche Information gelesen oder gehört hat? Allein über das Lesen ließe sich ein Einführungsbuch schreiben; dieses hier ist kein solches und so wird das Lesen an dieser Stelle im Rahmen des Sprachverstehens behandelt, Kapitel 5.5, wo auch weitere Literatur genannt ist.

5.1 | Die Schritte des auditiven Sprachverstehens im Überblick

Sprachverstehen ist eine kognitive Informationsverarbeitung, an deren Anfang ein Schallereignis steht und am Ende eine Sachverhaltsvorstellung, und zwar im Erfolgsfall diejenige, die der Sprecher zu verstehen geben wollte. Die beiden großen Prozesskomponenten des Sprachverstehens sind die Worterkennung und das Äußerungsverstehen.

Die lexikalische Erkennung ist nicht einfach eine Assoziation eines Schalleindrucks mit einem Stück begrifflichen Wissens, wie im Folgenden deutlich werden wird. Die akustische Form der Äußerung alleine lässt nicht erkennen, wie viele Wörter sie enthält. Sie lässt auch nicht erkennen, wie viele Silben und Segmente sie enthält und sie lässt schon gar nicht erkennen, welche Laute, Silben und Wörter im Einzelnen zu hören sind. Der Weg zur Worterkennung führt also über eine Reihe von Zwischenschritten. Zunächst muss in dem individuellen Schallereignis, produziert vor dem Hintergrund anderer Geräusche, von einer individuellen Person mit individuellem Stimmklang, in gerade dieser Lautstärke, Geschwindigkeit, Tonhöhe etc., überhaupt eine lautliche Form erkannt werden. Das setzt voraus, dass alles Individuelle, Einmalige aus dem Schallereignis herausgefiltert wird. Zum Zweiten, muss in dem scheinbar ununterbrochenen Signal Anfang und Ende von darin enthaltenen Wörtern entdeckt werden.

Das Äußerungsverstehen: Mit dem Auffinden eines Wortes im mentalen Lexikon ist aber noch nicht die Äußerung verstanden. Den Unterschied zwischen dem Zugriff auf ein Wort und dem Verstehen einer Äußerung kann man sich begreiflich machen, wenn man ihn als Bestandteil zweier verschiedener Aufgaben betrachtet. Nehmen wir an, ein Hörer soll entscheiden, ob die Lautformen *klar*, *Kabine*, *ist*, *bebig*, *nurf* u. a. Wörter des Deutschen sind. Um die Aufgabe zu lösen, muss das mentale Lexikon konsultiert werden. Ist das gehörte Item darin enthalten, so ist das Ergebnis der Bearbeitung ein »Ja«, falls nicht, ein »Nein« und die Sache ist erledigt. Ist die Aufgabe hingegen zum Beispiel zu entscheiden, ob die Äußerung *Kabine ist klar* zu einer Abbildung passt, unter der sie steht, so muss nach dem Erkennen des Wortes noch seine Stellung im Satz, die Beziehung seiner Bedeutung zu den Bedeutungen der anderen Wörter hergestellt werden und ihr Bezug zu einem Sprecher, zu den Objekten, der Zeit, dem spezifischen Ort usw., die mit der Äußerung gemeint sind.

5.2 | Lexikalische Erkennung

5.2.1 | Lauterkennung

Das kommunikative Ziel von Sprecher und Hörer ist, dass der Hörer die Absicht und die Sachverhalte erkennt, die ihm durch die Äußerung des Sprechers zu verstehen gegeben werden. Dazu muss der Hörer der lautlichen Form den Inhalt der Äußerung entnehmen und diesen dann mit seinem Wissen über Kontext, Situation und die Welt im Allgemeinen verknüpfen. Das erste Problem ist nun, dass die **Schallform** nicht die **Lautform** ist. Die Lautform muss dem akustischen Ereignis erst entnommen werden. Die abstrakte lautliche Repräsentation, die es dem Hörer ermöglicht, auf sein lexikalisches Wissen zuzugreifen, ist, wie wir im Kapitel über das Sprechen gesehen haben, eine metrisch gegliederte Kette von phonemischen Segmenten. Also muss das vom Gehör aufgenommene Schallereignis in eine Form überführt werden, die einen Bezug zu solchen Segmenten ermöglicht. Dieser Abbildungsprozess scheint auf den ersten Blick zwei weitere Schritte zu umfassen, nämlich die **Segmentierung**, also das Ermitteln, welches Stück Schall überhaupt einem Segment entspricht, und die **Kategorisierung**, also das Feststellen, um welche lautliche Einheit es sich handelt.

Als **Lauterkennung** wird der Prozess der Sprachverarbeitung bezeichnet, durch den im Sprachsignal Sprachlaute entdeckt und kategorisiert werden.

Zum Begriff

Akustische Eigenschaften des sprachlichen Schallsignals

Da ein Schallereignis natürlich zunächst einmal ein physikalisches Phänomen ist, betrifft der allererste Schritt des Sprachverstehens die Umwandlung von Schallwellen in Nervensignale. Diesen Schritt der sensorischen Wahrnehmung werden wir in unserer Darstellung allerdings überspringen. Relevant für uns ist jedoch, dass sich mit Mitteln der **akustischen Phonetik** genau die Eigenschaften des Schallsignals darstellen und analysieren lassen, die auf das Hörorgan bei der Sprachwahrnehmung treffen, also im Prinzip den Input für die sensorische Wahrnehmung abbilden. Abbildung 5.1 zeigt die akustische Repräsentation der Äußerung *Kabine ist klar*.

Ein **Spektrogramm** ist die bildliche Darstellung des Frequenzspektrums eines akustischen Signals. Gesprochene Sprache ist ein Signal, das sich aus vielen einzelnen Frequenzen zusammensetzt und zwar in einem Frequenzbereich zwischen etwa 20 Hertz (Hz) und 20.000 Hz. Mit Hilfe eines Spektrogramms lässt sich diese Zusammensetzung (im Verlauf der Zeit) sichtbar machen. Sichtbar werden hier u. a. sogenannte **Formanten**. Unter Formanten versteht man Energiekonzentrationen in bestimmten Frequenzbereichen, die während der Artikulation durch ein spezifisches Zusammenspiel der beteiligten Artikulationsorgane zustande kommen.

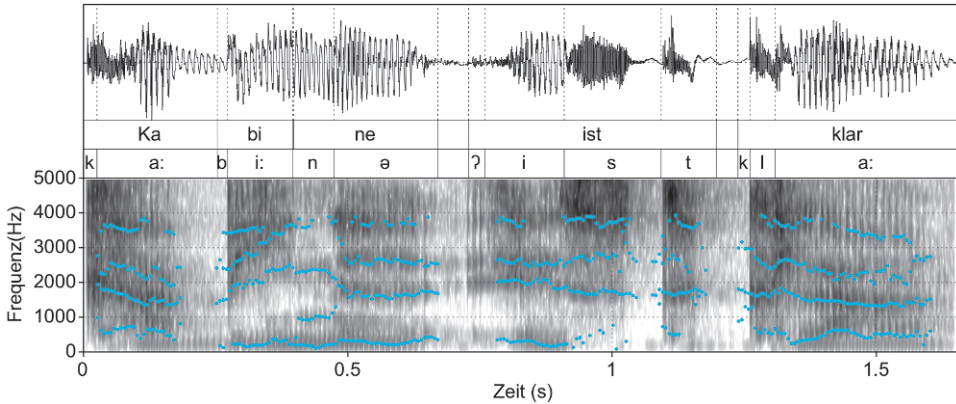


Abb. 5.1: Spektrogramm mit dargestellten Formanten für die Äußerung *Kabine ist klar* durch Sprecher A (je größer die Energiekonzentration in einem Frequenzbereich, desto dunkler die Darstellung)

Die quer laufenden Linien in Abb. 5.1 heben die Formanten im Bereich bis 5000 Hz hervor. Es wird sehr deutlich, dass sich die **Formantenfrequenzen** und das **Verhältnis der Formantenfrequenzen** untereinander für alle drei Vokale zum Beispiel im Wort *Kabine* deutlich unterscheiden. Beide Merkmale – Formantenfrequenz und Verhältnis der Formanten – gelten im Allgemeinen als wichtig für die Unterscheidung von Vokalqualitäten. **Übergangseigenschaften** von Formanten hingegen hängen von der Art und Weise ab, wie der Luftstrom bei der Artikulation durch den Verschluss und die Lösung der Artikulationsorgane verändert wird, also bei der Produktion von silbeninitialen Konsonanten bzw. Konsonantenclustern. In Abb. 5.1 beispielsweise sieht man deutlich, dass beim letzten Wort der Äußerung, also *klar*, die deutlichen Formantenfrequenzen für das enthaltene /a:/ erst relativ spät einsetzen, nämlich nach dem stimmlosen Plosiv /k/ und dem lateralen stimmhaften /l/.

Neben den Übergangseigenschaften der Formanten besteht ein weiteres Merkmal, das für die Unterscheidung von Konsonanten eine Rolle spielt in dem zeitlichen Verhältnis zwischen der Lösung eines Verschlusses (*Burst*) und der **Stimmeinsatzzeit** (*voice onset time*, VOT). In Minimalpaaren wie /pa/ und /ba/ ist die Stimmeinsatzzeit ein distinktives phonetisches Merkmal. Bei /pa/ setzt die Stimme deutlich später ein als bei /ba/.

Varianz durch Sprechermerkmale: Eine nicht unplausible Vermutung wäre nun, dass die akustischen, im Schallsignal enthaltenen Eigenschaften direkt von einem Hörer für die Erkennung von sprachlichen Einheiten genutzt werden. Problematisch ist aber, dass ein individuelles Sprecherereignis akustisch von einer Reihe von weiteren Faktoren geprägt ist. Dazu gehören Sprechermerkmale wie die Stimmhöhe, die Klangfarbe der Stimme des Sprechers aber auch akustische Besonderheiten der Sprechsituation (z. B. affektive Zustände oder Flüstern etc.). Abbildung 5.2 zeigt noch einmal den Satz »Kabine ist klar«, diesmal allerdings von einem anderen Sprecher geäußert. Allein der visuelle Vergleich zeigt, dass die Formanten in beiden Darstellungen nicht identisch sind. Individuelle und situationsbedingte Eigenschaften werden zusammengefasst oft als **indexikalische Eigenschaften** des Schallereignisses bezeichnet.

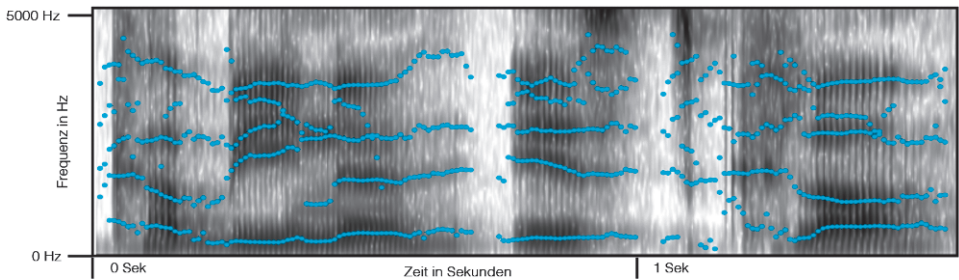


Abb. 5.2: Spektrogramm mit dargestellten Formanten von Sprecher B für die Äußerung *Kabine ist klar*

Varianz durch lautliche Umgebung: Sowohl der Segmentierung als auch der Kategorisierung steht noch ein weiterer Umstand im Wege und das ist die erhebliche **Variationsbreite**, mit der die Laute realisiert werden. Ein Vokal wird in bestimmten lautlichen Umgebungen offener, in anderen geschlossener artikuliert, unter Akzent etwas höher, unter Nebenakzent weniger hoch. Kurz gesagt, die lautlichen Eigenschaften genau eines Segments ist in großem Umfang von der Artikulation eines vorausgegangenen und eines nachfolgenden Segments abhängig. Man spricht hier auch vom Einfluss der **Koartikulation**.

Nicht-Linearität: Wie man mit Hilfe einer Sprachanalyse-Software wie PRAAT oder Speech Analyzer schnell feststellen kann, ist es in vielen Fällen nicht möglich, einzelne lautliche Segmente aus einem Sprachsignal zu isolieren. Man kann ein zuvor aufgenommenes Stück Schall zwar in einzelne kleine Stückchen zerschneiden, aber diese lassen sich nicht ohne Weiteres als ein bestimmtes lautliches Segment identifizieren. Der Grund dafür ist, dass sich die akustischen Eigenschaften benachbarter Segmente überlagern, bzw. in einander übergehen. Isoliert man beispielsweise genau das Stück Schall, das am Anfang der zweiten Silbe im Wort *Kabine* zu hören ist, also die Sprengung des Verschlusses von /b/ in der Silbe /bi:/, ist dieser Laut selbst gar nicht als Sprachlaut zu identifizieren. Erst im Zusammenhang mit dem folgenden Vokal, wird das /b/ als solches erkennbar. Mehr noch, es sind die Formantenübergänge zu einem folgenden Vokal, an denen man zwischen verschiedenen (in bestimmter Hinsicht aber ähnlichen) Verschlusslauten unterscheiden kann (z. B. /b/ und /d/).

Verarbeitung des sprachlichen Schallsignals

Kategoriale Wahrnehmung: Der Mensch ist mit einem »kognitiven Programm« ausgestattet, das dafür sorgt, dass viele Sinneseindrücke automatisch bestimmten Kategorien zugeordnet werden. Obwohl mittlerweile nachgewiesen wurde, dass der oft als »kategoriale Wahrnehmung« bezeichnete Mechanismus nicht nur während der Sprachverarbeitung zu beobachten ist, wie ursprünglich angenommen (Liberman 1957; Liberman/Harris/Hoffman/Griffith 1957), sondern auch bei anderen Aufgaben auftritt, wie der Gesichtserkennung (vgl. Beale/Keil 1995) dem Erkennen von Gesichtsausdrücken (vgl. Bimler/Kirkland 2001) oder dem Erkennen von

Tonintervallen bei der Wahrnehmung von Musik (vgl. Krumhansl 1991), hat seine ursprüngliche Entdeckung die Sprachrezeptionsforschung entscheidend geprägt und zu drei wichtigen Befunden geführt (Holt/Lotto 2010).

- **Kategorisierungseffekt:** Präsentiert man Versuchspersonen immer einen aus einer Reihe künstlich manipulierter Sprachlaute, die sich in Hinsicht auf genau eine akustische Dimension **schrittweise** unterscheiden (z. B. die schrittweise Manipulation der Stimmeinsatzzeit) und die Aufgabe besteht darin, die Stimuli als eines von zwei unterschiedlichen Phonemen zu klassifizieren (z. B. als /p/ oder /b/), dann zeigt sich, dass der Übergang der Klassifizierung von einem Laut zum anderen trotz der schrittweisen Manipulation abrupt und nicht graduell ist (s. Abb. 5.3).
- **Diskriminationseffekt:** Präsentiert man Versuchspersonen Paare von Lauten, die sich hinsichtlich einer akustischen Dimension schrittweise unterscheiden, und die Aufgabe besteht darin, anzugeben, welcher der beiden Laute mit einem dritten Laut identisch ist, zeigt sich ebenfalls ein diskontinuierliches Muster.
- **»Labelling«-Effekt:** Stehen Versuchspersonen vor der Aufgabe unterschiedliche /ba/-Laute zu unterscheiden, so ist die Performanz signifikant schlechter (auf Zufallsniveau) als wenn /ba/-Laute von /pa/-Lauten unterschieden werden sollen.

Relativierung der kategorialen Wahrnehmung: Nun ist es so, dass hinsichtlich der kategorialen Wahrnehmung aufgrund vieler Befunde nach ihrer »ursprünglichen Entdeckung« einige Einschränkungen gemacht wer-

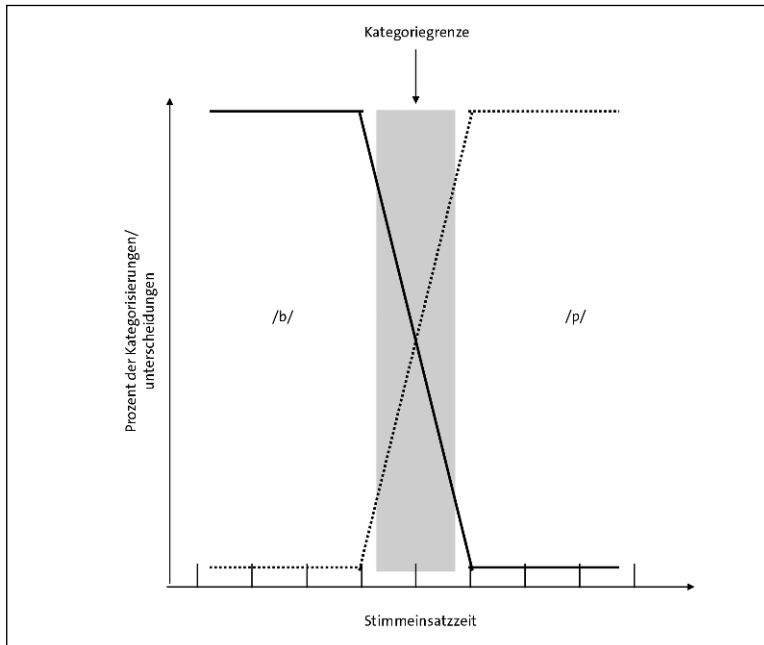


Abb. 5.3: Kategoriale Wahrnehmung: Typische Performanz von Versuchspersonen bei der schrittweisen Manipulation der Stimmeinsatzzeit (voice onset time, VOT) zweier Verschlusslaute

den mussten. Zum einen zeigte sich, dass Reaktionszeiten in einer Diskriminationsaufgabe, dann höher waren, wenn sich die zu kategorisierenden Lautsegmente in der Nähe der Kategoriegrenzen befanden (Pisoni/Tash 1974). Wäre die Wahrnehmung tatsächlich kategorial, würden man eine derartige Verzögerung nicht erwarten. Zum anderen ist der Mechanismus nicht für alle Sprachlaute gleich effektiv.

Perzeptueller Magneteffekt: Bei Vokalen zeigt sich, dass Versuchspersonen deutlich weniger gut bei Unterscheidungs- bzw. Kategorisierungsaufgaben abschneiden. Anders gesagt, der Übergang von einem Vokal zu einem anderen ist weniger kategorial als vielmehr kontinuierlich. Des Weiteren sind Hörer bei Vokalen deutlich besser beim Unterscheiden verschiedener Exemplare derselben Kategorie, zum Beispiel beim Unterscheiden verschiedener Realisierung von /i/-Lauten (Fry/Abramson/Eimas/Liberman 1962). Durch Patricia Kuhl und Kollegen (1991; 1992) wurde in diesem Zusammenhang der Begriff der **Prototypikalität eines Lautes** geprägt. Die Wissenschaftler stellten fest, dass Probanden, wenn sie verschiedene künstlich bearbeitete Varianten eines Vokals (das /i/ in »Piep«) mit Hilfe einer 7-Punkt-Skala als gutes, weniger gutes oder schlechtes Exemplar bewerten sollten, übereinstimmend solche Exemplare als gut bewerteten, die sich in Bezug auf bestimmte phonetische Merkmale ähnelten. Die Übereinstimmung der Probanden war auch bei den als schlechtes Exemplar bewerteten sehr hoch. Daraus schlossen die Wissenschaftler eben, dass es prototypische und nicht-prototypische Exemplare von Vokalen gibt. Im Anschluss an diese Feststellung konnte nachgewiesen werden, dass die Fähigkeit, zwischen verschiedenen Exemplaren von Vokalen zu unterscheiden, davon beeinflusst ist, ob die zu unterscheidenden Exemplare zu der Gruppe der prototypischen Varianten gehören oder nicht. Solche die zu den Varianten des Prototyps gehören, können schlecht voneinander unterschieden werden, solche die zu den Varianten des Nicht-Prototyps gehören, hingegen gut. Dieser Effekt wurde als **perzeptueller Magneteffekt** bezeichnet, weil ein Prototyp eines Lautes verschiedene Varianten wie ein Magnet anzieht und dabei mit Mitteln der akustischen Phonetik objektiv feststellbare Unterschiede durch das Wahrnehmungssystem neutralisiert werden. Bei der Unterscheidung verschiedener Varianten eines Nicht-Prototyps tritt der Effekt nicht auf.

Adaption: Ein interessanter Befund hinsichtlich der menschlichen Fähigkeit, sprachliche Laute zu kategorisieren ist, dass sie offenbar extrem sensibel gegenüber dem Kontext ist. Die Befunde in Ingvalson (2008) zum Beispiel zeigen deutlich, dass Kategoriegrenzen zwischen Lauten davon beeinflusst sind, wie viele verschiedene Laute in einem Experiment kategorisiert werden sollen (/r/ und /l/ gegenüber /r/ und /l/ und /w/) (vgl. Holt/Lotto 2010).

Auch der direkte phonetische Kontext beeinflusst wie ein bestimmtes Stück Schall kategorisiert wird. In einer Studie von Mann (1980) konnte gezeigt werden, dass ein und dieselbe Silbe einmal als /ga/ wahrgenommen wird, und zwar, wenn sie auf /al/ folgt, und ein anderes Mal als /da/, nämlich dann, wenn ihr /ar/ vorausgeht. Ladefoged/Broadbent (1957) konnten demonstrieren, dass dasselbe Target-Wort einmal als *bit*, einmal als *bet* wahrgenommen wird, und zwar in Abhängigkeit von den akusti-

Erkennen
von Vokalen

schen Eigenschaften eines vorausgehenden Satzes. Manipuliert wurden in diesem vorausgehenden Satz die sogenannten F1-Frequenzen.

Für viele sicherlich nicht verwunderlich ist, dass die Lautwahrnehmung auch durch visuell wahrgenommene Hinweise beeinflusst wird. Besonders eindrucksvoll wird dies durch den sogenannten **McGurk-Effekt** demonstriert.

Experiment **McGurk-Effekt**

McGurk und MacDonald (1976) präsentierten Versuchspersonen gleichzeitig akustische und visuelle Reize. Die akustischen waren einfache Silben, die visuellen waren Videosequenzen von einer Schauspielerin, die diese Silben artikulierte. Das besondere war, dass die gehörten Silben nicht zum Video passten. Sah man beispielsweise, dass die Frau die Silben /ba/-/ba/ artikulierte, hörte man die Silben /ga/-/ga/, sah man /pa/-/pa/, hörte man /ka/-/ka/. Die Aufgabe der Versuchspersonen bestand darin, einfach zu wiederholen was sie hörten. Der Hauptbefund dieser Studie ist darin zu sehen, dass fast alle Versuchsteilnehmer den Mismatch zwischen visuellem und auditivem Input quasi durch einen Kompromiss auflösten. 98 % aller erwachsenen Teilnehmer produzierten die Silben /da/-/da/ als Reaktion auf den ga-Sehen/ba-Hören-Stimulus.

Wie Hörer Sprachlaute kategorisieren bzw. zwischen verschiedenen Sprachlauten unterscheiden, basiert offenbar nicht auf fixierten, absoluten Kategoriegrenzen. Vielmehr werden diese an die aktuellen Anforderungen in der Umgebung angepasst.

Zum Begriff

Unter **kategorialer Sprachwahrnehmung** versteht man eine spezifische Reaktion des Wahrnehmungssystems auf sensorischen lautlichen Input, die dadurch gekennzeichnet ist, dass wahrgenommene Unterschiede innerhalb einer Kategorie komprimiert und/oder wahrgenommene Unterschiede zwischen Kategorien separiert werden (vgl. Harnard 2003).

Theorien der Lautwahrnehmung

Alle Theorien der Lautwahrnehmung müssen der mangelnden Invarianz im Sprachsignal (**Invarianzproblem**) sowie dem Phänomen der **Koartikulation** auf der einen Seite und der Fähigkeit des Sprachverstehenssystems, einzelne Segmente dennoch zu kategorisieren, auf der anderen Seite Rechnung tragen. Dabei lassen sich die verschiedenen Theorien dadurch unterscheiden, welche Informationen für die Kategorisierung von Lauten als entscheidend angesehen werden. Einige Theorien nehmen an, dass alle Hinweise, die dem Lautstrom entnommen werden können, verwendet werden. Andere gehen davon aus, dass bestimmte Informationen als irrelevant betrachtet werden und daher unberücksichtigt bleiben. In Hinblick

auf Theorien der zweiten Gruppe sprechen McMurray/Jongman (2011) von Theorien, die in der einen oder anderen Form **Kompensation** bzw. **Normalisierung** annehmen. Darunter versteht man alle Mechanismen, die dazu dienen, Varianz im Sprachsignal zu neutralisieren. Wir werden im Folgenden einige Theorien der Lautwahrnehmung vorstellen und zur besseren Orientierung angeben, ob sie Kompensation/Normalisierung annehmen oder nicht.

Sprachproduktionsbasierte Theorien: Liberman/Mattingly (1985) und Liberman/Whalen (2000) schlugen auf der Basis unzähliger Untersuchungen zur kategorialen Lautwahrnehmung im Rahmen der sogenannten **Motor-Theorie der Lautwahrnehmung** vor, dass der Grund dafür, dass Menschen Laute in Kategorien wahrnehmen, damit zusammenhängt wie Menschen Laute produzieren. Deutschsprachige z. B. können eben nur entweder ein /b/ oder ein /p/ produzieren, aber eben nichts, was dazwischenliegt. Die Annahme der Forschergruppe um Alvin Liberman war, dass die Kategorisierung eines Lautes in der Erkennung des vom Sprecher *intendierten* Motorikprogramms liegt, das bei der Sprachproduktion seine Artikulation auslöst, also der **artikulatorische Gesten**. Konsequenterweise wurde dann auch angenommen, dass dort wo die Sprachproduktion kategorial ist, auch die Sprachrezeption kategorial sein muss, also bei Konsonanten und dass, wo die Sprachproduktion kontinuierlich ist, die Rezeption kontinuierlich ist, also bei Vokalen. In der Motor-Theorie der Lautwahrnehmung wird Kompensation angenommen, denn eine intendierte artikulatorische Geste ist für verschiedene Laute bzw. Silben invariant, so damals die Annahme. Varianz im Sprachsignal komme erst durch die Ausführung der artikulatorischen Gesten zustande. Ein Kompensationsmechanismus ist der Theorie somit inhärent.

Sprachproduktionsbasierte Theorien heute

An der Behauptung, das motorische System der Artikulation sei an der Sprachwahrnehmung beteiligt, wird weiterhin festgehalten (vgl. Galantucci/Fowler/Turvey 2006). Darin können sich die Proponenten entsprechender Theorien auch teilweise durch die Ergebnisse einer Messung mit transkranieller Magnetstimulation (TMS) von Fadiga/Craigheo/Buccino/Rizzolatti (2003) bestätigt sehen. Bei den Versuchspersonen wurden Aktivierungen motorevozierter Potentiale gemessen, wie sie bei der motorischen Steuerung der Zunge beim Artikulieren der gehörten Sprachbeispiele stattfinden.

Auf den ersten Blick problematisch für produktionsbasierte Theorien scheint allerdings, dass bereits bei Neugeborenen Hinweise auf die Fähigkeit zur Lautunterscheidung festgestellt wurden und das obwohl sie zu diesem Zeitpunkt noch nicht in der Lage sind, die Laute, die sie unterscheiden können, selbst zu produzieren. Die Fähigkeit zur Lautunterscheidung ist demnach nicht zwangsläufig gleichzusetzen mit der Fähigkeit zur Lauterkennung.

Zur Vertiefung

In **exemplarbasierten Theorien** (vgl. Johnson 1997; Goldinger 1998; Pierrehumbert 2001; Hawkins 2003) wird davon ausgegangen, dass Hörer

viele verschiedene Exemplare sprachlicher Einheiten speichern und dementsprechend aktuellen Input mit den gespeicherten Einheiten vergleichen. Jedes Exemplar ist dabei ein ganzer »Chunk« von Informationen, der kontextabhängige Varianz und die Bedingungen, unter denen diese Varianz zustande kommt, beinhaltet. Kategorien, die für die Sprachverarbeitung genutzt werden, kann man sich als Sammlung verschiedener Exemplare vorstellen, die alle dasselbe Kategorie-Label tragen. Innerhalb einer Kategorie teilen alle Exemplare Ähnlichkeiten hinsichtlich einer bestimmten Dimension.

Offensichtlich verzichten derartige Theorien auf Kompensation für die mangelnde Invarianz des Schallsignals und auf Kompensation für die Koartikulation. Ganz im Gegenteil, jegliche Information des Inputs wird benötigt, um den Vergleich mit den gespeicherten Exemplaren erfolgreich gewährleisten zu können bzw. um die Kategorien durch das Hinzufügen neuer Exemplare robuster zu machen.

Auch exemplarbasierte Theorien nehmen einen Zusammenhang mit der Sprachproduktion an. So konnte in einigen Studien nachgewiesen werden, dass bestimmte phonetische Details in kürzlich wahrgenommener Sprache auch in der Sprachproduktion des Hörers erscheinen (vgl. Goldinger 1998; 2000).

Theorien der Merkmalsdetektion (vgl. Stevens 1972) widersprechen der postulierten mangelnden Invarianz des Sprachsignals und gehen davon aus, dass die Lautkategorisierung primär durch die Erkennung (Detektion) invarianter akustischer Merkmale oder Gruppen von Merkmalen geschieht, zum Beispiel durch spezifische akustische Eigenschaften von Bursts (Öffnung eines Verschlusses) und Formantenübergängen bzw. deren Zusammenspiel.

Klarerweise muss eine Theorie der Merkmalsdetektion nachweisen, dass invariante Merkmale wirklich für alle in einer Sprache auftretenden Laute existieren und dies ist bisher noch nicht vollständig gelungen. Kompensation muss in Theorien der Merkmalsdetektion zweifellos angenommen werden, denn irrelevante Eigenschaften wie Sprechtempo, Stimmhöhe, etc. überlagern charakteristische phonetische Merkmale von Lauten.

Experiment **Experiment zum Nachweis invarianter phonetischer Merkmale**

Stevens/Blumstein (1978) erzeugten für ein Experiment auf künstliche Weise Konsonant-Vokal-Silben. Die Konsonanten (hier die Verschlusslaute /b/, /d/, /g/) wurden dabei entsprechend den Ergebnissen zeitlich hochauflösender Spektralanalysen erzeugt, mit deren Hilfe festgestellt werden konnte, dass sich, je nach Artikulationsort, spezifische Energiekonzentrationen innerhalb der ersten 26 ms (!) nach Öffnung des Verschlusses zeigen. Für Bilabiale (/b/) waren das Energiekonzentrationen in einem niedrigen Frequenzbereich, für Velare (/g/) in einem mittleren und für Alveolare (/d/) in einem hohen. Die Verschlusslaute, die, wie gesagt, entsprechend dieser Analysen erzeugt wurden, wurden dann mit verschiedenen Vokalen kombiniert, die ebenfalls künstlich generiert wurden. Sollte es sich bei den festgestellten akustischen Eigenschaften der drei verschiedenen Konsonanten um invariante Merkmale handeln, so die

Argumentation der Autoren, dann sollten Versuchspersonen die Konsonanten unabhängig von den spezifischen Vokalen erkennen, mit denen sie kombiniert wurden. Und genau das zeigten die Ergebnisse einer Kategorisierungsaufgabe.

5.2.2 | Die Worterkennung

Das Erkennen verschiedener Laute im Schallstrom bedeutet aber noch lange nicht, dass sich hierdurch automatisch das Erkennen von Wörtern ergibt. Welche Laute gehören zu einem Wort, beziehungsweise welche Laute markieren Grenzen zwischen verschiedenen Wörtern? Wenn dieselben Laute unterschiedliche Wörter ergeben, zum Beispiel (*er*) *macht* und (*die*) *Macht*, welches muss für das Verstehen einer Äußerung ausgewählt werden? Fragen wie diese betreffen, die Aufgaben des Worterkennungsprozesses. Darüber hinaus stellen sich aber natürlich auch Fragen über den genauen (zeitlichen) Ablauf des Vorgangs, zum Beispiel: Wird er gestartet, sobald der erste Laut erkannt ist oder wird auf mehr Information gewartet? Wird ausschließlich auf Informationen zurückgegriffen, die durch die Lauterkennung zur Verfügung gestellt wird, oder macht das Worterkennungssystem auch von zusätzlichem linguistischem und möglicherweise sogar nicht-linguistischem Wissen Gebrauch, und wenn ja zu welchem Zeitpunkt? Bevor wir uns diesen Fragen genauer zuwenden, betrachten wir einige empirische Befunde im Bereich der Worterkennung.

Methoden und Befunde

Gating-Experiment: Auf den ersten Blick würde man vielleicht vermuten, dass man ein Wort natürlich immer erst vollständig hören muss, bevor man entscheiden kann, welches es ist. Dass dies nicht so ist, verdeutlichen Befunde aus sogenannten Gating-Experimenten (Grosjean 1980). In solchen Experimenten kann der Gesamtprozess der Worterkennung gewissermaßen künstlich verlangsamt werden und zwar dadurch, dass die Schallkette einer gesprochenen Äußerung in kleinen, zeitlich separiert aufeinander folgenden Portionen von z. B. 30 ms präsentiert wird. Eine Versuchsperson würde so zum Beispiel zunächst nur die ersten 30 ms eines Wortes hören, dann die ersten 60 ms, dann die ersten 90 ms usw. Fordert man sie dabei auf, anzuzeigen, sobald sie anhand der gehörten ›Portion‹ Schall ein Wort eindeutig erkennt, so gibt die dem Schallstück entsprechende Lautkette Aufschluss darüber, wie viel Lautinformation – gemessen vom Anfang eines Wortes – für die Identifizierung eines Wortes im mentalen Lexikon benötigt wird.

Das Ergebnis derartiger Gating-Experimente – so heißt die stückweise Präsentation einer Schallfolge oder auch einer Buchstabenfolge vom Wortanfang her – ist überraschend. Wird zum Beispiel das Wort *ziemlich* stückweise präsentiert, wird es erkannt, bevor es ganz gehört ist. Sobald das Schallstück, das das /l/ darstellt, verarbeitet ist, zeigen die Versuchspersonen an, dass sie ein Wort erkannt haben und nennen fast ausnahmslos das Wort *ziemlich*. Das Wort wird also erkannt, bevor es ganz gehört ist.

Stückchenweise
Präsentation
von Wörtern

Wichtig an diesem Ergebnis ist neben dem Erkennen des Wortes auch, dass die Stelle erkannt ist, an der das nächste Wort beginnen muss, nämlich nach *-lich*.

Wörter werden
häufig erkannt,
bevor ihr Ende
gehört wird

Für die vorzeitige Worterkennung gibt es zwei Erklärungen, eine sprachliche und eine kognitive. Die sprachliche ergibt sich aus einer Eigenschaft vieler Wörter, nämlich der, ein redundantes lautliches Ende zu haben. Redundant ist ein Zeichen, das im gegebenen Zusammenhang keinen zusätzlichen Informationsgehalt hat. Der Informationsgehalt des Lexems ist, das Lemma eindeutig zu bezeichnen. Und diese Information ist bei vielen Wörtern schon durch ein Stück aus der Anfangskette des Lexems vermittelt; ein Stück der Endkette ist redundant. Das letzte lautliche Segment der nicht-redundanten Anfangskette markiert den Punkt, an dem das Wort eindeutig von allen anderen Wörtern der Sprache verschieden ist. Dieser Punkt im Wort ist der **Diskriminationspunkt** des Wortes (*uniqueness point*). Er ist rein linguistisch bestimmt. Bei vielen Wörtern liegt er allerdings erst auf dem letzten lautlichen Segment. Ein Lexem eines solchen Wortes ist nicht redundant, wie z. B. das von *Kabine*. Erst der finale Schwa-Laut unterscheidet das Wort von *Kabinett*.

Beobachtungen bei der Durchführung von Gating-Experimenten zeigen, dass die Stelle im Wort, an der Versuchspersonen es richtig erkennen, nicht immer mit dem Diskriminationspunkt zusammenfällt. Der **Worterkennungspunkt** (*recognition point*) liegt in Experimenten mal vor und mal hinter dem Diskriminationspunkt. Die Stelle in einem Gating, an der die Versuchsperson eine Entscheidung über das Zielwort trifft, sei sie korrekt oder nicht, heißt **Isolationspunkt** (*isolation point*). Ein interessanter zusätzlicher Befund ist, dass der Isolationspunkt sowohl vom Kontext, in dem ein Wort auftritt als auch von der Wortfrequenz abhängt (vgl. Grosjean 1980). Methodische Details des Gating-Paradigmas und ihrer Durchführung behandeln Marslen-Wilson/Tyler (1980).

Schattensprechen (Shadowing): Geübte Personen können Wörter, die sie über Kopfhörer hören, so schnell nachsprechen, dass sie bis zu 250 ms an den Input herankommen (Marslen-Wilson 1973). Das zeigt, dass das Erkennen des Wortes, das nachgesprochen werden soll und die Zeit, die für die Aktivierung der für das Nachsprechen benötigten artikulatorischen Gesten zusammen unter Umständen eben nicht länger als 250 ms dauert. Gerade dann, wenn die nachzusprechenden Wörter selbst länger als 250 ms sind, verdeutlicht dieser Befund erneut, dass Wörter für ihre Erkennung nicht vollständig, also bis zu ihrem Ende gehört werden müssen.

Ein weiterer interessanter Befund ergibt sich, wenn geübte Schattensprecher während einer Nachsprechaufgabe auf ein Wort stoßen, dass anders endet, als sie erwarten. In solchen Situationen lässt sich beobachten, dass Hörer das Wort reproduzieren, das sie erwarten und nicht das Wort, das sie gehört haben.

Phonemrestauration: Hörer sind offenbar in der Lage, lautliche Segmente, die sie nicht wahrgenommen haben oder nicht kategorisieren können, zu ergänzen, und zwar so, dass aus der unvollständigen Repräsentation und der Ergänzung ein Wort entsteht, das zum Kontext passt. Dies wurde von Richard Warren und Kollegen (vgl. 1970; 1974; 1984) im Rahmen verschiedener Experimente demonstriert (s. Kasten).

Auch neurophysiologische Daten liegen in Bezug auf das Phänomen der Phonemrestauration vor. Eine ERP-Studie von Sivonen et al. (2006) verdeutlicht unter anderem, dass das Gehirn stärker auf unerwartete Wörter am Ende eines Satzes reagiert als auf hocharwartbare aber »verwischte« Wörter, also solche, bei denen der Onset durch ein Störgeräusch überlagert ist.

Experiment zur Phonemrestauration (Warren 1984)

Experiment

Versuchspersonen hörten ein aufgenommenes Stück Schall, welches, wenn vollständig wahrgenommen, ein Wort ergibt. Das initiale Phonem wurde allerdings entfernt und durch ein Husten ersetzt. Das so entstandene Stück Schall war für sich genommen nicht mehr als Wort erkennbar. Dasselbe auf diese Weise manipulierte Signal wurde dann in verschiedenen Kontexten Versuchspersonen präsentiert: *It was found that the #eel was on the__*. Das #-Symbol steht hier für den manipulierten Onset des Wortes und das »__« steht hier als Platzhalter für verschiedene Wörter, nämlich (1) »axle«, (2) »shoe«, (3) »orange« oder (4) »table«. Es zeigte sich, dass die Versuchspersonen das unvollständige Signal (#eel), je nach Kontext unterschiedlich ergänzten. Im Kontext von (1) als *wheel*, bei (2) als *heel*, bei (3) als *peel* und bei (4) als *meal*. Dieser Effekt wird als Phonemrestauration bezeichnet. Bemerkenswert ist in diesem Zusammenhang, dass die Information, die für die Phonemrestauration benötigt wird, erst nach dem unvollständigen Element zur Verfügung steht, was man so interpretieren muss, dass die Worterkennung offenbar nicht unmittelbar geschehen muss, sondern auch »nachträglich« erfolgen kann.

Eye-Tracking zeigt Wettbewerb: Auch mit Hilfe des Eye-Tracking lässt sich die Worterkennung untersuchen. Im sogenannten Visual-World-Paradigma sehen Versuchspersonen visuelle Stimuli, zum Beispiel bildliche Darstellungen von Objekten, und hören gleichzeitig einen auditiven Stimulus, in dem direkt auf die Objekte Bezug genommen wird. Als relativ robuster Effekt gilt, dass Versuchspersonen dazu tendieren, ihren Blick auf Dinge zu richten, mit denen sie sich gerade beschäftigen. Aus diesem Grund kann durch die Messung der Augenbewegungen ermittelt werden **wann** relativ zum Onset des auditiven Stimulus visuelle Aufmerksamkeit auf mit dem Audio-Stimulus auf die eine oder andere Art und Weise verknüpften visuellen Stimulus gerichtet wird.

In einer einflussreichen Studie von Allopenna/Magnuson/Tanenhaus (1998) sahen Versuchspersonen zum Beispiel visuelle Stimuli mit insgesamt vier Objekten, zum Beispiel einem Becher (engl. *beaker*), einem Käfer (*beetle*), einem Lautsprecher (*speaker*) und einer Kutsche (*carriage*). Gleichzeitig hörten die Versuchspersonen (englische Muttersprachler) die Aufforderung »Pick up the beaker. Now put it ...«. Die Namen von drei der vier Objekte stehen in einem bestimmten Zusammenhang: *Beaker* und *beetle* beginnen mit denselben Lauten. *Beaker* und *speaker* reimen sich. Es wurden die Augenbewegungen der Versuchspersonen gemessen, während sie die Aufforderung hörten. Es zeigte sich dabei folgendes Muster: Am

Anfang der Präsentation des Wortes *beaker* stiegen sowohl die Blickanteile, die auf dem Becher (*beaker*) gemessen wurden, als auch die auf dem Käfer (*beetle*) gemessenen. Sobald am Ende von *beaker* die phonologische Information aber nicht mehr mit *beetle* zusammenpasste, sanken die Blickanteile auf dem Käfer, wohingegen sie auf Becher weiter anstiegen. Gleichzeitig wurde allerdings ein Anstieg der Blickanteile auf dem Lautsprecher (*speaker*) festgestellt. Dieser Effekt war allerdings deutlich schwächer ausgeprägt. Zusammengefasst zeigen diese Ergebnisse, dass sowohl Onset-Kompetitoren als auch Reimkompetitoren während der auditiven Worterkennung eine Rolle spielen und dass erstere früher als letztere in Wettbewerb treten, dieser Wettbewerb aber auch früher beendet wird.

Fazit: Natürlich gibt es noch viele weitere Befunde zur Worterkennung. Aus den hier dargestellten, lässt sich allerdings schon einmal Folgendes festhalten:

- Die Worterkennung beginnt, sobald der erste Laut identifiziert wurde.
- Während des Worterkennungsprozesses stehen verschiedene Kandidaten miteinander im Wettbewerb.
- Sobald Information zur Verfügung steht, die mit einem Kandidaten inkompatibel ist, scheidet dieser Kandidat aus.
- Wörter können erkannt werden, noch bevor sie vollständig gehört sind.
- Die Worterkennung ist kein reiner Bottom-up-Prozess; Top-down-Prozesse helfen dabei (a) nicht identifizierbare Laute zu ergänzen und (b) die Wahrscheinlichkeit eines Kandidaten aus dem Kontext abzuleiten.

Prozesse der Worterkennung

Initialer Kontakt,
Selektion,
Integration

Grundsätzlich gehen viele Forscher heute davon aus, dass die Worterkennung ein Prozess ist, bei dem sprachlicher Input mit mehreren potenziellen lexikalischen Kandidaten verglichen wird und dass am Ende des Prozesses ein Kandidat ausgewählt wird, der einerseits zum sprachlichen Input und andererseits zu bereits verarbeitetem sprachlichen Material passt. Dabei wird meistens zwischen drei Prozessen unterschieden (vgl. Dahan/Magnuson 2006), nämlich zwischen dem ›initialen Kontakt mit dem Lexikon‹, der ›Selektion‹ sowie der ›Integration‹.

Unter **initialem Kontakt** wird der Prozess erfasst, bei dem durch den sprachlichen Input mehrere alternative lexikalische Repräsentationen aktiviert werden. Anders gesagt, hier wird die Lautform aller Wörter aktiviert, die mit dem Input kompatibel sind. Unter **Selektion** versteht man den Prozess, bei dem ermittelt wird, welche der aktivierten lexikalischen Repräsentationen (Wortkandidaten) am besten zum Input passen. Mit **Integration** meint man den Evaluationsprozess, bei dem unter den aktivierten Kandidaten derjenige ausgesucht wird, der unter den gegebenen Umständen am wahrscheinlichsten ist, d. h. sich mit anderen lexikalischen Einheiten am besten zu größeren Strukturen (Phrasen/Sätzen) verknüpfen lässt.

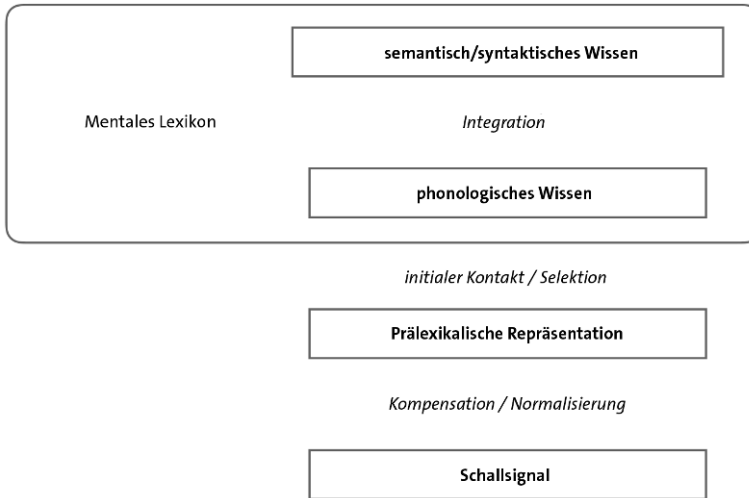


Abb. 5.4: Darstellung der ablaufenden Prozesse bei der Worterkennung

Initialer Kontakt

Prälexikalische Repräsentation: In der Mehrheit der Modelle zur Worterkennung wird davon ausgegangen, dass die im mentalen Lexikon gespeicherte Form eines bestimmten Wortes eine abstrakte lautliche Repräsentation ist, d. h. nicht jede Möglichkeit der lautlichen Realisierung eines Wortes ist im mentalen Lexikon gespeichert (vgl. McQueen 2007). Ausnahmen bilden hier natürlich die bereits oben angesprochenen exemplarbasierten Theorien. Hinsichtlich des initialen Kontaktes stellt sich dann natürlich die Frage, wie mit dem bereits thematisierten Invarianzproblem bzw. dem Problem der Nicht-Linearität des Sprachsignals umgegangen wird. Die allgemeine Annahme in diesem Zusammenhang ist, dass zunächst eine sogenannte prälexikalische Repräsentation generiert wird. Diese sehr kurzlebige und dem Bewusstsein nicht zugängliche Repräsentation sorgt für das, was oben mit den Begriffen ›Kompensation‹ bzw. ›Normalisierung‹ bei der Lauterkennung bezeichnet wurde.

Mit **prälexikalische Repräsentation** bezeichnet man eine Informationsebene, die zwischen dem akustischen Signal und der abstrakten lautlichen Repräsentation von Wörtern im mentalen Lexikon vermittelt.

Zum Begriff

Welche **Informationseinheiten auf prälexikalischer Ebene** allerdings anzunehmen sind, darüber herrscht keine Einigkeit. Auf den ersten Blick erscheint das **Phonem** hier wohl als der am besten geeignete Kandidat, denn es sind ja bekanntlich Phoneme, die linguistisch gesehen die kleinsten bedeutungsunterscheidenden Einheiten sind. Deutlich wird dies zum Beispiel durch ein Experiment von Marslen-Wilson und Zwitserlood (1989), in dem Probanden zuerst über Kopfhörer einen auditiven Stimulus hörten und direkt im Anschluss entscheiden sollten, ob eine kurze Zeit

später visuell präsentierte Buchstabenfolge ein Wort ist oder nicht (cross-modales Priming). Es zeigte sich, dass das visuell präsentierte Wort *bij* (niederländisch für Biene) schneller als Wort erkannt wird, wenn zuvor *honing* (niederländisch für Honig) gehört wird als wenn vorher *woning* (Wohnung) gehört wird. Beide Prime-Wörter unterscheiden sich nur durch den initialen Laut. Aus diesen Ergebnissen wurde geschlossen, dass beim Hören von *woning* der Zugriff auf *honing* blockiert ist, der Mismatch zwischen dem initialen Phonem des Zielwortes (/h/) und des Testwortes (/w/) also entscheidend für die Beobachtung ist. In späteren Experimenten konnten dann weitere Verfeinerung zur Bedeutung des Phonems bei der Worterkennung gemacht werden, zum Beispiel, dass ein Mismatch, der später in einem Wort auftritt weniger gravierende Folgen hat (vgl. Allopenna et al. 1998).

Woraus besteht
die prälexikalische
Repräsentation?

Suprasegmentale Informationen: Gegen die Annahme, dass das Phonem die (einzige) Informationseinheit auf prälexikalischer Ebene ist, spricht allerdings die Beobachtung, dass die Worterkennung von akustischen Merkmalen beeinflusst wird, die nicht phonemunterscheidend sind, so zum Beispiel von der Vokaldehnung (im Englischen). Salverda/Dahan/McQueen (2003) zeigten Versuchspersonen visuelle Stimuli, auf denen neben zwei Distraktoren solche Objekte waren, deren Name aus einer gleichen initialen Phonemkette bestanden, z. B. *Ham* (Schinken) und *Hamster* (Hamster). Die Versuchspersonen sahen denselben visuellen Stimulus in zwei Bedingungen. In der einen wurde der auditive Stimulus so gestaltet, dass die erste Silbe aus *Hamster* durch die erste und einzige Silbe aus *Ham* (Schinken) ersetzt wurde. In der anderen Bedingung blieb *Hamster* unverändert. Es zeigte sich, dass Versuchspersonen, wenn sie das manipulierte Wort *Hamster* hörten, stärker auf die Abbildung eines Schinkens schauten, was wohlgemerkt nicht das Zielwort war, als wenn sie das nicht manipulierte Wort *Hamster* hörten. Wichtig ist hier, dass sich die Vokaldauer des /a/ beider Tokens von *Ham* unterscheidet (das /a/ ist länger in *Ham* als in *Hamster*). Zu ganz ähnlichen Befunden kamen auch Davis/Marslen-Wilson/Gaskell (2002), die die Vorhersagbarkeit von Wörtern wie *Cap* und *Captain* mit Hilfe einer Gating-Aufgabe untersuchten. Beide Studien verdeutlichen, dass eben auch suprasegmentale Informationen Teil der prälexikalischen Information sein müssen.

Spektrale Informationen und distinktive Merkmale: An dieses Argument anknüpfend, nehmen einige Autoren an, dass die prälexikalischen Informationseinheiten essenziell aus spektralen Informationen bestehen, also zum Beispiel Informationen über Formanten, die zeitliche Dauer von Formanten, Formantenübergänge, Formantenkonfigurationen, die akustischen Qualitäten von Verschlussöffnungen usw. (Stevens 1972). Andere Autoren nehmen an, dass spektrale Informationen zu Mustern zusammengefasst werden, die wiederum die Erkennung distinktiver Merkmale wie Artikulationsort, Stimmeinsatzzeit oder Artikulationsmodus erlauben. Der initiale Kontakt zum mentalen Lexikon erfolgt aus diesen beiden Perspektiven also mittels Informationen, aus denen sich linguistisch gesehen Phoneme erst zusammensetzen.

Auch die **Silbe** wurde als Informationseinheit der prälexikalischen Repräsentation vorgeschlagen (Mehler/Dommergues/Frauenfelder/Segui

1981; Cutler/Mehler/Norris/Segui 1986). In diesem Zusammenhang sind zwei Tatsachen von großem Interesse. Die eine ist, dass eine beliebige Lautfolge als Ganze im Input schneller erkannt wird, wenn sie eine Silbe der Sprache bildet, als wenn sie nur einen Teil einer Silbe bildet. In einer **Laut-Monitoring**-Aufgabe (Achte darauf, wann in der Äußerung, die du hörst, die Lautfolge ›xy‹ auftaucht), haben französische Versuchspersonen die Lautfolge [pa] schneller in dem Wort *palace* als in *palmier* entdeckt; ebenso [ka] in *carotte* schneller als in *carton*. Der Befund spricht dafür, dass die Silbengliederung die Segmentierung des Lautstroms beim Verstehen mitbestimmt (Mehler et al. 1981, S. 300).

Die zweite Tatsache ist, dass die Erkennung von Morphemen im Schallstrom von der prosodischen Gliederung beeinflusst wird. Das einsilbige englische Wort *mint* z. B. wird als Bestandteil eines Non-Worts korrekter erkannt, wenn es betont und von einer unbetonten Silbe gefolgt ist, (Beispiel /MINTɛf/; Großschreibung bedeutet ›akzentuiert‹), als wenn es von einer stark betonten Silbe gefolgt ist (/MINTEIF/). Die Annahme ist, dass bei Wahrnehmung einer starken Silbe das System zu einer Silbentrennung neigt, also /MINTEIF/; dadurch ist das Zielwort nicht mehr direkt zu erkennen. Infolge der Trennung kann das Zielwort *mint* nur durch nachträgliche Zusammenführung der ersten Silbe, *min* mit dem Anlaut der zweiten, *t*, erkannt werden. Das geschieht weniger bei der Wahrnehmung einer schwachen Silbe, /MINTɛf/ (vgl. Cutler/Norris 1988).

Die beiden Befunde sprechen nach Auffassung von Cutler und Mitarbeitern für eine **metrisch** gesteuerte Segmentierungsstrategie, durch die der ankommende Schallstrom schon in einer frühen Phase der lexikalischen Erkennung, unter Umständen schon vor der Phonemerkennung, in Einheiten segmentiert wird, die für die weiteren Schritte der Worterkennung ausgenutzt werden (vgl. Cutler/Norris 1988; McQueen/Norris/Cutler 1994).

Inhalt der prälexikalischen Repräsentation: Ohne dass an dieser Stelle der einen oder anderen Konzeptionen der Vorzug gegeben werden soll, können wir festhalten, dass die prälexikalische Repräsentation sowohl segmentale als auch suprasegmentale Informationen beinhaltet, die beide dem Schallereignis entnommen werden. Die **segmentale prälexikalische Information** ist die Information über die im Schallstrom enthaltenen Einzellaute. Zur **suprasegmentalen prälexikalischen Information** gehört all das, was über die im Schallsignal enthaltenen Einzellaute hinausgeht. Das ist zum Beispiel Information darüber, welche Einzellaute welche Position in einer Silbe besetzen, oder Information über die lexikalisch motivierte Dehnung von Vokalen (*Bann* vs. *Bahn*); in Tonsprachen wie dem Chinesischen eben Information über den Ton (*shí shì shí* – nur das richtige Erkennen der Töne führt zum Erkennen der Wörter *10 ist 10*) sowie die Position eines Wortes in der Intonationsstruktur der Äußerung.

Selektion

Unter Selektion versteht man die Aktivierung von Wörtern im mentalen Lexikon, die zunächst auf Grundlage ihrer lautlichen Gestalt mit der Input-Repräsentation kompatibel sind, egal wie diese theoretisch konzipiert ist, als prälexikalische Repräsentation oder als spektrografische Repräsentation

des akustischen Signals. Die Aktivierung solcher Kandidaten beginnt, sobald die ersten Einheiten (lautliche Segmente oder ganze Silben) identifiziert wurden. Wenn also beispielsweise die Silbe /bi/ erkannt wurde, werden alle Wörter, die zu diesem Input passen, aktiviert, also unter anderem *Biber, Biene, Bibel, Bison, Biologie, Bielefeld* usw. Sobald nachfolgender Input mit einem Kandidaten inkompatibel ist, scheidet dieser aus dem Wettbewerbsprozess aus. Eine große Herausforderung stellt nun allerdings die Tatsache dar, dass der Schallstrom quasi kontinuierlich ist und sich nicht unmittelbar bestimmen lässt, ob z. B. /bi/ tatsächlich den Anfang eines Wortes markiert oder möglicherweise eine Silbe in der Mitte eines Wortes darstellt, wie in *abbiegen*. Es stellt sich also ein Segmentierungsproblem.

Erkennen von Wortgrenzen

Segmentierung: In der geschriebenen Sprache (zumindest im Deutschen und vielen anderen) werden Wörter durch Leerzeichen voneinander getrennt. Derartig eindeutige Hinweise darauf, wo ein Wort beginnt und wo es endet, sind in der gesprochenen Sprache allerdings nicht enthalten (vgl. McQueen 2007). Wie wird der Inputrepräsentation entnommen, wo sich Wortgrenzen befinden? Prinzipiell greifen Hörer hierbei auf drei Informationsquellen zu, nämlich auf:

1. Informationen, die im Sprachsignal selbst enthalten sind, z. B. suprasegmentale Informationen wie Lautdauer, Prosodie oder Metrik. Hier lassen sich allerdings Unterschiede zwischen Sprachen feststellen. So konnte etwa nachgewiesen werden, dass englische und niederländische Probanden den Wechsel von starken und schwachen Silben nutzen um Wortgrenzen zu entdecken, Hörer des Französischen und Japanischen hingegen nicht.
2. Informationen, die dem einzelsprachspezifischen phonologischen Wissen entnommen werden. Hierzu zählt phonotaktisches Wissen über die erlaubten Besetzungen der Silbenpositionen, Wahrscheinlichkeiten über Lautsequenzen an Wortgrenzen oder auch Vokalharmonie.
3. Informationen, die aus dem Wettbewerb zwischen den verschiedenen Kandidaten bereitstehen. Die Idee, die sich dahinter verbirgt, ist, dass die mit dem Input kompatiblen Kandidaten die Möglichkeiten, wo sich Wortgrenzen befinden, gewissermaßen automatisch einschränken.

Wenn Sprecher tatsächlich diese drei Typen von Informationen nutzen, um Wortgrenzen zu entdecken, kann natürlich die Frage gestellt werden, ob es hierbei zu einer Hierarchisierung bzw. Gewichtung kommt, schon deshalb, weil die Auswertung der unterschiedlichen Informationsquellen unter Umständen widersprüchliche Ergebnisse liefert. Mattys/White/Melhorn (2005) haben hierzu direkte experimentelle Evidenz von englischsprachigen Versuchspersonen vorgelegt, der gemäß lexikalischen Informationen, also Information, die sich aus dem Wettbewerb ergibt, der Vorzug gegenüber segmentalen Informationen aus dem Sprachsignal gegeben wird. Diesen wiederum scheint größere Bedeutung zuzukommen als signalbasierten suprasegmentalen Informationen.

Eine weitere Frage, die sich angesichts des Vorhandenseins multipler Hinweise zur Bestimmung von Wortgrenzen stellt, lautet: Wie werden diese Hinweise integriert? Norris/McQueen/Cutler/Butterfield (1997) haben hierfür einen speziellen Algorithmus vorgeschlagen, mit dessen Hilfe

das Worterkennungssystem überprüft, wie gut die Wortgrenzen verschiedener Wortkandidaten mit, zum Beispiel, metrischen Informationen kompatibel sind. Passt ein Wortkandidat nicht zur metrischen Struktur, sinkt automatisch sein Aktivationsniveau.

Die **Selektion** als Teilprozess der Worterkennung besteht in der Aktivierung von Wortkandidaten im mentalen Lexikon, die mit der Inputrepräsentation kompatibel sind.

Zum Begriff

Integration

Die erfolgreiche Selektion des Kandidaten, der am besten zur Inputrepräsentation passt, stellt noch nicht den Abschluss des Worterkennungsprozesses dar. Am deutlichsten wird dies am linguistischen Phänomen der Homophonie. Wörter mit ganz verschiedener Bedeutung können sich hinsichtlich ihrer phonologischen Eigenschaften gleichen. Beispiele für das Deutsche sind *M/macht*, *H/haut* oder *Bank*. Es fehlt also noch die Entscheidung darüber, mit welcher Bedeutung die erkannte phonologische Einheit verknüpft werden muss. Im Fall von *M/macht* und *H/haut* ist klar, dass diese Entscheidung relativ leicht unter Beachtung der syntaktischen Umgebung getroffen werden kann. Da die lautlichen Eigenschaften von *haut* und *macht* beide mit der morphosyntaktischen Markierung für die 3. Person Singular kompatibel sind, müsste die syntaktische Umgebung eben genau so gestaltet sein, dass sie an der entsprechenden Position ein Verb erlaubt, das über diese morphosyntaktischen Eigenschaften verfügt. Dies setzt natürlich voraus, dass die Prozesse, die für die Erkennung der Satzsyntax verantwortlich sind, entsprechende Informationen bereits zur Verfügung stellen können.

Im Fall von *Bank* (Geldinstitut vs. Sitzmöbel) ist die Lage anders. Die lautliche Gestalt ist morphosyntaktisch mit allen Instanzen beider Bedeutungen kompatibel, nämlich mit dem Nominativ, Genitiv, Dativ und Akkusativ Singular. Die syntaktische Umgebung selbst kann die Entscheidung demnach nicht herbeiführen. Es sind also Bedeutungsinformationen notwendig, die beispielsweise aus dem unmittelbaren Satzkontext ermittelt werden müssen. *Die Bank hat nachmittags geschlossen* z. B. macht die Bedeutung ›Sitzmöbel‹ unwahrscheinlich. Es kann aber auch notwendig werden, Informationen aus dem größeren Kontext mit einzubeziehen. Aus dem Satz *Dann bin ich zur Bank gegangen* in Isolation kann auch die Bedeutungsinformation der Wörter in der Umgebung nicht dabei helfen, den richtigen Kandidaten auszuwählen. Hier muss ein Hörer auf Informationen aus dem vorausgegangenen Diskurs oder auf sein Weltwissen zurückgreifen.

Die **Integration** als Teilprozess der Worterkennung besteht in der Verknüpfung genau eines von mehreren aktivierten und miteinander konkurrierenden lexikalischen Kandidaten mit dem syntaktischen, semantischen oder Diskurskontext.

Zum Begriff

Die zeitliche Koordinierung von initialem Kontakt, Selektion und Integration

Die Unterscheidung von initialem Kontakt, Selektion und Integration gestattet es, die einzelnen Arbeitsschritte bei der Worterkennung zu differenzieren und zu beschreiben. Darüber, wie diese Prozesse in der Zeit ablaufen, wurde bisher allerdings nicht viel gesagt. Obwohl wir in einigen Situationen ein Wort aus dem Kontext heraus vorhersagen können, seine lautliche Form also gar nicht hören müssen, beginnt die Worterkennung im Normalfall zumindest mit dem teilweisen Aufbau der Inputrepräsentation. Die Identifikation einiger Segmente und suprasegmentaler Eigenschaften im Schallstrom muss den anderen Prozessen also ein Stück weit vorausgehen.

Informationsfluss: Die Frage nach dem zeitlichen Ablauf der verschiedenen Prozesse berührt noch eine andere Frage, die für die aktuelle Forschung nach wie vor von großer Relevanz ist: Wie modelliert man am besten den Informationsfluss zwischen den Prozessebenen? Kann die Worterkennung als reiner ›bottom-up‹-Prozess beschrieben werden, bei dem autonome hochspezialisierte Prozesskomponenten den Input nach und nach so umwandeln und jeweils an eine nächsthöhere Prozessebene weitergeben, bis am Ende die Worterkennung erfolgt ist? Oder kann die Worterkennung überhaupt nicht erklärt werden, wenn nicht zumindest teilweise Information auch von den oberen Prozessebenen nach unten fließen kann, Information also zumindest teilweise auch ›top-down‹ fließt?

Die Entscheidung darüber ist neben der Frage, ob das Worterkennungssystem auf eine prälexikalische Ebene angewiesen ist oder nicht, ein Hauptunterscheidungsmerkmal zwischen den verschiedenen derzeit vorhandenen Modellen der Worterkennung.

Zur Vertiefung

Einige Modelle der Worterkennung

Das Kohortenmodell (Marslen-Wilson/Tyler 1980; Marslen-Wilson/Welsh 1978) war lange Zeit eines der einflussreichsten Modelle der Worterkennung. Es nimmt an, dass mehrere Wortkandidaten gleichzeitig aktiviert werden, und zwar sobald das erste Phonem erkannt ist. Das Modell ist stark interaktiv. Alle zur Verfügung stehenden Informationen werden genutzt, um Kandidaten auszuschließen bis am Ende das richtige Wort übrigbleibt.

TRACE (McClelland/Elman 1986): In diesem Netzwerk-Modell werden drei Ebenen (*layer*) angenommen, auf denen sich die für die Worterkennung relevanten Informationen (*featural*, *phonemic* und *lexical*) befinden. Feature-Knoten werden vom Input aktiviert. Diese aktivieren Phonem-Knoten, welche ihrerseits mit Wort-Knoten verbunden sind. Umstritten ist eine Annahme, die das TRACE-Modell macht: Wort-Knoten senden per Feedback Aktivierung an ihre eigenen Phonem-Knoten.

Neighborhood Activation Model/ PARSYN (Luce, 1986; Luce/Pisoni, 1998): Den Kern dieses Modells bildet ein Algorithmus, mit dessen Hilfe bestimmt wird, welche Wortkandidaten bzw. ›Nachbarn‹ wie stark durch den Input aktiviert werden (*global similarity metric*). Dies geschieht über

den Vergleich von Phonemen. Wörter sind dann ›Nachbarn‹, » ... if they differ by no more than one phoneme, whether by deletion (cat: at), addition (cat: scat, cast, cattle), or substitution (cat: bat, cot, cab).« Die Stärke der Aktivierung eines ›Nachbarn‹ wird zusätzlich von der Wortfrequenz (log-frequency) beeinflusst.

Eine ausführliche Beschreibung vieler verschiedener Worterkennungsmodelle findet sich in Magnuson/Mirman/Harris (2012).

5.3 | Die Analyse der Satzstruktur

Das menschliche Satzanalysesystem, auch **menschlicher Parser** (*human parser*) genannt, erarbeitet aus der durch die Worterkennung aktivierten syntaktischen Information, Teilen der prosodischen Information und der Reihenfolge der Wörter im Satz dessen syntaktische Struktur. Das ist, technisch gesprochen, die Gesamtheit der syntaktischen Beziehungen in einem gegebenen Satz. Einleitend werden wir versuchen, unser Problembewusstsein zu schärfen, indem wir die erstaunliche Komplexität der syntaktischen Beziehungen anhand eines Beispiels herausarbeiten. Davon ausgehend machen wir uns in Umrissen ein Bild vom grammatischen Wissen über das in Kapitel 2 schon Erwähnte hinaus. Im dritten Schritt werfen wir Blicke auf die Arbeitsweise des Parsers anhand exemplarischer, empirischer Befunde und vollziehen, auf sie bezogen, die großen Schritte der Theoriebildung über das Syntaxanalysesystem nach.

5.3.1 | Die Herausforderung an das Parsing-System

Die natürliche Sprache ist keine formale Sprache wie etwa die Programmiersprachen ALGOL oder FORTRAN oder die Sprache der formalen Logik. Sie ist nicht am Reißbrett erdacht und definiert worden, sondern Hand in Hand mit der Entwicklung der Urmenschen zum Homo sapiens vor etwa 150.000 Jahren entstanden und seither stetig verändert worden. Das Ergebnis ist – jedenfalls im Bereich der formalen, grammatischen Eigenschaften – ein hoch differenzierter Apparat aus syntaktischen Kategorien mit Kombinationsregeln. Die Beherrschung dieses ›Apparats‹ gelingt dem menschlichen Gehirn auf eine bis heute nur bruchstückhaft bekannte Weise aber im Normalfall ohne nennenswerte Anstrengung in der frühen Kindheit. Ganz anders verhält es sich mit dem Versuch, diese Fähigkeit des Menschen wissenschaftlich zu explizieren. Die Ausgangslage der dahingehenden Bemühungen ist von zwei massiven Herausforderungen gekennzeichnet, von der Komplexität der syntaktischen Oberfläche und von der syntaktischen Mehrdeutigkeit.

Die Komplexität der syntaktischen Oberfläche

Ein Satz ist eine Folge von Wörtern aber nicht jede Folge von Wörtern ist ein Satz. In einem syntaktisch wohlgeformten Satz ist jede Wortform mit allen anderen syntaktisch verknüpft, unmittelbar oder mittelbar, und Aus-

druck dieser Verknüpfung sind grammatische Morpheme und die Wortstellung.

Zur Illustration

Beispiel für die Komplexität der syntaktischen Struktur

In einem Zeitungsbericht über eine Unwetterkatastrophe in Darwin, Australien, stand der folgende Satz:

(5–1) *Eine Frau, der der Sturm alles genommen hat, hatte zwei Tage vor Beginn der Katastrophe ihre Tochter, die im Süden des Landes wohnte, überredet, zu einem Besuch zu kommen.*

Damit in diesem Satz jedes Wort die Form und die Position bekommt, die es hat, musste der Sprecher ungefähr achtzig verschiedene syntaktische Relationen berücksichtigen und zum Ausdruck bringen, die meisten davon mehrmals im Satz. Wird eine davon nicht beachtet, erscheint an einer berechenbaren Stelle etwas Ungrammatisches, was übrigens vom Hörer mikroskopisch genau registriert und gegebenenfalls auch korrigiert werden kann. Jede Wortform steht also im Einklang mit einer großen Zahl syntaktischer Strukturerefordernisse, was sich z. B. anhand der Wortform *Landes* veranschaulichen lässt. Diese Form und ihre Position sind bestimmt durch die folgenden syntaktischen Relationen:

- Der Wortstamm muss ein Nomen sein und nicht z. B. eine Präposition, sonst entstünde etwas Ungrammatisches: **... im Süden des von wohnte...*
- Das Nomen muss der Klasse der Nomina angehören, die sich mit einem Artikel zu einer Nominalphrase verbinden und nicht alleine schon eine solche bilden, wie etwa Eigennamen: **...im Süden des Berlins wohnte ...*
- Das Genus des Nomens muss Neutrum sein. Ein feminines Nomen würde nicht mit der Form des Artikels *des* kongruieren: **im Süden des Gegend*
- Der Numerus des Nomens muss Singular sein.
- Der Kasus muss Genitiv sein.
- Das Nomen muss dem Artikel folgen und nicht vorangehen: **... im Süden Landes des wohnte ...*
- Das Nomen muss der Verbform *wohnte* vorausgehen. Es bildet mit dem Artikel zusammen eine Nominalphrase, deren Bestandteile (Konstituenten) nicht durch das phrasenexterne finite Verb unterbrochen werden können.
- Das Nomen *Landes* steht rechts von *wohnte*, weil die Nominalphrase, der es angehört, zum sog. Mittelfeld des Nebensatzes gehört, und das Mittelfeld geht im Nebensatz dem final stehenden finiten Verb voran.
- Das Nomen *Landes* steht rechts von *Süden*, weil die Phrase, der es angehört, eine attributive Relation zu *Süden* erfüllt und die Determinatorposition in der dominierenden Nominalphrase *dem Süden* durch *dem* besetzt ist. Würde diese Tatsache nicht berücksichtigt, würde die nicht wohlgeformte Folge **... im des Landes Süden ...* entstehen.
- *Landes* steht rechts von *die*, dem Relativpronomen, weil das Relativpronomen die satzinitiale Position einnimmt, die den Nebensatz mit der dominierenden Konstituente verbindet.

- *Landes* steht rechts von *Tochter*, weil der Nebensatz, dem es angehört, als ganzer ein Attribut zu *Tochter* ist und seiner Position folgt. Die Relation ist der zwischen *Landes* und *Süden* ähnlich.

Wir brechen die Aufzählung hier ab. Sie hat wohl hinreichend erkennbar gemacht, dass jede syntaktische Eigenschaft einer Wortform und ihre Position in der Wortfolge sich aus dem Ensemble von syntaktischen Relationen und ihrem Zusammenspiel in dem gegebenen Satz ergibt.

Syntaktische Mehrdeutigkeit

Die syntaktischen Beziehungen zwischen den Wörtern im Satz sind Ausdruck ihrer inhaltlichen Bezüge. Dieselben Wörter in anderen syntaktischen Beziehungen drücken einen anderen Inhalt aus, wie (5–2) zeigt.

- (5–2) *Eine Frau [...] hatte vor Beginn der Katastrophe zwei Tage ihre Tochter [...] überredet, ...*

Die syntaktische Struktur muss im Verstehensvorgang erkannt werden, damit der Inhalt der Äußerung richtig verstanden wird. Der Hörer verarbeitet die eingehende Folge von Wortformen ebenso detailliert, wie der Sprecher sie konstruiert; der Hörer hat allerdings ein Problem, das der Sprecher nicht hat, das **Mehrdeutigkeitsproblem**.

Das **Mehrdeutigkeitsproblem** besteht darin, dass die Informationen, die zur Produktion des Satzes geführt haben, dem Produkt entnommen werden müssen. Das Analyseproblem wäre trivial, wenn einer jeden Wortform alle aktuell relevanten Informationen direkt und eindeutig zu entnehmen wären. Das ist aber nicht der Fall. Die Form *Tochter* z. B. kann Nominativ, Genitiv, Dativ und Akkusativ Singular sein. Der tatsächliche Kasus im Satz (5–1) ergibt sich erst aus Informationen, die später im Satz folgen. Bis dahin kann das Parsing-System keine sichere Entscheidung treffen.

Temporäre Mehrdeutigkeit ist eine Eigenschaft von Wortformen, die die Analyse sehr aufwändig macht, denn es müssen bis zur vereindeutigenden Information (disambiguierende Information) entweder alle Möglichkeiten einkalkuliert bleiben oder es muss eine vorläufige Entscheidung getroffen werden, die, wenn sie falsch war, später wieder revidiert werden muss (**Reanalyse**).

Zu der Schwierigkeit infolge von temporärer (lokaler) Mehrdeutigkeit tritt ein Zeitproblem hinzu. Der Parser muss nämlich möglichst ohne zeitliche Verzögerung arbeiten, weil ihm anderenfalls nachfolgende Informationen entgehen.

Des Weiteren kann das Analyseproblem noch dadurch verschärft werden, dass die syntaktische Information, die der Hörer dem Satz entnehmen kann, letzten Endes nicht nur eine, sondern mehrere Strukturen zulässt. Eine Wortfolge ist dann also nicht nur temporär mehrdeutig, sondern **endgültig mehrdeutig**. Im Deutschen bilden z. B. präpositionale Phrasen eine häufige Quelle für endgültige Mehrdeutigkeit. Ein Beispiel ist (5–3).

- (5–3) *Bertrand Russell wurde wegen Demonstrierens vor dem Parlament verhaftet.*

Mehrdeutigkeit
ist hoch sprach-
spezifisch

Mit der Phrase *vor dem Parlament* kann der Sprecher den Ort der Verhaftung oder den Ort des Demonstrierens gemeint haben. Syntaktisch ist die Phrase also entweder attributiv mit *Demonstrierens* oder adverbial mit *verhaftet* verknüpft. Da der Sprecher wohl eines von beiden und nicht beides ausdrücken wollte, stellt sich dem Hörer beim Verstehen also die Aufgabe, die gemeinte Lösung von der nicht gemeinten zu unterscheiden. Dazu sind zusätzliche Informationen aus dem weiteren Kontext erforderlich. Dies illustrieren die Sätze der Gruppe (5–4).

- (5–4) (a) *Der Ober bringt die Suppe schnell herein.*
 (b) *Der Ober bringt die Suppe heiß herein.*
 (c) *Der Ober bringt die Suppe gelangweilt herein.*

Hinsichtlich ihrer lexikalisch-syntaktischen Information und Reihenfolge unterscheiden sich die drei Sätze nicht. Und doch ist das adjektivische Element (*schnell, heiß, gelangweilt*) in jedem Satz jeweils mit einem anderen Element verknüpft als in den beiden anderen Sätzen: *schnell* bezieht sich auf *bringt*, *heiß* auf *Suppe* und *gelangweilt* auf *Ober*. Diese Bezüge gehen nicht aus der syntaktischen Information hervor; die ist ja, wie gesagt, in allen Sätzen dieselbe. Dass trotzdem kein Verstehensproblem auftritt, kann also nur daran liegen, dass die intendierte Struktur vom Hörer unter Hinzuziehung von außersyntaktischem Wissen erkannt wird.

Fazit: Um den Inhalt eines Satzes zu verstehen, muss seine syntaktische Struktur verstanden werden. Die syntaktische Struktur umfasst die Beziehungen von Wortformen innerhalb ihrer jeweiligen Phrase (Konstituentenstruktur) und die Funktion der Wortform in der Phrase und letztlich im Satz (Funktionale Gliederung). Dem Parsing-System stehen für die Analyse die lexikalischen Informationen zur Verfügung, die zu jeder Wortform dem Lexikon zu entnehmen sind. Sie bilden ja das Ergebnis der Worterkennung. Im Falle syntaktisch mehrdeutiger Wortformen (Beispiel *Tochter* in 5–2) liefert die Worterkennung nicht nur eine Information, sondern alle lexikalischen Alternativen.

Kernfragen der Parsing-Forschung: Trotz all dieser Schwierigkeiten scheint der Hörer die syntaktische Analyse im Allgemeinen reibungslos und zügig zu bewältigen. Mitunter ist er sogar in der Lage, eine vom Sprecher abgebrochene Äußerung syntaktisch wohlgeformt so zu vervollständigen, wie sie der Sprecher geplant hatte. Diese Tatsache wirft eine Reihe spannender Fragen auf, die sich zu zwei Hauptfragen zusammenfassen lassen.

1. Von welchem Wissen macht der Parser Gebrauch, um die syntaktische Struktur eines Satzes zu rekonstruieren? (Wissensbasis)
2. Nach welchen Prinzipien arbeitet das Parsing-System?

Zur Illustration Beispiele für syntaktische Mehrdeutigkeit

Temporär (lokal) mehrdeutige Sätze:

- (5–5) *Hans versprach Maria keine Zigaretten mehr anzuvertrauen.* (potentieller *garden-path*-Effekt)
 (5–6) *Das sind die Professorinnen, die Hans gesehen haben.* (potentieller *garden-path*-Effekt)

- (5-7) *Peter hat die Ärztin im Krankenhaus gestern sehr geholfen.* (potentieller *garden-path*-Effekt; aus Hemforth/Strube (1999))
 (5-8) *Emma sah Fritz fallen.*
 (5-9) *Emma sah Fritz raubte die Kasse aus* (Satzzeichen weggelassen)

Endgültig mehrdeutige Sätze:

- (5-10) *Der Polizist verfolgte den Dieb mit dem Fahrrad.*
 (5-11) *Das Kind aß die Kekse auf dem Tisch.*
 (5-12) *Das Mädchen sah die Frau.*
 (5-13) *Er beobachtete viele Tiere auf dem Weg zur Waldlichtung.*

5.3.2 | Das syntaktische ›Wissen‹ des Parsers

Will man sich ein Bild davon machen, welches ›Wissen‹ das syntaktische Analysesystem besitzt, so ist von vornherein zweierlei strikt zu unterscheiden. Das eine ist das syntaktische Regelsystem, also der Parser im engen Sinne. Das andere ist das ›Wissen‹, wie nichtsyntaktische Informationen zur Erkennung der syntaktischen Struktur einer Äußerung verarbeitet werden, um im Fall von Mehrdeutigkeiten zwischen mehreren Parsing-Ergebnissen zu entscheiden.

Zum syntaktischen Wissen im engeren Sinne zählen:

- das Regelwissen über Elemente und Beziehungen in Konstituentenstrukturen
- das Wissen über syntaktische Funktionen (Subjekt, Attribut, Prädikat)
- universale syntaktische Prinzipien beim Aufbau einer syntaktischen Phrase

Sprachspezifisch
vs. universal

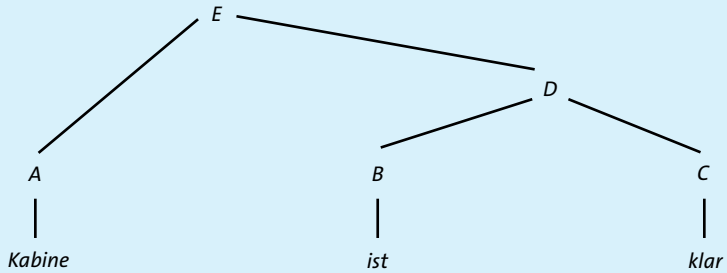
Konstituentenstruktur: Die Wortformen eines Satzes sind durch zwei Systeme von Beziehungen geordnet, ein System von **Präzedenzbeziehungen** und ein System von **Dominanzbeziehungen**. Erstere schränken die Möglichkeiten der linearen Abfolge, letztere die Möglichkeiten zur Phrasenbildung ein und beide sind durch ein generelles Prinzip aufeinander bezogen: Das **Lokalitätsprinzip**. Es besagt ganz allgemein, dass die Bestandteile (Konstituenten) einer Phrase einander benachbart sind.

Zur Illustration

Beispiele für die Konstituentenstrukturanalyse

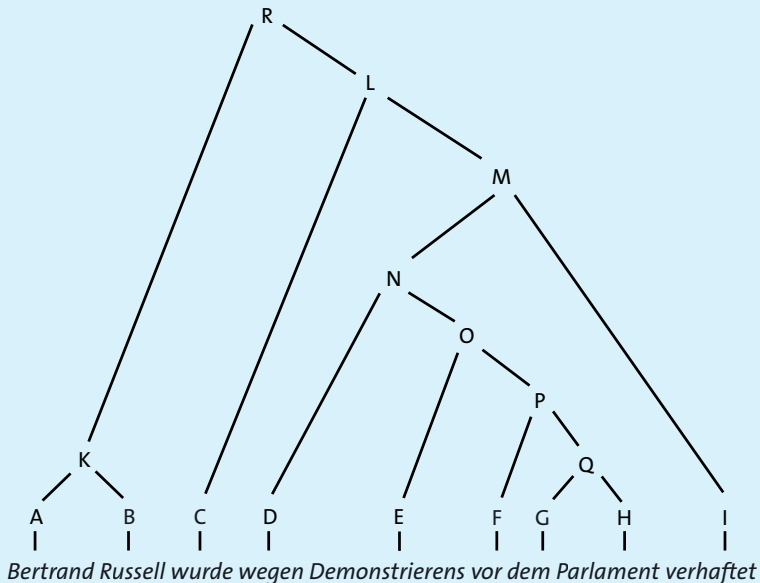
Wer z. B. die deutsche Sprache beherrscht, ist in der Lage, die Wortkette eines Satzes in Teilketten zu zerlegen, deren Elemente er intuitiv als enger zusammengehörig empfindet als mit Elementen anderer Teilketten also mit Elementen außerhalb der Phrasengrenzen. Die Zerlegung des Satzes *Kabine ist klar* ergibt in etwa die Struktur in Abb. 5.5.

Abb. 5.5: Intuitive Konstituentenstruktur des Satzes *Kabine ist klar*



Die Konstituentenstruktur ist also ein Ensemble hierarchischer Beziehungen. Die Kategorie E dominiert A und D direkt, D dominiert B und C direkt und E dominiert B und C indirekt. Die Umkehrbeziehung der Dominanz ist die **Ist-Konstituente-von-Beziehung**. Also B und C sind Konstituenten von D und zwar unmittelbare Konstituenten; A ist unmittelbare Konstituente von E, ebenso D. B und C sind Konstituenten von E. Ein längerer Satz hat eine komplexere Konstituentenstruktur. Die intuitive Struktur von (5-3) ist etwa die in Abb. 5.6.

Abb. 5.6: Intuitive Konstituentenstruktur von Satz (5-3). Die Bedeutung der Symbole an den Knoten ist außer Acht gelassen. Strukturelle Gemeinsamkeiten von Phrasen sind hier einfachheitshalber nicht ausgedrückt.



Da Satz (5-3) syntaktisch mehrdeutig ist, gibt es noch eine alternative Konstituentenstruktur; die lassen wir einstweilen außer Acht.

Empirische Evidenz dafür, dass die Konstituentenstruktur und damit also auch das zu Grunde liegende Regelwissen beim Verstehen eingeschaltet ist, liefern experimentelle Befunde aus sog. Klick-Experimenten, Messungen elektrischer Potenziale bei der Verarbeitung (ERP-Muster) und Verhaltensdaten von Reaktionen auf sogenannte *garden-path*-Strukturen.

Datenlage

- Im **Klick-Experiment** hören Probanden Sätze, in denen Klickgeräusche einmal zwischen, einmal innerhalb von Grenzen syntaktischer Phrasen eingefügt sind. Sofern die Teilkette, die eine Phrase bildet, auch in der Verarbeitung eine Einheit bildet, sollten Fremdkörper innerhalb der Phrase eher ignoriert oder in der Wahrnehmung an die Grenzen verschoben werden. Die Daten bestätigen den Einfluss von phrasalen Einheiten beim Satzverstehen. Ein klassisches Klick-Experiment berichten Holmes/Forster (1972).
- **ERP-Daten:** Die Reaktionen des Verstehensystems auf syntaktische Abweichungen lassen sich in spezifischen elektrischen Potentialen außen an der Hirnschale nachweisen. So äußert sich die Verarbeitung eines Satzes mit einem syntaktisch nicht-wohlgeformten Wortformenübergang (**Der Freund wurde im besucht*) in einem spezifischen negativen Potential im linken vorderen kortikalen Bereich nach etwa 100 – 200 ms und einem späteren positiven Potential nach etwa 600 ms (vgl. den Überblick in Friederici/Cramon 1999).
- **Holzweg-Effekt** (*garden path effect*): Die übliche Reaktion des Verstehenssystems beim Verarbeiten eines Satzes wie (5–5) reicht – je nach Schwere des Effekts – von ›Stolpern‹ über Neubeginn der Analyse bis zur Ablehnung des Satzes. Die Reaktion wird auf eine Irritation des Parsers zurückgeführt, die dadurch zustande kommt, dass ein Wort, das relativ spät in einem Satz erscheint, nicht mehr in die bis dahin schon aufgebaute, letzten Endes aber unzutreffende Struktur eingefügt werden kann, weil sie eben nicht mit den Konstituentenstruktur-Regeln (KS-Regeln) in Einklang gebracht werden kann. Dass vorübergehend eine zwar zulässige, tatsächlich aber unzutreffende Analyse durchgeführt worden ist, wird verschieden erklärt. Eine Erklärung argumentiert eben mit dem Verlauf des Analysevorgangs auf der Basis von KS-Regeln; dazu unten mehr.

Syntaktische Funktionen: Eine syntaktische Funktion ist eine spezifische komplexe Konfiguration innerhalb der Konstituentenstruktur des Satzes. Betrachten wir noch einmal Satz (5–1). Auf die Frage, wer wen überredet hat, würden die meisten Deutschsprachigen sicher spontan antworten, dass die Frau diejenige war, die die Tochter überredet hat. In der syntaktischen Beschreibung wird dieser Analyse Ausdruck verliehen, indem man sagt, *eine Frau* nimmt die Funktion des Subjekts im Hauptsatz ein und *ihre Tochter* die Funktion des direkten Objekts. Unter Bezug auf die Konstituentenstruktur des Satzes lässt sich die Subjektfunktion also als die Nominalphrase definieren, die der obersten Verzweigung ihres Satzes hierarchisch am nächsten steht. Dabei ist unerheblich, welche Position die Phrase in der linearen Abfolge des Satzes einnimmt. Das direkte Objekt ist in dieser Redeweise die Nominalphrase, die dem verbalen Element desselben Satzes hierarchisch am nächsten steht.

Komplexität als Problem

Kopfelemente und Argumente: Natürlich ist das alles komplizierter, als es auf den ersten Blick erscheint. Zunächst einmal kommt noch die Bedingung hinzu, dass Subjekts- und Objektsnominalgruppe in einer spezifischen syntaktischen Beziehung zum Verb stehen müssen. Jedes Verb und viele andere Wörter weisen im Lexikon die Information auf, dass sie – vom Subjekt abgesehen, das in jedem Satz gefordert ist – bestimmte syntaktische Ergänzungen in der Konstituentenstruktur **fordern**. Das Verb *finden* z. B. fordert neben sich eine nominale Ergänzung oder eine *dass*-Satzergänzung. Das Substantiv *Frage* fordert eine Ergänzung um eine Präpositionalphrase (PP) oder um einen *ob*-Satz. Jede Präposition fordert eine Ergänzung um eine Nominalphrase (NP). In der Grammatiktheorie heißen diese Ergänzungen **Argumente** und das Ensemble der Argumente heißt **Argumentrahmen**. Das Wort, das einen Argumentrahmen fordert, heißt relativ zu diesem Rahmen das Kopfelement, kurz der **Kopf**. Das Kopfelement hat auch noch andere syntaktische Wirkungen. Die wichtigste ist, dass die kategoriale Eigenschaft des Kopfes, also Verb oder Nomen oder Präposition zu sein, die kategoriale Eigenschaft der gesamten Phrase bestimmt, die es mit seinen Argumenten bildet. Ein Verb bildet also eine Verbphrase, ein Nomen eine NP, eine Präposition eine PP. Subjekt und Objekt müssen Argumente des Verbs sein – jedenfalls in einem gewissen Sinne.

Empirische Evidenz: Inwiefern ist dies alles nun für den Parsing-Vorgang von Bedeutung? Die Bedeutung ergibt sich daraus, dass angenommen wird, ein kompetenter Sprachbenutzer habe neben dem Konstituentenstruktur-Wissen auch ein davon zu unterscheidendes direktes grammatisches Wissen von den syntaktischen Funktionen. Er habe also einen integrierten, unbewussten Begriff jeweils von Subjekt und Objekt und verbinde mit diesen syntaktischen Funktionen Annahmen über ihre Ausdrucksseite, ohne all die genannten grammatischen definitatorischen Einzelheiten jedes Mal ausführlich zu analysieren und aus dem Ergebnis die Subjekts- bzw. Objektsbeziehung abzuleiten.

Diese Annahme ist nicht unplausibel. Sie wird schon durch die kursorische Beobachtung bestätigt, dass *eine Frau* im Satz (5-1) spontan als Subjekt verstanden wird. Ohne dieses Wissen zu unterstellen, wäre nicht plausibel, warum nicht *ihre Tochter* als Subjekt und *eine Frau* als Objekt verstanden werden. Der Form nach können beide NPs Nominativ und Akkusativ sein, und Objekterstsätze sind im Deutschen möglich, wie in *Die Kabine habe ich klar gemacht*.

Auch experimentelle Beobachtungen stärken die Annahme, dass das Wissen über die Existenz von Subjekt und Objekt direkt im Parsing-Prozess eingesetzt wird und zwar auf dem Weg einer Subjekt-vor-Objekt-Präferenz bei der Analyse der ersten Nominalgruppe im Satz. Subjekterst- und Objektzweitsätze werden leichter gelesen (vgl. Hemforth/Strube 1999, S. 259). Subjekt/Objekt-mehrdeutige Relativpronomen werden im ersten Analysedurchgang eher als Subjekt analysiert als als Objekt (vgl. Friederici 1998).

Grammatische Prinzipien: Die Konstituentenstruktur (KS) eines Satzes ist ein Ensemble von vielen syntaktischen Relationen. Würde das syntaktische Wissen eines kompetenten Sprechers/Hörers vollständig aus der

Kenntnis aller in seiner Sprache möglichen KS-Relationen bestehen, würde das zu der Annahme zwingen, dass er ein enorm umfangreiches und enorm differenziertes Repertoire solcher KS-Relationen beherrscht und beim Verstehen eines Satzes jedes Mal die aktuell zutreffenden davon identifiziert und anwendet, veranlasst durch die lexikalischen Eigenschaften der jeweiligen Wortform. Selbst bei vorsichtiger Schätzung müsste man mehrere Tausend syntaktischer KS-Bedingungen annehmen. Für die Beschreibung relativ kleiner Ausschnitte des Deutschen und des Englischen mit Mitteln einer KS-Syntax sind nach Brockhaus (1971, S. 78) etwa eintausend verschiedene Kategorien nötig; diese Syntax berücksichtigt keine adverbialen Gruppen, keine PPs und nur einen Teil von Relativsatztypen (Brockhaus 1971, S. 69). Eine KS-Syntax des Deutschen enthält also leicht mehrere tausend Kategorien und ein Mehrfaches davon an zulässigen KS-Relationen.

Die relative Leichtigkeit und die kurze Dauer des Spracherwerbs durch das Kind einerseits und die rasche und mühelose Produktion und Rezeption von Sätzen durch den kompetenten Sprecher begründen Zweifel daran, dass das syntaktische Wissen des Menschen eine KS-Syntax ist. Die Zweifel werden verstärkt durch die Feststellung, dass syntaktisch sehr verschiedene Sprachen hinsichtlich einiger struktureller Einschränkungen übereinstimmen und dass sich aus deren Zusammenspiel die KS-Möglichkeiten aller Sprachen vorhersagen lassen. Wie in Kapitel 3.5 erwähnt, scheint sich das Kind beim Spracherwerb ebenfalls von solchen Prinzipien leiten zu lassen.

Prinzipien des Phrasenaufbaus: Die linguistische Analyse der syntaktischen Vielfalt hat ergeben, dass der syntaktische Aufbau der Sätze durch eine relativ kleine Anzahl grammatischer Prinzipien eingeschränkt ist. Nicht alle sind gefunden und einige nicht unbezweifelt. Vergleichsweise gut bestätigt ist, dass der Vielfalt der Phrasen, einfachen und komplexen, NPs, PPs, Adjektivphrasen und allen anderen kategorial verschiedenen, einige wenige Aufbauprinzipien zu Grunde liegen. Die bekanntesten sind das Kopf-Prinzip, das Kopf-Vererbungsprinzip und das Ebenen-Prinzip.

- **Kopf-Prinzip:** Jede Phrase hat genau einen Kopf.
- **Kopf-Vererbung:** Die morphologischen Merkmale einer Phrase werden beim Kopf der Phrase realisiert.
- **Ebenen-Prinzip:** Auch die Vielfalt möglicher hierarchischer Gliederungen eines Satzes lässt sich unter Bezug auf ein Prinzip reduzieren, das Ebenen-Prinzip. Man geht davon aus, dass eine Nominalphrase eine komplexe nominale Einheit ist, eine Verbalphrase eine komplexe verbale usw. und nicht umgekehrt. Kurz: Wenn X eine grammatische Kategorie ist (Nomen, Verb, usw.), dann ist X^0 die lexikalische Ebene von X und X^{\max} die phrasale Ebene (auch XP, also NP, VP usw.). Das Ebenenprinzip besagt dann: Der Kopf einer Kategorie X^i ist eine Kategorie X^j , wobei j größer oder gleich 0 ist und kleiner oder gleich i (Formulierung nach Grewendorf/Hamm/Sternfeld 1987, S. 208).

5.3.3 | Vom grammatischen Prinzip zur Analysemaxime

Die relative Einfachheit und die Universalität oben beschriebener Prinzipien spricht dafür, dass sie einen Bestand unseres syntaktischen Wissens darstellen. Und damit taucht in der Parsing-Forschung die Frage auf, ob sie beim Prozess des syntaktischen Verstehens ausgenutzt werden. Ein solches Prinzip, dem in der Parsing-Forschung viel Aufmerksamkeit zukommt, schränkt das Verhältnis von Theta-Raster und Argument-Positionen einer Phrase ein.

Theta-Raster: Hat ein Verb mehrere Argumente, so werden sie dadurch unterschieden, dass jedem eine und genau eine thematische Rollenbeziehung zum Verb zukommt. Solche Rollen, in der Grammatiktheorie **Theta-Rollen** genannt, sind z. B. die Agens-Rolle, die Thema-Rolle, die Rezipient-Rolle. Ihrer Natur nach sind die Theta-Rollen Bedeutungsbeziehungen zwischen der Verbbedeutung und den Referenten der NPs in den Argumentpositionen. Das Ensemble der Theta-Rollen eines Verbs heißt Theta-Raster.

Weiteres zum Begriff der thematischen Rolle findet sich in jeder Einführung in die Grammatiktheorie, eine kürzere Darstellung bezogen auf das Deutsche z. B. in Grewendorf et al. (1987, S. 190ff.), eine ausführlichere in Stechow/Sternefeld (1988, S. 257ff.). Eine Unzahl von Details übergehend, lässt sich das Theta-Raster, auch Theta-Kriterium genannt, etwa folgendermaßen formulieren.

- a) Jeder Argumentposition einer Phrase kommt genau eine Theta-Rolle zu und jeder Theta-Rolle genau ein Argument.

Dieses Wissen, so Pritchett (1992), wird vom Analysesystem ausgenutzt; es leitet daraus eine Analysemaxime ab, die **Theta-Verknüpfung** (*theta attachment*):

- b) Es wird bei jedem Analyseschritt versucht, das Theta-Kriterium zu erfüllen und dabei wird das maximale Theta-Raster angesetzt.

Die Annahme in b) ist theoretisch notwendig, weil der Parser der Maxime schon folgen muss, bevor ihm in der linearen Abfolge des Satzes die Information über die tatsächlich in der Phrase gegebene Theta-Struktur zugänglich ist, z. B. bei spät platziertem zugehörigem Kopfelement. Pritchett zeigt, wie der Parser in Verfolgung dieser Maxime zuweilen erfolgreich arbeitet und zuweilen in die Irre geht. Seine Beobachtungen bezieht er aus der Verarbeitung englischer Sätze mit lokaler Mehrdeutigkeit.

Evidenz: Aus Ergebnissen von Leseexperimenten ist bekannt, dass scheinbar ähnliche Mehrdeutigkeiten (hier Objekt im Hauptsatz vs. Subjekt im Nebensatz) manchmal einen Holzweg-Effekt auslösen und manchmal nicht. Dazu muss man berücksichtigen, dass die Sätze im Experiment wortweise auf dem Bildschirm präsentiert werden und nicht, wie hier, satzweise. So soll der schrittweise Input der Wortformeninformation in den Parser gesichert werden, etwa wie beim Hören. Typische Satzbeispiele im Deutschen wären (5–8) und (5–9).

Die englischen Belege sind Sätze wie (5–14) und (5–15).

- (5–14) *Without her it would be impossible.* (kein Holzweg-Effekt)
Ohne sie wäre es unmöglich
- (5–15) *Without her contributions would be impossible*
 (bewusster Holzweg-Effekt)
Ohne sie wären Beiträge unmöglich

Dass Satz (5–14) leicht zu analysieren ist und (5–15) zunächst falsch analysiert wird und in einem weiteren Schritt reanalysiert werden muss, erklärt Pritchett eben mit der Theta-Verknüpfung.

Parsing auf dem
Holzweg

Betrachten wir zuerst Satz (5–14). Die Präposition *without* ist der Kopf der PP und hat eine Argumentposition und das Theta-Raster enthält eine thematische Rolle, ›Thema‹ genannt (*theme*). Die nächste Wortform ist *her*. Sie ist wortklassenmehrdeutig, kann Possessivartikel sein und Personalpronomen. Als Personalpronomen kann sie alleine eine NP bilden und mit einer Theta-Rolle verknüpft werden. Das geschieht auch gemäß Theta-Verknüpfungsmaxime. Damit ist das Theta-Raster von *without* abgearbeitet und die Analyse geht zur nächsten Theta-Domäne über und gelangt schließlich zur korrekten eindeutigen Struktur des Satzes.

Anders bei der Analyse von (5–15). Sie läuft bis *her* genau wie von (5–14) – natürlich, denn sie hat ja bis dahin denselben Input. Dann erscheint *contributions*. Hier bieten sich strukturell nun zwei Alternativen. Entweder *contributions* ist der Kopf einer neuen NP oder *her contributions* ist als NP zu analysieren und wird mit der zunächst nur *her* zugeordneten Theta-Rolle aus dem Raster von *without* verknüpft. Da für die neue NP *contributions* noch keine Theta-Verknüpfung verfügbar ist, und die Gruppe *her contributions* eine wohlgeformte NP zur Einnahme der Argumentposition von *without* ist und dafür auch eine Theta-Rolle zur Verfügung steht, wird gemäß der Maxime der Theta-Verknüpfung diese Analyse gewählt. Das stellt sich im weiteren Verlauf der Analyse als unzutreffend heraus, weil die so gerade gebildete PP zusammen mit *would* keine Phrase bildet.

Es wird neu analysiert, entweder ab der letzten Strukturalternative oder bei schwerem Holzweg-Effekt von vorne an und unter Vergleich aller möglichen Wege. Man spricht dann von **Zusammenbruch** und **Neustart**, nicht von Reanalyse. Um zur zutreffenden Struktur von (5–15) zu gelangen, muss der Parser jedenfalls die zunächst angenommene Theta-Verknüpfung von *her contributions* rückgängig machen, *contributions* einer Argumentposition außerhalb der *without*-PP zuordnen und auch einer PP-externen Theta-Rolle, die noch nicht verfügbar sind. Dieser Schritt, so Pritchett, ist kognitiv aufwändig und die Bedingungen, unter denen er erzwungen wird, liefern eine Erklärung für den Holzweg-Effekt.

Wir haben die Arbeit von Pritchett (1992) aus mehreren Gründen hier so ausführlich dargestellt. Sie ist ein Klassiker und fast jede spätere Arbeit über prinzipiengeleitetes Parsing knüpft an den einen oder anderen Punkt darin an. Sie liefert eine überzeugende Generalisierung, von der Theta-Verknüpfungsmaxime ausgehend, auf die Anwendung aller grammatischen Prinzipien beim Parsing:

- c) Generalized Theta Attachment: Every principle of the syntax attempts to be maximally satisfied at every point during processing (Pritchett 1992, S. 138).

**Maßnahmen zur
Holzweg-
vermeidung**

Nach dieser allgemeinen Maxime arbeitet das Satzanalyse-System also eigentlich kurzsichtig, indem es lokale Entscheidungen trifft, die sich später als falsch erweisen und aufwändige Reparaturen erfordern. Die Fakten sprechen aber dafür, dass es so ist. Das ist der Preis für die ansonsten rasend schnelle und störungsfreie Arbeit des Parsers. Und dass dieser Anteil so hoch ist, und das Verstehen so unauffällig reibungslos funktioniert, liegt zum einen daran, dass schon der Sprecher darauf achtet, störungsanfällige Äußerungen, also Mehrdeutigkeiten, zu vermeiden. Der zweite Grund ist, dass in die Analyse auch noch andere Informationen eingehen, von denen gleich die Rede sein wird.

Zur Vertiefung

Weitere Meilenstein-Arbeiten zum prinzipiengeleiteten Parsing

Pritchett vorausgegangen aber weniger generell in der Konzeption ist Amy Weinbergs (1988) Erklärung der Verarbeitung leerer Elemente (*empty categories*) im Parsing unter Bezug auf das *empty-category*-Prinzip (Eine leere Kategorie muss strikt registriert sein) und das Subjazenz-Prinzip (Eine grammatische Beziehung darf sich nicht über mehr als einen Grenzknoten erstrecken) (vgl. Stechow/Sternefeld 1988). Die von Weinberg (1988) betrachteten Sätze sind von der Art wie (5–16) und (5–17).

- (5–16) *Did you watch the movie without eating?*
Hast du gesehen den Film ohne zu essen?
(5–17) *Who did you meet without greeting?*
Wen hast du getroffen ohne (ihn/sie) zu grüßen?

Damit beide zutreffend analysiert werden, muss in (5–17) der Referent von *who* sowohl dem Verb *meet* als auch *greeting* als Argument zugeordnet werden, nicht aber dem Verb *eating* in (5–16) die NP *the movie*. Gorrell (1995) geht einer Ungenauigkeit in Pritchetts (1992) Erklärung von Reanalysebedingungen nach. Um zu erklären, dass oberflächlich ähnlich mehrdeutige Sätze unterschiedlich gut analysiert werden, postuliert Pritchett (1992, S. 101) die On-line Lokalisierungsbedingung für Reanalyse (*On-Line Locality Constraint*, OLLC).

- d) The target position (if any) assumed by a constituent must be ›governed‹ or ›dominated‹ by its source position (if any), otherwise attachment is impossible for the automatic Human Sentence Processor.

Mit ›source position‹ und ›target position‹ sind Ausgangs- und Zielposition einer Konstituente bei der Reanalyse bezeichnet. Die ›source position‹ ist also die erste, i. e. die unzutreffende, und die ›target position‹ ist die korrekte.

Diese Bedingung ist in Satz (5–16) erfüllt, was eine stillschweigende Reanalyse ermöglicht, in (5–17) hingegen nicht, was zu einem Holzweg-Effekt führt.

- (5–18) *Ian knows Thomas is a train* (kein Holzweg-Effekt)
Ian weiß Thomas ist ein Zug (Satzzeichen weggelassen)

- (5–19) *Ian put the candy on the table in his mouth.* (Holzweg-Effekt)
Ian tat das Bonbon auf den/dem (im Engl. kasusmehrdeutig)
Tisch in den Mund.

Thomas in (5–18) wird zunächst als direktes Objekt von *knows* analysiert. Die Reanalyse ist möglich, weil diese Position, nämlich die Argumentposition von *knows* die Zielposition, nämlich Subjekt in dem Objektsatz, dominiert. Das ist in (5–19) nicht der Fall, und das löst den Holzweg-Effekt aus. *On the table* wird zunächst als Präpositionalobjekt zu *put* analysiert, also als Schwesterknoten von *the candy*. Diese Position dominiert aber nicht die Zielposition als PP-Attribut von *candy*. Reanalyse im Sinne von Pritchett ist also ein Prozess, in den zwei verschiedene Strukturbeschreibungen eingehen und miteinander verglichen werden. Gorrell (1995) merkt dazu an, dass auf diesen Vorgang die grammatischen Begriffe ›dominieren‹ und ›regieren‹ nicht angewendet werden können (Gorrell 1995, S. 83), denn sie seien dem Sinn der Grammatiktheorie nach Relationen in einem Satz, nicht Relationen über Sätze hinweg. Er löst das Problem durch die Konstruktion eines Parsers mit zwei Prozessbereichen, einem Strukturbauer (*structure builder*) und einem Interpretierer (*interpreter*). Damit entfällt die Notwendigkeit für ein eigenes Teilsystem des Parsers, das für die Reanalyse zuständig ist, und das ganze Strukturbeschreibungen alternativer Analyseergebnisse vergleicht. In dieser Richtung weiter geht das Modell von Stuart/Crocker (1998).

5.3.4 | Nicht-syntaktisches Wissen

Prosodisches Wissen: Zur prosodischen Information zählen die Pausierung und die intonatorische Kontur der Äußerung, die sich ihrerseits wiederum aus mehreren Parametern ergibt. Ohne zusätzliche Markierung ist z. B. der folgende Satz mehrdeutig:

- (5–20) *Peter sagt Maria sei verreist.*

Peter kann Subjekt von ...*sagt* [*Maria sei verreist*] sein oder von ... *sei verreist*] sein. Dann ist *Maria* das Subjekt des Hauptsatzes, also ... [*sagt Maria*]. Orthographisch wird die zweite Lesart durch Kommata aufgezeigt:

- (5–21) *Peter, sagt Maria, sei verreist.*

In der gesprochenen Sprache kann die syntaktische Gliederung u. a. durch die Verteilung von Pausen aufgezeigt werden, in der Lesart (5–21) etwa an den Positionen der Kommata.

Spricht man den Satz in dieser Interpretation vor sich hin, stellt man fest, dass sich mit der Pausierung auch der Melodieverlauf ändert. Die zweite Silbe von *Peter* wird höher gesprochen als in der anderen Lesart. Das ist aber bei weitem nicht alles und für sich alleine auch wieder mehrdeutig.

Die Rolle der Prosodie in der Analyse der syntaktischen Struktur steht außer Frage – und nicht nur die der Äußerung, wie schon die eindrucks-

volle Sammlung von Beobachtungen in Johns-Lewis (1986) belegt. Das Zusammenspiel von Prosodie und syntaktischer Information ist allerdings so kompliziert, dass bisher nur für punktuelle Ausschnitte zuverlässige Beobachtungen vorliegen. Wir beschränken uns hier auf die Betrachtung eines illustrativen Beispiels, die Untersuchung von Weber/Grice/Crocker (2006).

Experiment

Eye-Tracking-Studie zum Einfluss prosodischer Informationen auf die syntaktische Analyse (Weber et al. 2006)

In dieser Eye-Tracking-Studie wurde die Annahme geprüft, dass die Kontur der initialen Teilkette eines Satzes der Form $NP_1 V NP_2$ in die On-line-Analyse der syntaktischen Funktion der lokal mehrdeutigen NP_1 eingeht und die Mehrdeutigkeit damit früh aufgelöst wird. Die Versuchsperson sah eine Abbildung mit mehreren Referenten in einer Szene und hörte dazu einen Satz, der ein auf die Abbildung bezogenes Ereignis beschrieb, z. B.: *Die Katze jagt womöglich den Vogel*. Die Intonation ist die der SVO-Lesart. Zu jedem experimentellen Item gab es eine zweite Satzvariante, *Die Katze jagt womöglich der Hund*, also OVS mit der zugehörigen Kontur, hier Fokusakzent auf *Katze*. Gemessen wurde die Verteilung der Fixationen auf die abgebildeten Referenten in der Hörphase vor der desambiguierenden NP_2 . Es fand sich ein signifikanter Einfluss. Beim Hören der OVS-Version wurde signifikant mehr auf den wahrscheinlichen Agenten, also den Hund, geblickt, als in der Variante mit der SOV Kontur und zwar, wie gesagt, vor der desambiguierenden NP_2 .

Häufigkeiten: Wer eine Sprache beherrscht und alltäglich verwendet, bildet mit der Zeit Erfahrungen über häufiger und weniger **häufig** verwendete syntaktische Konstruktionen. Dieses Wissen wird beim Parsing-Vorgang laufend aktiviert und trägt zur Beschleunigung der Strukturanalyse beim Satzverstehen bei. Seit Anfang der 1990er Jahre sind kognitionspsychologische Studien diesen komplexen Prozessen nachgegangen, wie in McKoon/Ratcliff (1998) unter der Bezeichnung *Memory Based Language Processing* (MBLP) beschrieben. Dass auch die Häufigkeit einer Konstruktion in großen Textkorpora einer Sprache die Leichtigkeit der Analyse bestimmt, zeigen die Befunde von Gibson und Mitarbeitern (vgl. Gibson/Pearlmutter 1994).

Kooperationsprinzip: Des Weiteren werden in diesem Zusammenhang Konstruktionen untersucht, die alle eines gemeinsam haben, nämlich, dass sie nicht Argumentphrasen, sondern **Adjunkt-Phrasen** betreffen. Das sind Konstituenten, die nicht von einer Kopfkategorie syntaktisch gefordert, sondern einer solchen sozusagen frei hinzugefügt sind. Typische Beispiele sind Adverbiale und Attribute. Sie können syntaktisch die Form einer NP, PP oder auch eines Nebensatzes haben. Hier sind einige Konstruktionen aus den Untersuchungen von Frazier/Clifton (1996, S. 43).

(5–22) *A table of wood that was from Galicia.* (Mehrdeutige Relativsatz-anknüpfung; Tisch oder Holz aus Galizien?)

- (5–23) *Some girl hit some boy last night who was ...* (Wer ist who?)
 (5–24) *John ate the broccoli raw/naked.* (Small-clause-Mehrdeutigkeit: Wer ist roh bzw. nackt?)
 (5–25) *The nurse weighed John and Mary.* (Koordination/Konjunktion-Mehrdeutigkeit)
 (5–26) *The doctor didn't leave because he was angry.* (Mehrdeutigkeit des Negationsskopos)

Aus Ergebnissen von Paraphrase-, Lesezeit- und Ratingaufgaben leiten Frazier und Clifton (1996) eine Annahme über den Analysevorgang bei der Verarbeitung solcher Sätze ab. Die Befunde stützen die Theorie, dass eine syntaktisch temporär nicht analysierbare Phrase zu der jeweils gerade bearbeiteten thematischen Verarbeitungsdomäne ›assoziiert‹ wird und zwar unter Ausnutzung struktureller und nicht-struktureller (i. e. nichtsyntaktischer) Prinzipien.

Eines dieser nicht-strukturellen Prinzipien kann im größeren Zusammenhang mit Prinzipien kommunikativen Handelns gesehen werden. Diese Sicht besagt, dass sprachliche Kommunikation eine Art und Weise gemeinsamen Handelns ist. Für dieses Handeln (*joint action*; vgl. Clark 1996) verlassen sich die Beteiligten auf generelle Prinzipien der Kooperation; exemplarisch sind die von Paul Grice (1975) aufgestellten Konversationsmaximen der Quantität, Qualität, Relevanz und Modalität. Der Kooperativität zuzuordnen wäre die psycholinguistische Strategie, als die jeweils aktuelle thematische Verarbeitungsdomäne die maximale Teilstruktur der zuletzt bearbeiteten Kategorie mit Theta-Raster zu behandeln. Frazier und Clifton (1996) nehmen also einen neuen Typ von Relationen im sprachlichen Wissen des Menschen an, die sog. *association*, das Kernstück ihrer *Construal*-Hypothese.

Weltwissen: Wie anhand von Satz (5–5) (*Hans versprach Maria ... anzuvertrauen*) erläutert, verursacht eine solche Konstruktion der Analyse Verarbeitungsprobleme. Eine syntaktisch mehrdeutige Wortform am Satz-anfang wird aufgrund von syntaktischen Analysepräferenzen in einer Lesart verstanden, die sich erst spät im Satz als unzutreffend erweist. Die Analyse endet in einer Sackgasse; in leichten Fällen kann sozusagen im Flug durch Reanalyse eines Teils der Struktur die zutreffende Struktur aufgebaut werden. Bei schwerer Irritation wird abgebrochen und neu gestartet. Hartnäckige *garden-path*-Sätze werden ohne explizite, bewusste Strukturbetrachtung gar nicht als grammatisch erkannt.

Zu den hartnäckigsten Holzweg-Effekten zählen im Englischen Sätze mit uneingeleitetem Relativsatz mit initialer finit/infinit-Mehrdeutigkeit der Verbform wie in (5–27).

- (5–27) *The horse raced past the barn fell / and fell.*

Die *and-fell*-Konstruktion wird leicht analysiert, die *fell*-Konstruktion von vielen Versuchspersonen als nicht wohlgeformt beurteilt. Dem Urteil liegt die Analyse zu Grunde, dass *raced* 3. Pers. Sing. Prät. Aktiv von *to race*, hier soviel wie rennen, jagen bedeutet. Die Struktur wird entsprechend aufgebaut mit *raced* in der Position des finiten Hauptverbs. Damit ist dann

**Hartnäckige
Holwege**

aber die final auftauchende Wortform *fell* (*stürzte*) grammatisch nicht vereinbar. Die intuitive Bestätigung für den Anfangsteil der Analyse scheint nun beim Muttersprachler so stark, dass die Mehrdeutigkeit von *raced*, das auch Partizip Perfekt (gejagt) sein kann, kognitiv nicht realisiert wird. In dieser (Partizip-Perfekt-) Lesart bedeutet der Satz soviel wie *Das Pferd, das am Stall entlang gejagt wurde, stürzte*.

Die Ursache für die Schwierigkeiten bei Sätzen wie (5–27) wurde in ihrem syntaktischen Bau gesehen, in der frühen **ungewöhnlichen Mehrdeutigkeit** und der spät auftauchenden vereindeutigenden Information.

Bemerkenswert ist nun die Beobachtung, dass dieselbe Konstruktion mit anderen Wörtern weniger Analyseprobleme bereitet, wie etwa Satz (5–28).

(5–28) *The land mine buried in the sand exploded.*

Die Landmine, die im Sand vergraben war, explodierte.

Auf die entsprechenden Untersuchungen referieren Tanenhaus/Spivey-Knowlton/Hanna (2000). Die Erklärung für das Ausbleiben des Holzweg-Effekts bei Sätzen wie (5–28) wird darin gesehen, dass der Parsing-Prozess nicht nur von den syntaktischen Informationen der Wortformen bestimmt wird, sondern auch von außersyntaktischen; das ist in (5–28) eben dem Weltwissen, dass *the land mine* die Rolle des Aktors von *buried* erfahrungsgemäß nicht so gut erfüllt wie die der Thema-Rolle, ein Umstand, der vom Verstehenssystem bei der syntaktischen Analyse dazu ausgenutzt wird, *the land mine* gleich als Subjekt eines uneingeleiteten passivischen Relativsatzes (*die im Sand vergraben war*) zu analysieren und nicht als Subjekt eines Aktivsatzes mit der finiten Form *buried* als Hauptverb. Eine Landmine ist eben eher bekannt als etwas, was vergraben wird und nicht als etwas, was selbst vergräbt. Damit ist die Position des Hauptverbs noch offen und wird dann ohne Schwierigkeit der finalen Form *exploded* zugewiesen.

Dieselbe Theorie würde auch den Holzweg-Effekt bei (5–27) erklären. Ein Pferd ist eben eher etwas, das selbst rennt, als dass es gejagt wird. Unter Berücksichtigung dieses Wissens wird es in erster Linie als Akteur und Subjekt eines aktivischen Satzes mit *raced* als Hauptverb analysiert.

Fazit: Zur Erkennung der Satzstruktur wird syntaktisches und außersyntaktisches Wissen eingesetzt; an außersyntaktischem ist dies nachgewiesenermaßen satzintern erschließbares semantisches Wissen und über den Satz hinaus Kontextwissen. Eine Bestätigung für diese Feststellung wird von der Beobachtung erbracht, dass sich die verschiedenen Wissensinhalte offenbar auf verschiedene syntaktische Analyseoperationen unterschiedlich auswirken, je nachdem wie informationshaltig die Ausdrucksmittel für die Auflösung anaphorischer und Theta-Rollen-Mehrdeutigkeit in der jeweiligen Sprache sind (vgl. Hemforth/Konieczny/Scheepers 2000).

5.3.5 | Strategien der Satzanalyse

Vor dem Hintergrund all dessen, was bisher über den Parsing-Vorgang dargestellt worden ist, die Aufgabe, den Input, die Wissensressourcen, kann man sich leicht vorstellen, dass die zeitliche Koordination der einzelnen Informationsverarbeitungsvorgänge keine triviale Angelegenheit ist. Die grundsätzlichen Alternativen werden von zwei theoretischen Entscheidungen bestimmt, nämlich davon, wie der Parser bei struktureller Mehrdeutigkeit vorgeht und davon, wann nicht-syntaktisches Wissen den Parsing-Prozess unterstützt.

Deterministische vs. probabilistische Arbeitsweise: Gelangt der Parser an einen Punkt, an dem ein Wort im Input mehr als eine mögliche Analyse erlaubt, also mehrere alternative syntaktische (Teil-)Strukturen entstehen würden, kommen zwei Arbeitsweisen in Frage. Das System kann sich für eine Alternative entscheiden und diese unter Registrierung des Zustandes, in dem die Entscheidung getroffen worden ist, weiterverfolgen – unter günstigen Umständen bis zu einer erfolgreichen Analyse. Da eine erfolgreiche Analyse nicht ausschließt, dass der Satz endgültig mehrdeutig ist, muss danach geprüft werden, ob noch eine weitere erfolgreiche Analyse möglich ist. Dazu muss der Zustand zum Zeitpunkt der Alternativenwahl wieder (rückwärts) aufgefunden werden und die Alternative ebenfalls berechnet werden. Die zweite Möglichkeit ist, im Falle mehrerer möglicher Analysen alle gleichzeitig weiter zu verfolgen, gegebenenfalls, bis mehrere erfolgreich abgeschlossen sind, was nur bei einem endgültig mehrdeutigen Satz der Fall sein wird. Bei Vorliegen eines temporär mehrdeutigen Satzes wird bei Auftauchen disambiguierender Information die Analysealternative, die gemäß der Grammatik damit nicht vereinbar ist, abgebrochen.

Beide Strategien, wie sie hier dargestellt sind, haben gemeinsam, dass bei jedem Schritt eine Ja/Nein-Entscheidung im Einklang mit den Syntaxregeln getroffen wird. Sie heißen deshalb beide **deterministisch**.

Nun kann man sich natürlich einen Parser vorstellen, in dem mehrere Verknüpfungsmöglichkeiten zwischen zwei Kategorien danach unterschieden sind, wie **wahrscheinlich** in einem gegebenen Fall ihre Verwendung in der Sprache ist. Systeme mit nicht-deterministisch bewerteten Verknüpfungen heißen **probabilistisch**. Über die Eigenschaften und die Arbeitsweise von Parsern als mathematische Algorithmen informieren Naumann/Langer (1994), über die psycholinguistischen Anwendungen Hemforth/Konieczny (2007).

Zeitpunkt und Art des an der Analyse beteiligten Wissens: Die zweite theoretische Entscheidung betrifft die Frage, welches Wissen am syntaktischen Analyseprozess zu welchem Zeitpunkt beteiligt ist. Ein System, das zu jedem Zeitpunkt alle syntaktischen Alternativen in die Berechnung einbezieht und keine auf Grund von außersyntaktischem Wissen ausschließt, heißt **uneingeschränkt** (*unrestricted*).

Werden hingegen syntaktisch gegebene Optionen unter Berücksichtigung nicht-syntaktischer Informationen ausgeschlossen, heißt das System **eingeschränkt** (*constraint-based parsing*). Die Einschränkung kann wiederum als Ja/Nein-Entscheidung getroffen werden oder als Gewichtung, also deterministisch oder probabilistisch.

Parsing und
Wahrscheinlichkeit

Seriell
oder interaktiv?

Die Rolle von Erwartungen beim Satzverstehen: Die Geschwindigkeit, mit der der Parser die Satzstruktur erkennt, ist erklärungsbedürftig. Ein Anhaltspunkt wird darin gesehen, dass das Parsing-System nicht strikt inkrementell arbeitet. Wie soll man das verstehen? Es gibt Beobachtungen, die die Annahme stützen, dass der Parser aus dem gehörten Input mehr Information ableitet als lediglich die, die für die Einfügung des bisher gehörten Segments in die Satzstruktur erforderlich ist. Unter gewissen Umständen werden zu einem gegebenen Zeitpunkt des Vorgangs zusätzlich sozusagen Vorhersagen über die weitere Satzstruktur aktiviert. Liegen bestimmte Teilstrukturen fest, aktiviert das System Erwartungen über die Struktureigenschaften des noch nicht gehörten bzw. gelesenen Segments. Man spricht von *prediction* und entsprechend von *predictive parsing*.

Schließlich werden Analysestrategien danach unterschieden, ob die mit dem Input gegebenen Informationen, also lautliche, syntaktische, semantische und kontextuelle, pro Analyseeinheit zeitlich hintereinander (**seriell**) oder ohne zeitliche Einschränkungen bei einem Analyseschritt gemeinsam maximal ausgenutzt werden (**parallel und interaktiv**).

5.3.6 | Modelle

Parsing-Prinzipien

Das Modell von Mitchell P. Marcus: Das bekannteste Symbolverarbeitungsmodell ist der sog. **D-Parser** von Mitchell P. Marcus (1980), nach seinem Autor häufig und respektvoll auch Marcus-Parser genannt. Das ist ein deterministisches, seriell, uneingeschränktes Modell der menschlichen syntaktischen Analyse. Seine Leistung entspricht allerdings nicht der des menschlichen Parsers, wie der Holzweg-Effekt zwingend belegt. In Annäherung an die Verhältnisse im kognitiven Analysesystem des Menschen wird im D-Parser ein Speicherraum angenommen (*buffer*), der auf 3 Einheiten begrenzt ist. Gelangt die Analyse trotz Auswertung der Informationen in diesem *buffer* an einen Punkt, an dem die neu ankommende Wortform mit keiner Regel mehr in die bis dahin aufgebaute Analyse eingefügt werden kann, bricht die Analyse ab und stellt auf diese Weise den Holzweg-Effekt dar. Details der Formalisierung, Implementierung und Arbeitsweise eines seriellen deterministischen Parsers (und anderer) stellt Hellwig (1989a und 1989b) dar.

Das garden-path-Modell: Im Wesentlichen gleich (deterministisch, seriell, schwach semantisch eingeschränkt) arbeitet das Analysemodell von Frazier (1979). In ihm wird angenommen, dass jede Wortform strukturell sofort in die Konstituentenstruktur eingefügt wird, bei Mehrdeutigkeit unter Verfolgung von gewissen Strategien. Die beiden wirkungsstärksten Strategien sind die **Minimal-Attachment**-Strategie und die **Late-Closure**-Strategie.

Minimal-Attachment schreibt vor, bei Vorliegen mehrerer Verknüpfungsmöglichkeiten diejenige zu bevorzugen, die im Rahmen der gegebenen Grammatik minimal ist, gemessen an der Zahl von zusätzlichen Verzweigungen, die in die Struktur aufgenommen werden müssen, um die aktuelle Konstituente einzufügen. Damit wird z. B. erklärt, dass bei der Analyse von *The horse raced past the barn fell* die Hauptsatzanalyse (*raced*

als finites Hauptverb) der ›Uneingeleiteter-Nebensatz-Analyse‹ vorgezogen wird. Letztere würde nämlich das Einfügen einer zusätzlichen NP-Verzweigung erfordern, erstere nicht.

Late-Closure schreibt vor, neu ankommende Wortformen so lange wie mit dem Syntaxwissen des Parsers vereinbar in die aktuell bearbeitete Phrase einzufügen. Diese Strategie erklärt die Neigung des Parsing-Systems zur sog. tiefen Verknüpfung (**low attachment**); eine Beispielstruktur ist (5–29).

(5–29) *While Mary is mending the sock / (it) fell off her lap.*

The sock kann in dem Satz ohne *it* Objekt von *mending* sein oder Subjekt des Hauptsatzes. *Late Closure* schreibt die Objektanalyse vor, weil sie noch möglich ist und damit die VP weiter ausgebaut wird. Sie stellt sich dann als unzutreffend heraus, was die Verstehensschwierigkeit solcher Sätze erklärt. Wirkt sich die Anwendung beider Strategien widersprüchlich aus, hat *Minimal Attachment* Vorrang vor *Late Closure*.

Der *garden-path*-Effekt wird damit erklärt, dass das Analysesystem in zwei Schritten arbeitet. Im ersten Schritt wird von einem Teilsystem des Parsers eine Folge von Wortformen zu Phrasen verknüpft, die aber zunächst unverbunden bleiben. Im zweiten Schritt verbindet ein anderes System die phrasalen Kategorien zu einem Satzganzen, wobei auch nicht-syntaktisches Wissen herangezogen wird. Das schrittweise Analyseverfahren wird als ›Zwei-Schritt-Analyse‹ bezeichnet (*two-step-analysis*).

Hat nun infolge von *Minimal Attachment* und *Late Closure* eine unzutreffende erste Analyse stattgefunden, so wird deren Ergebnis an den zweiten Schritt weitergegeben, sofern die disambiguierende Information nicht so nah ist, dass sie noch in den Speicherinhalt des Arbeitsspeichers passt; dessen Kapazität ist auf ca. 5 Einheiten begrenzt. Liegt die eindeutige Information außerhalb dieser Länge von Wortfolgen, kommt es zu einer falschen Satzgliedanalyse und der Fehler wird zu spät erkannt; eine stillschweigende, glatte Reanalyse ist nicht mehr möglich. Die Analyse wird abgebrochen.

Constraint-based-Parser: Das bekannteste Beispiel für eine hinsichtlich aller wesentlichen Merkmale alternative Arbeitsweise des Parsers ist ein probabilistisches, vielfach eingeschränktes, interaktives und paralleles Modell (Trueswell/Tanenhaus 1994; MacDonald/Pearlmutter/Seidenberg 1994). Es ist nicht-deterministisch, d. h. mehrere Analyseentscheidungen konkurrieren miteinander und die insgesamt wahrscheinlichste kommt zum Zug (probabilistisches Parsing). Bei einer Analyseentscheidung werden immer alle syntaktischen Alternativen berücksichtigt. Die Zahl der in Betracht gezogenen Alternativen ist allerdings durch Auswertung lexikalischer und kontextueller Informationen sowie Informationen über die Vorkommenshäufigkeit eingeschränkt (*constraints*). Das System arbeitet **interaktiv, also nicht seriell**, indem eben Informationen von sog. höheren Wissens Ebenen in die Analyseentscheidungen eingehen und zu keinem Zeitpunkt ausschließlich syntaktische Informationen berücksichtigt werden. Es ist schließlich **parallelverarbeitend**, weil bei mehrdeutiger In-

Alternative
Modelle

formationslage das Parsing-System alle strukturellen Alternativen gleichzeitig bearbeitet und beurteilt.

Beim Auftreten mehrerer Analysealternativen, z. B. a) reduzierter Relativsatz und b) Hauptsatz wie in (5–27), wird die Entscheidung für die eine oder die andere Struktur durch die Berechnung von Wahrscheinlichkeiten beeinflusst. Diese ergeben sich aus den theoretisch angenommenen *Constraints*. Im gegebenen Fall wirken auf die Analyseentscheidung u. a. der Umstand, wie wahrscheinlich es im Allgemeinen ist, dass das erste Nomen im Satz die Agens-Rolle des Verbs einnimmt oder eine andere thematische Rolle. Auch, dass z. B. *the horse* Agens von *race* ist, hat für sich genommen eine feststellbare Wahrscheinlichkeit, und die ist anders als die, dass es *Patiens* von *race* ist. Tanenhaus et al. (2000) nennen diesen Umstand den **Thematic Bias**.

Ein zweiter *Constraint* für die Analyseentscheidung ist der Umstand, wie wahrscheinlich generell eine Hauptsatzalternative gegenüber einer Nicht-Hauptsatzalternative ist. Ein dritter *Constraint* ist die Häufigkeit der Verbform *raced* als finite und als infinite Form usw. Je nach konkretem Modell wird eine unterschiedliche Anzahl von *Constraints* angenommen.

Entscheidend ist, dass alle *Constraints* bei jedem Schritt der Analyse zur Berechnung der Gesamtwahrscheinlichkeit für die eine oder andere Analyseentscheidung beitragen und zwar laufend neu und möglicherweise anders beim Beginn des Analysevorgangs, als beim Erkennen der Verbform, und wieder anders als beim Erkennen der PP usw. Je ähnlicher sich die Präferenzen für die beiden Alternativen sind, desto länger dauert die Entscheidung für die eine oder die andere Struktur.

Ein funktional-neuroanatomischen Modell: Der Ablauf eines Teils dieser Schritte einschließlich der vorangehenden Laut- und Worterkennung ist in einem funktional-neuroanatomischen Modell von Friederici/Cramon (1999, S. 339) rekonstruiert. Es macht über die Reihenfolge hinaus auch Aussagen über die absoluten Zeitspannen (ZEIT-Achse) und die kortikalen Areale der Funktionen unter Bezug auf die Arealbezeichnungen nach Brodmann (BA).

Evidenz

Welche Daten stützen alle diese Behauptungen? Dass der Parser online arbeitet, ist schon früh erkannt worden, z. B. durch Reaktionszeitdaten (Frazier 1979), Eye-Tracking-Daten (Eberhard/Spivey-Knowlton/Seidivy/Tanenhaus 1995) und EEG-Ableitungen (Garnsey/Tanenhaus/Chapman 1989). Die Befunde sind vielfach bestätigt und durch neue Ergebnisse präzisiert worden. Diejenigen, die dem Leipziger funktionalen neuroanatomischen Modell (s. Abb. 5.7) zugrunde liegen, sind in Friederici/Cramon (1999) neben vielen weiteren erwähnt. Die Unabhängigkeit der syntaktischen Analyse bei nicht-ambigen Sätzen ergab sich schon aus Eye-Tracking-Daten von Rayner/Carlson/Frazier (1983) und bestätigt sich im ERP-Experiment von Hagoort (2003).

Bisher stellt sich Parsing als eine Aktivität dar, die ausschließlich auf gegebenen Input reagiert. Schon Alltagsbeobachtungen zeigen aber, dass, wer eine Äußerung hört oder liest, zeitgleich mit der Verarbeitung des Inputs automatisch strukturierte Erwartungen über die noch folgende Teilkette anstellt. Wie erklärt sich das? Es wurde schon auf die Alltagsbeobachtung hingewiesen, dass wir als Hörer in der Lage sind, beim Stocken

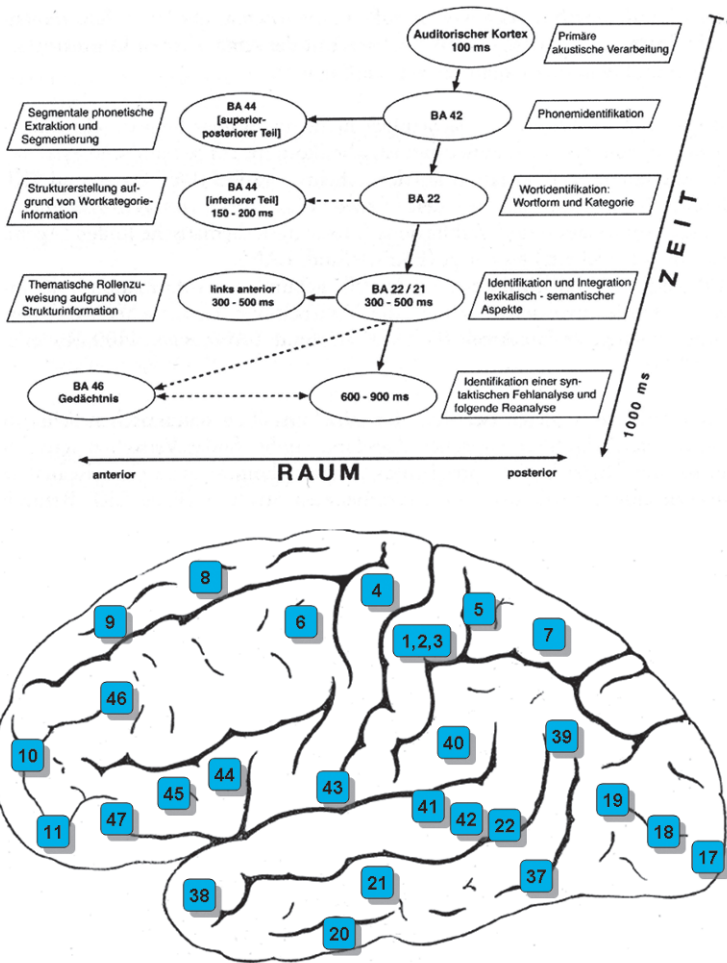


Abb. 5.7: Neurokognitives Modell des Sprachverstehens nach Friederici/Cramon (1999; Abb. 2, S. 339; wiedergegeben mit freundlicher Genehmigung des Hogrefe Verlags); darunter zur Erläuterung die Brodmann-Areale, linke Hemisphäre

des Sprechers mitten im Satz die wahrscheinlich intendierte Fortsetzung seiner Äußerung ohne großes Zögern selbst zu produzieren. Das Phänomen, dass, der Verstehensvorgang laufend von Aktivitäten des Produktionssystems begleitet wird, ist auch experimentell gut bestätigt. So haben Iijima/Sakai (2014) in mehreren Experimenten, bei denen die Magnetenzephalographie (MEG) eingesetzt wurde, nachgewiesen, dass Probanden japanische Sätze unterschiedlich schnell verstanden haben, je nachdem, ob der aktuelle Input förderliche oder weniger förderliche Merkmale für die vorangehende Bildung von Fortsetzungserwartungen enthielt.

5.4 | Verstehen der Satzbedeutung

Die Sprache ist das Medium zur Teilnahme an der Kommunikation und der Dichtung. Bei der Wahrnehmung dieser Funktion im Alltag sind wir uns der Umstände der besonderen Beziehung der Lautform und der syntaktischen Gliederung einer Äußerung zu einer mentalen Repräsentation eines Sachverhalts, also der Bedeutung dieser Form, normalerweise nicht bewusst. Erst angesichts eines uns unbekannten Wortes werden wir explizit auf die Bedeutungsseite eines Ausdrucks aufmerksam; gelegentlich auch beim Kreuzworträtsel-Lösen oder, wenn wir auf der Wissenschaftsseite unserer Tageszeitung erfahren, dass man als erwachsener Sprachbenutzer ca. 30.000 Wörter mit Form und Bedeutung in seinem sprachlichen Wissen gespeichert hat. Kaum gehen unsere Überlegungen darüber hinaus, z. B. zu der Frage, wie wir dieses sprachliche Wissen einsetzen, um die Bedeutung eines Satzes oder den Inhalt eines ganzen Romans oder auch nur eines Witzes der Folge der Wörter im Satz bzw. dem Aufbau eines Textes entnehmen.

Sprachliches und
außersprachliches
Wissen

An den Schritten des Bedeutungsverstehens sind zwei Arten von Wissen beteiligt: Lexikalisch/konzeptuelles Wissen (s. Kap. 2) und prozedurales Wissen. In Hinblick auf das Verstehen der Satzbedeutung ist unter letzterem ein Repertoire von Funktionen zu verstehen, das dafür sorgt, phonologisch und syntaktisch identifizierte lexikalische Einheiten eines Satzes zunächst mit der semantischen Information im mentalen Lexikon zu verbinden und anschließend in die semantische Struktur der jeweils komplexeren Einheit Phrase, Teilsatz oder Satz, zu integrieren (**postlexikalische Integration**). Das Ergebnis einer erfolgreichen Integration ist zunächst eine mentale Repräsentation der Satzbedeutung und dann ein mentales Abbild der Message, die vom Sprecher kommuniziert werden sollte, inklusive seiner Intention.

Prozedurales Wissen

Da sich die Bedeutung eines Satzes daraus ergibt, welche Wörter in dem Satz vorkommen und wie diese syntaktisch miteinander verbunden sind, muss das Bedeutungsverstehenssystem offenbar über Regeln verfügen, die diese Abbildung automatisch herstellen. Angenommen wird seit Chomsky (1981) neben der phonologischen und der syntaktischen Struktur eines Satzes eine sogenannte **Logische Form**. Sie ergibt sich nach Regeln, die die Konstituentenstruktur eines Satzes (Segmente und Funktionen), die Thetarollen der nominalen Phrasen und der Subjekts- und Objektsätze zusammen mit den Bedeutungen der Wörter in einen Ausdruck einer logischen Formensprache übersetzen. Das Ergebnis der Prozedur ist also ein logischer Ausdruck dessen lexikalische Segmente die im mentalen Lexikon gespeicherte Bedeutung aktivieren.

Darüber, wie genau der Anteil von Bedeutung verstanden wird, der über die logische Form eines gegebenen Satzes hinausgeht, ist momentan kaum stichhaltige empirische Evidenz verfügbar. Dass da aber noch ein »Mehr« an

Bedeutung ist, das über die bloße Bedeutung der Wörter eines Satzes und ihrer Verbindung hinausreicht, wird im nächsten Abschnitt deutlich.

Lexikalisch/konzeptuelles Wissen

Die Bedeutungen von Wörtern im mentalen Lexikon sind nicht alle von gleicher inhaltlicher »Beschaffenheit«. *Haus, Katze, Schreibtisch* verbinden sich mit der Vorstellung von Objekten, *laufen, trinken, schlafen* mit Konzepten von Tätigkeiten, *groß, einsam, arm* mit Konzepten von Eigenschaften von Objekten. Wiederum davon verschieden sind die Bedeutungen von *in, zwischen, war* gegenüber *ist*. Neurolinguistische Evidenz für die Existenz derartiger Kategorien im mentalen Lexikon liefern u. a. Ghio/Tetamanti (2010).

Modalitäten von Bedeutungswissen: In der jüngeren neurolinguistischen Forschung wurde gezeigt, dass neben den Informationen der ontologischen Kategorisierung nach Objekten, Eigenschaften, Zuständen, Vorgängen, Handlungen und Beziehungen im mentalen Lexikon auch Spuren von Auswirkungen der Wege enthalten sind, auf denen Bedeutungen von Wörtern erworben wurden, Dilkina/Ralph/Matthew (2013) bezeichnen diese als Ergebnisse von Erwerbsmodalitäten; solche sind die **Wahrnehmung** (visuell, auditiv, taktil), auf dem Weg über die **Funktionalität**, d. h. aus Erfahrungen im Gebrauch und in der Interaktion, drittens »**enzyklopädisch**«, d. h. aus der Erfahrung des räumlichen und zeitlichen Vorkommens und schließlich »**verbal**«, d. i. aus der sprachlichen Kommunikation. Gestützt wird diese Annahme durch den Nachweis von Unterschieden in der Aktivierung verschiedener Hirnareale beim Verstehen von Wörtern.

Handlungswissen: Mitte des letzten Jahrhunderts lag der Universität Oxford die Dissertation eines amerikanischen Schülers der Philosophen Austin und Strawson namens John Rogers Searle vor; sie trug den Titel *Sense and Reference*. Sie gilt aus heutiger Sicht zusammen mit einer Schrift von Austin, *How to do things with words* (Austin 1975) als ein Grundstein der sog. Sprechakththeorie (Searle 1969). Diese Theorie gilt der Tatsache, dass mit sprachlichen Äußerungen Handlungen vollzogen werden. Prototypische sprachliche Handlungen sind Befehlen, Mitteilen, Fragen. Die Tatsache, dass wir einer Äußerung zusätzlich zu ihrer Bedeutung entnehmen können, welche Handlung damit vollzogen wird, begründet die Annahme, dass zur Kompetenz eines Sprachbenutzers auch das Wissen über den Zusammenhang von sprachlichen Ausdrücken und den damit vollziehbaren Handlungen gehört. Searle nennt diese Funktion von Äußerungen die **illokutionäre** Funktion.

Aktuelles Situationswissen: Soweit war bisher nur von der Bedeutung von Sätzen die Rede. Wird ein Satz in einer kommunikativen Situation als Äußerung verwendet, findet zusätzlich zum semantischen Verstehen noch die aktuelle Interpretation statt. Sie umfasst u. a. alle Vorgänge des referentiellen Verstehens. Das ist das Verstehen von Redeteilen, die auf kontingente Objekte, Personen, Eigenschaften und Zeitintervalle verweisen. Einschlägige Wörter sind alle mit deiktischer Bedeutung wie *er, ich, jetzt, gestern, hier, das da*.

Sprachverstehen
im situativen
Kontext

In diesen Schritt des Sprachverstehens geht das temporäre Wissen um die lokalen und globalen Umstände der jeweiligen Äußerungssituation ein, also Individualität von Personen und Objekten, das Hier und Jetzt der Kommunikation.

Ein Modell, das der Tatsache Rechnung trägt, dass man einerseits die Bedeutung eines Satzes, ferner der damit getätigten Äußerung die Bezüge auf die außersprachliche Situation UND die damit beabsichtigte Handlung entnehmen kann, ist Herbert Clarks *Using Language* (Clark 1996).

Die Interaktion von Parsing und Bedeutungsverstehen

Die Darstellung, wie aus den lexikalischen Elementen und der Syntax eines Satzes die logische Form extrahiert wird, könnte den Eindruck erwecken, die semantische Analyse schließe sich zeitlich an die syntaktische Analyse an. Tatsächlich war dies auch die Sicht der frühen Parsing-Modelle, wie des *garden-path*-Modells von Frazier (1979). Verbesserte Methoden zur Messung des zeitlichen Ablaufs des Satzverstehens lassen heute aber keinen Zweifel mehr daran, dass die syntaktische Analyse und das Verstehen der Satzbedeutung vielmehr zeitlich eng verknüpft sind.

Betrachten wir Sätze wie:

- (5-30) *Der Juwelier poliert sein Silber jeden Tag. **Er** braucht **dafür** immer drei Stunden.*
 (5-31) *Der Juwelier poliert sein Silber jeden Tag. Es glänzt dann wieder wie neu.*

Im Vergleich dazu:

- (5-32) *Der Juwelier poliert sein Silber jeden Tag. **Sie** braucht dafür immer drei Stunden.*
 (5-33) *Der Juwelier poliert sein Silber jeden Tag. **Er** glänzt dann wieder wie neu.*

Alle vier Sätze sind syntaktisch wohlgeformt. Die Sätze (5-33) und (5-34) weisen aber eine inkohärente Verwendung der anaphorischen Pronomen auf. *Sie* ist auf keine der Nominalphrasen im vorangehenden Satz zu beziehen. Ebenso ist *Er* im Folgesatz (5-34) im Kontext des Satzes nicht plausibel zu interpretieren. Zu erwarten ist nun, dass das Verstehenssystem bei der Verarbeitung der fraglichen Segmente irritiert reagiert.

Eine Messmethode ist schon oben erwähnt worden: Die Messung der elektrischen Aktivitäten der beteiligten neuronalen Module. Bei der Verarbeitung von (5-33) stößt das System auf eine syntaktische Verletzung, in (5-34) auf eine semantische Ungereimtheit, etwas salopp gesagt. In zahlreichen ERP-Studien hat sich bestätigt, dass das System auf syntaktische Probleme anders reagiert als auf semantische. Erstere, also Satz (5-33) im Vergleich zu Satz (5-31), lösen eine signifikante Verstärkung früher links anteriorer negativer (LAN) und späterer positiver Potenziale (P600) aus. Auf semantische Störungen dagegen, also (5-34) gegenüber (5-32), reagiert das System mit einer späteren negativen Kurve (N400).

Dass sich diese Methode bewährt, zeigen die Befunde zahlreicher Experimente in verschiedenen Sprachen; exemplarisch für das Englische ist u. a. (Barkley/Kluender/Kutas 2015). Die übereinstimmende Erkenntnis ist ferner, dass die Reaktion auf die Inkohärenz **spontan** und im unmittelbaren zeitlichen Zusammenhang mit dem Auftreten des kritischen Segments in der syntaktischen bzw. semantischen Analysephase erfolgt. Das ist eine überzeugende Datenlage zur Beantwortung der generellen Frage nach dem zeitlichen Zusammenspiel von syntaktischer und semantischer Verarbeitung. Die beiden Teilsysteme arbeiten simultan und interaktiv.

Bedeutungsverstehen bei anaphorischen Ausdrücken

Inwieweit das System auch auf satzinterne und satzübergreifende anaphorische Inkongruenzen verschieden reagiert, zeigen Colonna/Schimke/Hemforth (2015) anhand von Daten aus einem Fragebeantwortungs- und einem Visual World-Experiment.

Zu ergänzen ist noch, dass diese Befunde noch präzisiert werden können. Es bleibt nämlich die Frage, zu welchem Zeitpunkt genau die semantische Analyse eines Segments beginnt. Bei der Bearbeitung anaphorischer Ausdrücke kann das der Zeitpunkt bei der Erkennung eines Pronomens sein. Es könnte aber auch sein, dass das Sprachsystem aus dem zuvor verstandenen Inhalt eine Erwartung dazu erzeugt, welche syntaktischen und semantischen Eigenschaften die unmittelbar kommenden Teile der Äußerung aufweisen. Dass das Verstehenssystem diese prädiktive Kapazität hat, ist nicht mehr fraglich; offen ist allerdings noch, welche spezifischen Bedingungen von Fall zu Fall vorhersagerelevant sind und welche weniger. Ein Handbuchartikel, der besonders die neuronale Evidenz zum Ablauf von Verstehensprozessen zusammenfassend präsentiert ist Brown/Hagoort/Kutas (2000).

Zur Vertiefung

Übersicht der Verarbeitungsschritte beim Satzverstehen

Aus den vielen Einzelbefunden, die vorliegen, ergibt sich ein grobes Mosaik der Verarbeitungsschritte, die beim Verstehen eines einfachen Satzes in einem Text ablaufen.

Das Auffinden der Satzstruktur ist Satz für Satz ein konstruktiver Prozess, an dessen Beginn die Kenntnis des bisher verstandenen Textes und an dessen Ende die syntaktische Struktur steht, mit der die Wörter des neuen Satzes sich hierarchisch verbinden.

1. Das Parsing ist ein flüchtiger Prozess, in dem Information laufend aufgenommen, in Strukturfragmente integriert, kurz verfügbar gehalten wird und rasch wieder verblasst. Wie welches neue Segment zum Aufbau der Struktur beiträgt, entscheidet der Parser unter Verwertung von Information aus drei Quellen, (1) den Struktur Erwartungen, die sich aus dem vorangegangenen Schritt ergeben (*prediction*), (2) dem Input, d. h. dem aktuellen Lemma, und (3) dem lebenslang erworbenen Wissen grammatischer, semantischer und pragmatischer Strukturen sowie spezifischen Vorkommenshäufigkeiten (*Constraints*).
2. Der Informationsstand im Parser ändert sich laufend, also auch die Erwartungen, wodurch der Parser zu phasenweise unterschiedlichen Ver-

Satzverstehen
kompakt

arbeitungsprozessen am Anfang, im Innern und am Ende des Satzes veranlasst wird, auch bei syntaktisch gleichartigem Input.

3. Syntaktisch eindeutiger Input wird vom Parser anders ausgewertet als syntaktisch mehrdeutiger, erwarteter Input anders als unerwarteter.
4. Die mit einem erkannten Wort lexikalisch verbundene Information wird zeitlich sukzessiv ausgewertet.
5. Bei eindeutiger Lemma-Information führt der Parser folgende Schritte nacheinander durch:
 - Registrierung der Wortartinformation gemäß Lexikoneintrag; ggf. mehrere,
 - Registrierung der funktionalen Kategorien wie Numerus, Kasus, Genus, Tempus, und der Argumentstruktur des Wortes (Subkategorisierung),
 - Registrierung der lexikalischen Bedeutung,
 - Integration der Informationen in die aktuell bearbeitete syntaktische Phrase,
 - Zusammenfügung (Aggregation) zur Satzstruktur (*wrap-up*).
6. Syntaktisch mehrdeutiger Input wird on-line desambiguiert. Dabei werden prosodische und semantische Eigenschaften des Inputs und ›höheres‹ Wissen einbezogen.
7. Bei ungünstiger Konstellation kann die Desambiguierung sich als falsch herausstellen, z. B. bei früher lokaler Mehrdeutigkeit und später desambiguierender Information (Holzweg-Effekt).
8. Gegebenenfalls Reanalyse.
9. Ableitung der Satzbedeutung aus der syntaktischen Struktur und den lexikalischen Konzepten (Inhaltsanalyse).
10. Verknüpfung der Inhalte mit Diskurs- und Weltwissen (referenzielles Verstehen).
11. Schlussfolgerungen auf Sachverhalte, die im Diskurs nicht ausgedrückt sind (*inferencing*).
12. Zeitlich laufend interaktiv abgestimmter Zugriff auf syntaktische UND semantische Informationen beim Satzverstehen.
13. Zusätzliche Verarbeitung von Situations- und Weltwissen bei Äußerungsverstehen.

5.5 | Lesen

5.5.1 | Lesen als Kommunikationsform

Die Parameter
der schriftlichen
Sprachver-
arbeitung

Die schriftliche Kommunikation ermöglicht sprachliche Verständigung unter Bedingungen, in der mündliche Face-to-Face-Verständigung nicht möglich ist. Es sind dies alle Konstellationen, in denen den Kommunizierenden:

- die gleichzeitige Teilnahme an der Kommunikation nicht möglich ist,
- ein akustischer Kanal nicht gegeben ist, bzw. kein visueller bei Gebärdensprache,
- einem der beiden Kommunikationspartner die mündliche Kommunikation nicht zu Gebote steht – und keine gebärdensprachliche Alternative.

Dafür bringt das Schreiben zusätzlich mit sich:

- die automatische Speicherung des Textes und damit
- die Weitergabemöglichkeit,
- den einfachen Einsatz nicht-sprachlicher graphischer Zeichen (Skizze, Icons),
- die Redigierungsmöglichkeit.

Umgekehrt bringt die Bereicherung der Kommunikationsmöglichkeiten auch Einschränkungen gegenüber dem mündlichen Face-to-Face-Setting mit sich:

- kein spontaner Wechsel der Gesprächsrolle,
- kein Feed-back,
- bei fehlendem Sichtkontakt kein Einsatz nicht-sprachlicher Zeichen (Mimik, Gestik, Deixis),
- langsamere Übertragungsrate.

5.5.2 | Lesen als Prozess der Sprachverarbeitung

Der kognitive Vorgang des Lesens umfasst eine große Zahl nach Art und Funktion verschiedener Prozesse. In erster Näherung zu unterscheiden sind die **visuelle Wahrnehmung** und die **Verarbeitung des Wahrgenommenen**. Erstere richtet sich nicht nur auf sprachliche Zeichen, sondern auf die gesamte Oberfläche unserer räumlichen Umgebung, soweit sie Lichtstrahlen im Wellenbereich des Lichtspektrums von ca. 380 bis 760 nm ($1 \text{ nm} = 10 \times 10^{-9} \text{ Meter}$) aussendet oder reflektiert. Aus Platzgründen beschränken wir uns in diesem Buch auf die Leseprozesse auf Satzebene. Höhere Prozesse, zum Beispiel des Textverstehens, behandeln Kintsch/Welsch (1991).

Die physiologische Grundlage des Sehens ist das visuelle System: ›Außen‹ das bewegliche Auge; im ›Innen‹ der Sehweg und das Sehzentrum im visuellen Kortex. Jede mehrbändige Enzyklopädie beschreibt ausführlich das Sehorgan des Menschen; wir erwähnen hier also nur das für den Lesevorgang Wissenswerte.

Lesen als Sehen

Das Licht wird vom Auge durch die kreisförmige Pupille aufgenommen. Sie ist eine Öffnung in der bikonvexen Linse, wird von der Iris, der Regenbogenhaut, gebildet und kann durch deren Bewegung im Durchmesser blendenähnlich verändert werden (Adaption). Auch die Linse kann zur Schärfenregulierung an die Entfernung eines Objekts angepasst werden (Akkommodation). Im Innern des Auges wird das einfallende Licht auf die rückseitige Innenwand des Auges gelenkt. Dort trifft es auf die Netzhaut (Retina) mit ihren Millionen von lichtverarbeitenden Zellen. Die sehschärfste Region der Retina ist infolge höchster Dichte an Farbsehzellen die Fovea, im Zentrum des sog. gelben Flecks (Makula) gelegen. Sie nimmt nur einen sehr kleinen Teil der Netzhaut ein, ca. 0,02 %. Das macht es erforderlich, das Auge zum Scharfsehen immer auf den Bereich im Gesichtsfeld zu richten, dem die visuelle Aufmerksamkeit gilt.

Foveales, parafoveales und peripheres Sehen: Als Gesichtsfeld eines Auges wird der wahrnehmbare Bereich bei Geradeaussehen ohne Bewe-

gung bezeichnet; es beträgt horizontal ca. 150° , der foveale Bereich knapp 2° . Wahrnehmung ist aber auch in dem parafovealen Bereich des Gesichtsfeldes möglich, der sich an den fovealen beiderseits anschließt und rund 5° beträgt (mehr dazu unten).

Das gesamte Gesichtsfeld reicht aber noch weiter. Starke distale, d. h. äußere Reize, z. B. Objekte, von denen ein starkes Signal ausgeht (Bewegung, helle Strahlen), nimmt das menschliche Auge auch im **peripheren** Gesichtsfeld noch wahr. Das schließt sich an das parafoveale beiderseits an und reicht für das rechte Auge rechts bis zum schläfenseitigen Rand (temporaler Rand) und links bis zum nasalen Rand, beim linken Auge sinngemäß umgekehrt. Natürlich hat das Gesichtsfeld auch eine vertikale Ausdehnung; sie beträgt etwa 60° nach oben und 70° nach unten. Insgesamt hat das Gesichtsfeld eines Auges also die Form eines Ovals, das horizontal zum temporalen Rand etwas weiter reicht als zum nasalen, je nach Nasenform. Bei beidäugigem Sehen überschneiden sich die Gesichtsfelder beider Augen zentral um ca. 120° ; das gesamte beidäugige Gesichtsfeld beträgt also rund 180° . Der gesamte durch zusätzliche Augenbewegung wahrnehmbare Bereich heißt **Blickfeld**; er beträgt beim Menschen ca. 240° .

Die räumlichen Bedingungen des Lesens: Was ergibt sich daraus nun für die Dynamik des Lesevorgangs? Zunächst einmal folgt daraus, dass beim Lesevorgang immer mehr Information im Blickfeld liegt als nur die jeweils foveal fixierten Schriftzeichen. Bei der üblichen Lesehaltung von 30 bis 40 cm Entfernung vom Text und bei einer Schriftgröße von 10 pt (Punkt; 1 pt = 0.0353 cm) und entsprechender Schriftweite erfassen wir foveal drei bis vier Buchstaben, parafoveal nach rechts weitere ca. 16 und nach links ca. 4 Stellen. Die seitliche Ausrichtung der Lesespanne ist nicht physiologisch begründet. Sie wird mit dem Lesenlernen kognitiv installiert und ist bei einem Rechts-Links-Verlauf der Schrift entsprechend linksgerichtet. Die Sehschärfe nimmt nach außen hin asymptotisch bis auf ca. 5 % an den Rändern ab und beträgt bei 5° z. B. nur noch 40 %, bei 20° etwa 10 %. Beim hinten seitlich aus dem Auge austretenden Sehnerv enthält die Retina einen lichtunempfindlichen Bereich, den sog. blinden Fleck, bei etwa 15° bis 17° links (rechtes Auge) bzw. rechts (linkes Auge) der Fovea, in Blickrichtung gesehen.

Da der Text im Normalfall vor unseren Augen fest steht, müssen wir die Augen bewegen, um mehr wahrzunehmen als drei bis vier Zeichen. Dazu wird das Auge durch einen ›Sprung‹ (Sakkade) weiterbewegt. Der Landeplatz ist vorab automatisch kalkuliert, was bis zu 200 ms beanspruchen kann, während die Ausführung selbst nur etwa 30 ms dauert, was so gut wie keine Wahrnehmung ermöglicht.

Methoden

Messung der Lesezeit, der Lesebewegung und des Leseerfolgs: Das Lesen umfasst zwei Hauptprozessgruppen, die visuelle Wahrnehmung und die visuelle Verarbeitung. Die visuelle Wahrnehmung geschieht auf dem Wege der Fixation einer Buchstabenfolge und der Weiterbewegung des fovealen Gesichtsfeldes durch Augenbewegung (Sakkade). Beides kann man messen.

Als aufschlussreich für die Messung der Lesezeit haben sich die folgenden Maße erwiesen:

- Fixationsdauer, ggf. Erstfixation unterschieden von Blickdauer, d. h. die Summe der Fixationszeiten bei mehrfacher Fixation auf einem Wort
- Lesezeit des ersten Lesedurchgangs (beim Lesen längerer Einheiten)

Die Lesebewegung wird gemessen in Maßen wie:

- Zahl der Sakkaden (mittlere Sakkadenlänge 8–9 Buchstabenpositionen),
- Zahl der Regressionen (rückwärts gerichtete Sakkaden),
- die Lesezeit beim zweiten Lesen von einer interessierenden Region des Input,
- die Lesezeit für eine sprachliche Region bis zum Weiterlesen im Text (*go-past duration*),
- die Lesezeit für eine sprachliche Region vor einer Regression (*regression-path duration*).

Und schließlich lässt sich noch der Leseerfolg messen, indem man prüft, ob das Erkannte mit dem Stimulus übereinstimmt; das gängige Maß ist ›% korrekt‹.

Die Lesespanne wird mit verschiedenen Verfahren der Stimuluspräsentation gemessen. Die meist verwendeten sind die Fenstermethode und die Grenzmethode. Beide erfordern eine spezifische Geräteausstattung mit Programmen zur Augenbewegungsmessung (Eye-Tracking-System). Solche Systeme erlauben es, den Text fensterweise auf dem Bildschirm zu präsentieren (**moving window**-Technik) und die Fenstergröße dabei frei zu wählen. Auch die Bereiche außerhalb des Fensters links und rechts können auf bestimmbare Weise besetzt werden. Die englische Bezeichnung *moving window* bezieht sich auf die technische Manipulation, das Erscheinen des nächsten Fensters automatisch auszulösen, wenn der Blick durch Sakkade auf eine neue Position wandert. Mit dieser Technik lässt sich die Lesespanne ermitteln, indem man über mehrere Leseabläufe hin die Fenstergröße variiert und diejenige ermittelt, die zur besten Leseleistung führt. Umgekehrt kann auch das foveale Fenster maskiert werden und gleichzeitig das rechts benachbarte Fenster den lesbaren Text enthalten. Das Lesbare kann dann nur parafoveal wahrgenommen werden und die Leseleistung kann somit Aufschluss darüber geben, wie die Leseleistung bei parafovealer Wahrnehmung ist.

Parameter des
Lesevorgangs

Die **Grenzmethode** macht ebenfalls von dem Grundprinzip Gebrauch, das Gesichtsfeld rechts von der fixierten Region verschieden zu besetzen. Anders als bei der Fenstermethode wird aber nicht mit einem beweglichen Fenster gearbeitet. Vielmehr ist der Text einer Zeile ein Stück lang korrekt gegeben und rechts von einer vorher festgelegten Grenze manipuliert z. B. *Fritz trainiert gerade draußen im | Garten*. Überschreitet der Blick diese Grenze, wird das Zielsegment, z. B. *Garten*, sichtbar, wo zuvor lediglich *Garxxx* oder *Gxxxxx* oder *Becher* sichtbar waren. Die Fixationszeit auf diesem Segment lässt Schlussfolgerungen über die Qualität der parafovealen Wahrnehmung beim Lesen zu. Ausführlich werden diese Methoden dargestellt und erörtert in Inhoff/Rayner (1996), Rayner/

White/Kambe/Miller/Liversedge (2003) und Rayner/Liversedge/Henderson/Ferreira (2004).

Der Leseverlauf

Schriftverarbeitung: Der Input des Leseprozesses ist visuelle Information von graphischen Figuren, der Output die Aktivierung des gelesenen Wortes im mentalen Lexikon. Wer erfolgreich lesen gelernt hat, hat für jede lexikalische Einheit im mentalen Lexikon eine schriftliche Form gespeichert, ggf. eine mehrdeutige, wie z. B. < Band > für *das Band* und *die Band* (Musikgruppe) und *der Band* (Buch). Nun sind aber nicht alle Texte in nur einer und immer derselben Schrift geschrieben. Dennoch erkennen wir die **Buchstaben**, seien sie in Slimbach wie hier oder etwa in **Frutiger** geschrieben. Wir würden sie auch erkennen, wenn sie in einer uns zuvor unbekannten Schrift (z. B. rotis semi sans) geschrieben wären. Also muss die graphematische Repräsentation abstrakter sein als die von uns jeweils gerade wahrgenommene graphische Realisierung. Anders betrachtet: Wir verfügen über die Fähigkeit, graphisch verschiedene Realisierungen desselben Graphems zu erkennen.

Buchstaben: Wie geschieht das? Buchstaben – wie auch Laute – sind Zeichen, die in einigen Merkmalen übereinstimmen und in anderen nicht. Der Buchstabe ›R‹ weist wie das ›K‹ den linken senkrechten Strich auf, wie das ›B‹ den oberen rechten Halbbogen und wie das ›L‹ wiederum den unteren rechten Querstrich von oben links nach unten rechts. Die Buchstaben eines jeden Alphabets lassen sich so nach einem Satz spezifischer Merkmale analysieren. Mithin enthält das mentale Lexikon eine Merkmaleebene, eine Graphemebene und eine Wortebene. Und da die schriftliche Verarbeitung an den Merkmalen ansetzt, müssen zwischen Elementen der Ebenen Verbindungen bestehen, schematisch veranschaulicht in Abb. 5.8.

Alle Modelle nehmen an, dass die Schriftverarbeitung an diesen Merkmalen ansetzt, die erkannten Merkmale aktiviert werden und kaskadierend die Grapheme aktivieren, mit denen sie verbunden sind. Das am stärksten aktivierte Graphem wird als das gegebene erkannt und die Aktivierung

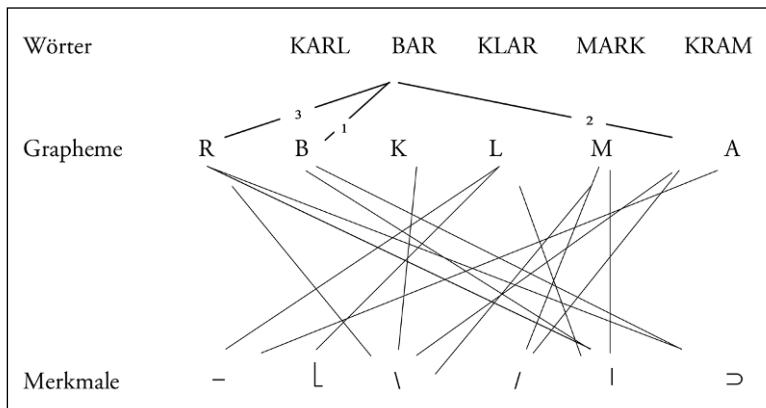


Abb. 5.8: Schema graphematischen Wissens im mentalen Lexikon (Verbindungen unvollständig)

setzt sich auf die Wortebene fort. Nicht abschließend klar ist (a) die Rolle der Position des Zielbuchstabens im Wort, (b) die Identität seiner Nachbarn, (c) die Richtung der Aktivationsausbreitung. Findet Aktivierung nur *feed-forward* statt oder interaktiv, auf allen Ebenen kaskadierend oder von der Buchstaben- zur Wortebene nicht mehr kaskadierend?

Dass es hier noch viele offenen Fragen gibt, geht nicht auf widersprüchliche, sondern auf unterschiedliche Daten zurück. Das System arbeitet bei Einzelbuchstabenerkennung anders als bei der Erkennung von Buchstaben im Wort, bei tachistoskopischer, d. h. sehr schneller Präsentation des Stimulus anders als bei normaler Lesegeschwindigkeit, beim lauten Lesen anders als beim Korrekturlesen.

Gründliche Überblicksdarstellungen über diese und mehr Details der Schriftverarbeitung sind Balota (1994) und Lupker/Snowling/Hulme (2005). Es wird in diesem Abschnitt übrigens nicht sorgfältig zwischen ›Buchstabe‹ und ›Graphem‹ unterschieden (vgl. Fuhrhop/Peters 2013).

Visuelle Worterkennung: Die nächste spannende Frage ist, wie das Lesesystem das Zielwort erkennt. In Kapitel 2.2.1 ist der Wortüberlegenheitseffekt erwähnt, der darin besteht, dass die Erkennungszeit für ein Wort kürzer ist als die Summe der Erkennungszeiten für die Grapheme, aus denen es zusammengesetzt ist. Evidenz dafür bildeten die Ergebnisse des Experiments von Reicher (1969). Die Stimuli in dem Reicher-Experiment waren vier Buchstaben lang (*Maus*, *Maul* etc.). Das entspricht etwa der fovealen Lesespanne, und diese Feststellung wirft natürlich die Frage auf, ob die Worterkennung nicht auch von der Wortlänge abhängt und von welchen Faktoren sonst noch.

Im Zentrum der Erklärungen stehen die folgenden Fakten:

1. Der Wortüberlegenheitseffekt wird damit erklärt, dass ein Wort infolge seiner graphischen Kontur als ganze Form und somit schneller erkannt wird als die Reihe der Buchstaben von links nach rechts; das führt zu einer vergleichsweise schnelleren Aktivierung der Wortbedeutung, was die Reicher-Daten erklärt. Für diese Theorie spricht auch die Tatsache, dass durchgehend klein geschriebene Wörter besser verarbeitet werden als Wörter in Großbuchstaben. Die kleinen Buchstaben weisen durch ihre Oberlängenunterschiede besser unterscheidbare Konturen auf als die Majuskelformen derselben Wörter, die die Oberlängenunterschiede neutralisieren.

2. Der Wortlängeneffekt: Längere Wörter werden weniger schnell erkannt als kürzere (vgl. Just/Carpenter 1980).

3. Der Wortfrequenzeffekt besteht darin, dass häufige Wortformen besser bearbeitet werden als weniger häufige.

4. Der semantische Priming-Effekt besteht darin, dass bei einer lexikalischen Entscheidungsaufgabe besser geantwortet wird, wenn der Präsentation des Zielwortes diejenige eines semantisch verwandten Wortes als Prime vorangegangen ist, z. B. *WOLF* – *HUND* statt *WOLF* – *BAUM*. Der Effekt tritt auch bei einfacher Wortnennungsaufgabe ein. Der Befund spricht dafür, dass semantische Information bei der Worterkennung mitverarbeitet wird, was bei den Non-Wort-Items nicht möglich ist. Der Effekt

Lesen:
Leicht oder
schwierig?

stärkt die Annahme, dass bei der Worterkennung interaktive Aktivierung stattfindet.

5. Der Priming-Effekt bei maskierter Wiederholung besteht darin, dass ein Zielwort besser bearbeitet wird, wenn ihm in der Präsentation unmittelbar und kurz dasselbe Wort – allerdings in einer anderen Schreibweise – vorangeht, also *hund* – *HUND*. Auch hier ist die Erklärung, dass der Prime die Bedeutung voraktiviert und so – durch Top-Down-Aktivierung – auch das Zielwort schon voraktiviert ist.

6. Der Mehrdeutigkeitseffekt besteht darin, dass mehrdeutige Wörter schneller erkannt werden als eindeutige. Dieser Effekt wird damit erklärt, dass das lexikalische Wissen eine Netzwerkstruktur aufweist und dass Aktivierung in diesem Netz zwischen den Knoten in beide Richtungen fließen kann (Feedback). Bei der Wahrnehmung einer mehrdeutigen Wortform werden mehrere Konzepte aktiviert und aktivieren per Feedback stärker die bearbeitete Wortform, was zum Zeitpunkt des Erkennens zu einer schnelleren Reaktion führt; man beachte allerdings die Verzögerungswirkung der Mehrdeutigkeit beim Worterkennen im Kontext.

Modelle

Diese und weitere Befunde haben zu unterschiedlichen Modellannahmen geführt. Die beiden großen Gruppen von Modellen sind Suchmodelle und Aktivationsmodelle.

Suchmodelle erklären den Worterkennungsvorgang grundsätzlich als einen Prozess, in dem eine außerlexikalische Suchform mit den Einträgen des mentalen Lexikons verglichen wird, bis es entweder zu einer Übereinstimmung kommt oder die Suche ohne Erfolg beendet ist (Non-Wort). Die erwähnten Befunde werden im Suchmodell dadurch erklärt, dass die Suchgeschwindigkeit von Anordnungen des durchsuchten Materials im Lexikon abhängt, die die Suche leiten. Solche Ordnungsparameter sind Länge, Frequenz, Bedeutungsähnlichkeit etc.

Aktivationsmodelle modellieren die Worterkennung als einen autonomen kaskadierenden und – je nach Modell – interaktiven Aktivationsfluss in der mehrdimensionalen Netzstruktur des mentalen Lexikons. Eines der besonders ausgearbeiteten Lesemodelle ist das *Parallel Distributed Processing*-Modell (PDP) von Seidenberg/McClelland (1989). Es unterscheidet sich von anderen Aktivationsmodellen, da im Lexikon keine eigenen Repräsentationen von Graphemen, Wörtern und Konzepten angenommen werden, sondern subsymbolische »Merkmale«, die so miteinander verknüpft sind, wie es eben dem lexikalischen Wissen entspricht. Die Erklärungsstärke des PDP-Modells liegt darin, dass es Zeitunterschiede des Suchens präziser vorhersagen kann.

Es ist allerdings schwierig, die Modelle letztlich zu bewerten, weil es kein Modell gibt, das alle Fakten am besten erklärt. Nehmen wir etwa den Mehrdeutigkeitseffekt. Er wird durch die reiche Aktivationsausbreitung in einem Aktivationsmodell mit Feed-back besser erklärt als in einem Suchmodell. Ein solches erklärt aber den Wortüberlegenheitseffekt besser als ein PDP-Modell, in dem ja die Wortform als ganze gar nicht repräsentiert ist und mithin auch nicht Ziel einer gesamtheitlichen Erkennung eines

Wortformschemas sein kann. Umgekehrt wiederum erklären alle Suchmodelle die Fakten weniger plausibel, die auf eine zweiseitige Aktivationsausbreitung zurückzuführen sind. Wie kann man hier weiterkommen?

Besonders aussichtsreich erscheint die rechnergestützte Simulation. Auch Lesemodelle sind formal so ausgearbeitet worden, dass sie als leistungsfähige Computersimulationen programmiert werden konnten. Indem ein Simulator eben das menschliche Leseverhalten simuliert, kann mit ihm im Prinzip jede Leseaufgabe unter verschiedenen Bedingungen berechnet und das Ergebnis mit dem menschlichen Leseverhalten verglichen werden, soweit dazu Daten aus psycholinguistischen Experimenten vorliegen (vgl. Kliegl/Grabner/Rolfs/Engbert 2004).

Besonders komplexe und leistungsfähige Simulationen verschiedener Ausschnitte des Leseprozesses sind: SWIFT (Engbert/Longtin/Kliegl 2002; Engbert/Nuthmann/Richter/Kliegl 2005), EZ Reader (Reichle/Rayner/Pollatsek 1999; 2003) und das *Dual-Route-Cascaded-Modell* (DRC) von Coltheart und Mitarbeitern (Coltheart/Rastle/Perry/Langdon/Ziegler 2001). Letzteres beinhaltet auch eine Simulation des lauten Lesens, also der Verbindung von schriftlicher und lautlicher Verarbeitung.

Kontext und parafoveale Wahrnehmung beim Lesen

Kontext: In der Mehrzahl der Fälle lesen wir nicht Einzelwörter, sondern einen Text, der eine oder mehrere Zeilen lang ist. Unter dieser Bedingung treten zu den Faktoren, die die Einzelworterkennung bestimmen, Umstände hinzu, die die Worterkennung wesentlich mitbeeinflussen. Das sind der Faktor ›Kontext‹ und die Lesespanne in der parafovealen Wahrnehmung. Die Lesebewegung setzt sich aus den beiden Ereignissen ›Fixation‹ und ›Sakkade‹ zusammen, und daraus leiten sich die zwei Hauptfragen der Leseforschung ab: Wie lange ist die Fixation und wohin wird gesprungen? Diese lassen sich auch in einem Satz formulieren: Wann und wohin erfolgt die Sakkade?

Über das ›Wann‹, also die Dauer der Fixation eines Wortes, und über die determinierenden Faktoren wurde oben schon berichtet: Wortlänge, Frequenz etc. Nachzutragen ist allerdings, dass nicht alle Faktoren der Worterkennung sich im Kontext ebenso auswirken wie in der Einzelworterkennung. Markant verändert ist die Wirkung der Mehrdeutigkeit. Während ambige Wörter einzeln schneller erkannt werden als eindeutige, nimmt ihre Verarbeitung in einem Kontext mehr Zeit in Anspruch als das Erkennen von eindeutigen Wörtern.

Lesen und Blickfeld

Einfluss des Kontexts auf Lesezeiten

Experiment

Eine Untersuchung von Rayner/Duffy (1986) galt dem Frequenzeffekt, der semantischen Komplexität und der Mehrdeutigkeit. Im hier interessierenden Experiment lesen Versuchspersonen (a) mehrdeutige und (b) in Länge und Frequenz vergleichbare eindeutige Substantive im Kontext eines zweiteiligen Satzes. Dessen erster Teil endet mit dem experimentellen Wort und der zweite Teil enthält die Information, die die Mehrdeutigkeit auflöst, sofern eine gegeben ist.

Die mehrdeutigen, hier nur zweideutigen Wörter, unterschieden sich außerdem in der Balanciertheit der Mehrdeutigkeit; es gibt (a1) ausgeglichene und (a2) nicht ausgeglichene mehrdeutige Nomina. Bei den ersteren sind die beiden Bedeutungen etwa gleich häufig, die nicht ausgeglichenen haben eine klar dominante Bedeutung und eine nicht dominante. Die theoretischen Annahmen waren, dass bei Wahrnehmung eines mehrdeutigen Wortes die Verarbeitung länger dauert als bei eindeutigen, die Verarbeitung eines ambigen Wortes mit einer dominanten Bedeutung (a2) kürzer dauert als die eines ausgeglichenen (a1).

Die Dauer wurde abgeleitet aus den Lesezeiten des zweiten Satzteils. Die Ergebnisse bestätigen die Annahme. Die durchschnittlichen Erstfixationszeiten der Zielwörter unter den verschiedenen Bedingungen (a1), (a2) und (b) sind nicht signifikant verschieden, aber die Blickzeiten auf den zweiten Satzteil für ausgeglichen ambige sind signifikant länger als für eindeutige und für nicht-ausgeglichen ambige.

Diese Studie veranschaulicht, dass die Wirkung eines wortinternen Faktors (hier die Mehrdeutigkeit) sich unter dem Einfluss weiterer Faktoren (hier Kontext) stark verändern, ja sogar umkehren kann.

Lesesteuerung

Parafoveal wahrgenommene Informationen: Wie wertet das Lesesystem die Information aus, die zwar wahrgenommen, aber noch nicht fokussiert ist? Machen wir uns noch einmal klar: Die parafoveale Information geht beim Lesen von den Buchstabenfolgen aus, die in diesem Augenblick nicht fokussiert werden; sie erreichen das visuelle System also in einem Moment, in dem andere, vorangehende Information fixiert und ebenfalls dem System zum Zweck des Verstehens im weiteren Sinne zugänglich gemacht wird. Theoretisch sind zwei verschiedene Annahmen möglich.

Berechnung der nächsten Sakkade: Die eine geht davon aus, dass parafoveale Information nicht im fovealen Gesichtsfeld ist und entsprechend schlecht, das heißt langsam, aufgenommen wird. Besinnen wir uns auf die beiden Teilprozesse des Lesens ›Wahrnehmung‹ und ›Verarbeitung‹, so könnte man folgern, dass parafoveale Segmente zwar wahrgenommen, nicht aber verarbeitet werden. Wozu dann die Wahrnehmung? Eine Verwendung des Wahrgenommenen könnte darin bestehen, die Weite der nächsten Sakkade zu berechnen, so dass diese zügig nach der Fixation eingeleitet werden kann. Immerhin kostet die Berechnung bis zu 200 ms. Für diese Berechnung würde das System auswerten, wie weit das nächste Leerzeichen entfernt und wie lang die darauffolgende Buchstabenkette ist. Bei der generellen Strategie, die Sakkade so zu berechnen, dass der Landeplatz etwas vor der Mitte des Folgewortes sein soll, würde diese Information ausreichen (vgl. Inhoff/Rayner 1996).

Parallele Verarbeitung mehrerer lexikalischer Einheiten: Die alternative Annahme geht dahin, dass dem parafoveal Wahrgenommenen mehr als nur äußerliche, graphische Information entnommen wird; das müsste dann schon das Ergebnis von höherer Verarbeitung sein, also Graphem- oder Wort- oder Bedeutungserkennung. Diese Theorie impliziert, dass während einer Fixation mehrere lexikalische Einheiten gleichzeitig wahrgenommen und verarbeitet werden können. Das ist z. B. die Grund-

annahme in dem oben erwähnten SWIFT-Modell (vgl. Kliegl/Engbert 2003).

Was tatsächlich der Fall ist, lässt sich derzeit nicht mit letzter Sicherheit sagen. Sicher ist, dass kurze und sehr häufige Wörter beim Lesen unter bestimmten Bedingungen übersprungen werden. Das sind in der Regel Funktionswörter wie Artikel, Konjunktionen, Auxiliare, also solche, die zudem noch syntaktisch besser erwartbar sind als Inhaltswörter.

Es wurde allerdings auch mehrfach ein Zusammenhang zwischen dem Überspringen eines Folgewortes und der Fixationsdauer des Vorgängewortes beobachtet. Das würde einerseits auf eine parafoveale Verarbeitung hinweisen, andererseits aber auch darauf, dass sie mehr Zeit kostet und demnach vielleicht doch nicht parallel zur fovealen Verarbeitung geschieht, sondern seriell. Warum aber dann nicht eine eigene Fixation auf diesem übersprungenen Wort? Eine Erklärung wäre, dass die Lesebewegung kognitiver Kontrolle unterliegt, was sich darin zeigt, dass das System mitunter ökonomisch arbeitet, indem es infolge parafovealer Verarbeitung erkennt, dass sich durch eine etwas längere Fixation auf dem Vorgängewort eine erneute Fixation auf dem Folgewort erübrigen kann. Der aktuelle Stand der Erkenntnis zu diesen Fragen ist in Rayner/Juhasz/Pollatsek/Snowling/Hulme (2005) präsentiert und kommentiert.

Schriftliche und phonologische Verarbeitung

Die bisherige Darstellung hat ausschließlich Fakten und Modelle behandelt, die sich um das stille Lesen bewegen. Wer Lesen gelernt hat und es routiniert beherrscht, kann aber das Gelesene schnell und scheinbar ohne Anstrengung zusätzlich auch noch aussprechen (**lautes Lesen**). Damit stellt sich die Frage, auf welchem Weg von der visuellen Information auf die phonologische zugegriffen wird.

Lesen vs. Hören

Die Repräsentationsfrage: Was wir über den Inhalt und den Aufbau des mentalen Lexikons wissen, legt unmittelbar eine Vermutung darüber nahe, wie die schriftliche Form in das lexikalische Wissen integriert ist, nämlich: Zusätzlich zu dem phonologischen Wissen und ggf. auch dem prälexikalischen Silbenlexikon enthält das mentale Lexikon die oben skizzierte vernetzte Information über Buchstabenmerkmale, Buchstaben, Grapheme und die schriftliche Repräsentation von jedem orthographisch bekannten Wort. Von diesem ›graphematischen‹ Lexem gibt es eine Verbindung zu nicht-schriftlichen Bestandteilen der lexikalischen Einheit, z. B. zu ihrem Konzeptknoten. Dieser ist ja, wie beschrieben, mit dem Lemma und dem ›phonologischen‹ Lexem vernetzt. Wird das ›graphematische‹ Lexem nun beim Lesen erkannt, wird das Konzept aktiviert und über dieses die Lautform, die dann wie beim Sprechen artikulatorisch realisiert wird. Wer Lesen lernt, lernt also diese zusätzliche Vernetzung. Gibt es aber auch einen nicht-trivialen Zusammenhang zwischen dem Vorgang des Lesens und der phonologischen Repräsentation der Wörter (Lexeme) im mentalen Lexikon? Sowie die Fragen präziser werden, sind die Antworten schon weniger naheliegend:

- Welche lexikalische Ebene ist mit der schriftlichen Repräsentation verbunden?

- Wie viel von der nicht-schriftlichen Information wird beim stillen Lesen aktiviert?
- Wie viel von der lautlichen Information wird beim Sprechen aktiviert?
- Wenn es mehrere Wege der Aktivierung gibt, welcher ist der primäre – unter welchen Bedingungen?

Buchstabe-Laut-Korrespondenz: Gehen wir zunächst noch einmal von Alltagsbeobachtungen aus: Wer laut lesen kann, kann auch Buchstabenfolgen laut lesen, die er noch nicht kennt, seien es Non-Wörter oder dem Leser eben bislang unbekannte Wörter, wie vielleicht das Wort *babig* aus Carnaps berühmter Heideggerkritik. Aus der schriftlichen Form <babig> erzeugt unser Sprachproduktionssystem regulär die Lautform [ba:biç]. Dass das regulär geschieht, zeigt an, dass wir infolge des Lesenlernens über Ausspracheregeln verfügen, deren Anwendung uns bei der Aussprache unbekannter Graphemfolgen leitet – und umgekehrt beim Buchstabieren unbekannter, gehörter Wörter. Kurz: Wir verfügen über die Kenntnis der Buchstaben-Laut-Korrespondenz in unserer Sprache, die auch Sonderfälle einschließt, wie mehrgliedrige Grapheme (<sch>, <ch>, <ee> etc.).

Diese Tatsache löst die Frage aus, ob die Graphem-Phonem-Übersetzung der Standardweg des Lesens ist und wenn nicht, welche anderen Wege es noch gibt.

Zwei Wege der Worterkennung

Graphem-Phonem- und Graphem-Konzept-Verknüpfung: Wenn es, wie oben dargelegt, einen Wortüberlegenheitseffekt auch beim Lesen gibt und dieser Effekt auf die Erkennung der graphischen Kontur der Buchstabenfolge zurückgeht, dann kann er nicht durch reguläre Graphem-Phonem-Übersetzung zu Stande kommen, weil ja eben keine einzelnen Grapheme aktiviert sind. Also ist von daher mit zwei Wegen zu rechnen, dem Graphem-Phonem-Übersetzen und dem direkten Graphem-Konzept-Zugang; ein solches System wird in der Fachliteratur als Zwei-Prozess-System (*dual process model*) bezeichnet.

Welche **Evidenz** liegt zugunsten der beiden Wege vor und welcher Weg hat beim Lesen Priorität? Gehen wir zunächst der Annahme des ›Übersetzungsverfahrens‹ nach. Hier stellt sich zunächst wieder die immer gleiche Frage: Verläuft die Aktivierung von Schrift- zu Lautform unidirektional oder interaktiv? Von einem System mit interaktiver Aktivierung wird bei Lesen des Wortes <Leib> die Lautform /laip/ aktiviert und infolge von Feedback der schriftliche Eintrag <Laib>, weil er derselben Lautform zugeordnet ist. Das ist bei allen homophonen Wörtern so, z. B. <mehr> – <Meer> = /me:v/, <Kälte> – <Kelte> = /keltə/, <Lehre> – <Leere>, <bis> – <Biss>. Wird einer Versuchsperson in einer lexikalischen Entscheidungsaufgabe ein Homophon und – zur Kontrolle – ein im Übrigen vergleichbares eindeutiges schriftliches Wort, z. B. <Lampe>, präsentiert, nimmt die Entscheidung bei Homographen im Durchschnitt mehr Zeit in Anspruch.

Dieser **Homophon-Effekt** der Worterkennung wird mit dem Zeitbedarf erklärt, den die Aktivierung von zwei Graphemfolgen im Unterschied zu nur einer bei eindeutigen Formen mit sich bringt. Diese Erklärung wäre noch mit der Annahme vereinbar, dass die Aktivierung ausschließlich unidirektional – *feed forward* – verläuft: Die Buchstabenfolge aktiviert eine

Lautfolge, diese aktiviert zwei (oder mehr) schriftliche Formen und dann findet ein Vergleich zwischen diesen und dem Stimulus statt, worauf das Zielwort erkannt wird.

Der graphematische Einfluss auf die phonologische Reimerkennung:

Schon früh in der Forschungsgeschichte zum Lesen ist allerdings ein Verhalten beobachtet worden, das zu einer Revision dieses *feed-forward*-Modells zugunsten eines interaktiven Modells Anlass gab, der graphematische Einfluss auf die phonologische Reimerkennung. Versuchspersonen hören ein Wort und kurz darauf ein weiteres; sie sollen entscheiden, ob die beiden sich reimen oder nicht. Beispiele: (1) <Reh> – <Zeh>, (2) <Reh> – <Klee>, (3) <Reh> – <Ruh>. Bei Paaren vom Typ (1) und (2) wird positiv reagiert, bei (3) negativ. Interessant ist nun, dass die positive Reaktion bei (1) schneller erfolgt als bei Paaren vom Typ (2), was damit erklärt wird, dass beim Hören automatisch die schriftliche Repräsentation aktiviert wird und ggf. die positive Entscheidung wegen der Unterschiede der schriftlichen Formen verzögert, wie im Fall <Reh> – <Klee> (Seidenberg/Tanenhaus 1979). Derselbe Effekt tritt bei schriftlicher Präsentation der Stimuli auf; die schriftliche Form ist also bei der Reimkontrolle immer wirksam, auch dann, wenn sie für die Aufgabenlösung unerheblich ist.

Fazit: Das Lesesystem und das phonologische System bilden zusammen ein sog. **Zwei-Prozess-System**. Beim Lesen wird automatisch phonologische Information aktiviert und beim Hören automatisch schriftliche (Weg 1). Dieser Weg ist extrem schnell. Nach dem derzeitigen Stand der Messungen dauert die graphem-initiierte Aktivierung des phonologischen Lexems nicht länger als 30 ms. Zusätzlich zu diesem ›Übersetzungsverfahren‹ ist die schriftliche Information direkt – unter Umgehung der phonologischen Route – mit der Konzeptebene verbunden, was den Wortüberlegenheitseffekt erklärt, den semantischen Priming-Effekt und besonders auch Daten aus Untersuchungen mit Dyslexiepatienten (s. dazu Kap. 6.1). Auf die Frage, ob einer der beiden Wege im System Vorrang hat und ggf. welcher Weg das ist und wie dieser Vorrang geregelt ist, gibt es zurzeit keine klare Befundlage. In einem ausgeklügelten Experiment mit französischen Wörtern haben Ziegler/Ferrand/Jacobs/Rey/Grainger (2000) einen zeitlichen Vorsprung des direkten Weges gemessen, in einem späteren Experiment beobachteten Van Orden/Kloos/Snowling/Hulme (2005) dann allerdings Kontraevidenz.

5.6 | Sprachverstehen – ein separates Teilsystem?

Vergleicht man die Komponenten des Sprachverstehens mit dem Prozess der Sprachproduktion stellt sich die Frage, ob der Verstehensvorgang nicht von demselben System bewerkstelligt wird wie die Sprachproduktion, indem einfach die Verarbeitungsrichtung umgekehrt wird, also bottom-up statt top-down. Die Frage ist im vorigen Kapitel schon angeklungen und es ist klar geworden, dass die beiden Verarbeitungsrichtungen sich nicht vollständig überlappen können, sicherlich aber auch nicht vollständig separiert sind.

Funktionsspezifisch und demnach verschieden sind natürlich die peripheren Input- bzw. Output-Organen und mit ihnen die dort ablaufenden Prozesse. Gehör und Augen sind von Artikulations- und Schreiborgan getrennt, also finden beim Hören und Lesen zweifellos andere Prozesse statt als beim Sprechen und Schreiben, wie auch zwischen mündlichen und schriftlichen Verarbeitungsprozessen ihrerseits. Weniger offensichtlich und im Detail letztlich auch noch nicht aufgeklärt ist das Ausmaß an Überlappung der beteiligten Wissensbestände, der Repräsentationsebenen und der Aktivations- und Verarbeitungsvorgänge zwischen den Repräsentationsebenen.

Beobachtungen zum Anteil der geteilten Ressourcen

Evidenz **Spracherwerb:** Da ist zunächst einmal die Tatsache, dass Verstehen und Sprechen nicht gleichzeitig gelernt werden, sondern mit einer zeitlichen Versetzung. Das Kleinkind versteht früher als es sprechen kann. Das gilt für die Entwicklung der Fertigkeit insgesamt wie auch für einzelne Wörter und Strukturen. Das könnte die Annahme stärken, dass die Verstehensprozesse andere sind als lediglich umgekehrte Produktionsprozesse. Zwingend ist dieser Schluss aber nicht, denn die zeitliche Versetzung könnte ja daran liegen, dass sich nur einfach die Artikulationsfähigkeit später entwickelt als die Hörfähigkeit, nicht aber die dahinterliegenden, kognitiven Prozesse selbst.

Für diesen Zweifel spricht, dass ältere Kinder mit entwickelter Artikulation produktiv und rezeptiv ähnliche Performanz zeigen (Palmer/MacLeod/Hunt/Davidson 1985). Gegen die Überlappungsannahme spricht wiederum, dass Rezeption und Produktion gleichzeitig stattfinden können. Es ist offenbar durchaus möglich, auf einer Party gerade jemandem ein Cocktailrezept zu erklären und gleichzeitig zu hören, wenn in dem Gespräch einer benachbarten Gruppe der eigene Name fällt oder ein anstößiges Wort. Und man kann sich selbst zuhören, während man spricht. Deutet das vielleicht darauf hin, dass sogar auch das lexikalische Wissen modalitätenspezifisch ist, ein Produktionslexikon, ein Verstehenslexikon, ein schriftliches und ein lautliches?

Verschiedene Frequenzinformationen für Laut- und Schriftform: Zumindest für die Trennung der phonologischen Repräsentation von der orthographischen spricht, dass die Frequenzinformationen für beide verschieden sind. Die Frequenz eines phonologischen Lexems lässt keine Rückschlüsse auf die Geschwindigkeit der visuellen Worterkennung zu und umgekehrt. Das zwingt zu der Annahme separater Speicher für die schriftliche und die lautliche Form der lexikalischen Einheit. Ob die Trennung sich aber auch auf die Lemma-Ebene erstreckt, ist wiederum fraglich. Die Tatsache, dass der lexikalische Zugriff bei der Produktion sowohl durch Hören als auch durch das Lesen eines semantisch ähnlichen Wortes gestört wird, spricht wiederum für die Annahme, dass beide Aufnahme- prozesse auf dieselbe lexikalische Repräsentation zugreifen.

Modalitätsspezifische Sprachstörungen: Für die Klärung der Überlappungsfrage ebenfalls aufschlussreich sind Beobachtungen des Sprachverhaltens von Sprachpatienten. Dass sprachliche Module unabhängig von-

einander gestört sein können, deutet auf Nicht-Überlappung hin. Nach Auswertung mehrerer sprachpathologischer Erkenntnisse schließt Harley (2001, S. 395) auf eine minimale Überlappung und postuliert vier modalitätenspezifische mentale Lexika mit nur einem gemeinsamen Modul, dem semantischen System.

Überlappung grammatischer Prozesse: Bei der Annahme, dass nicht nur die lautliche, sondern auch die syntaktische Verarbeitung prozedural ein Aktivationsausbreitungsvorgang ist, entfällt die Notwendigkeit, getrennte Teilsysteme anzunehmen. Aber was heißt das? Stellen nicht schon zwei getrennte Verbindungen zwischen einem syntaktischen Knoten auf der Wortartebene und einem auf der Phrasenebene, eine top-down und eine bottom-up gerichtete Verbindung zwei Systeme dar? Der Befund, dass eine Störung im syntaktischen Verstehen öfter mit eingeschränkter Leistungsfähigkeit des Kurzzeitgedächtnisses einhergeht, als mit einer Beeinträchtigung des allgemeinen Arbeitsspeichers, spricht für ein geschlossenes, abgekapseltes Satzerkennungsmodul, und die Tatsache, dass bei diesen Patienten die syntaktischen Prozesse der Produktion nicht entsprechend gestört sind, spricht für Nicht-Überlappung.

Überlappung der Wissenspeicher: Die Befunde über das sprachliche Wissen deuten darauf hin, dass syntaktisches und semantisches Wissen nicht modalitätsspezifisch sind, weder die lexikalische noch die grammatische Wissenskomponente. Die Verarbeitungswege Produktion und Rezeption werden weithin als separat angenommen. Aber an der Satzverarbeitung setzen Versuche an, diese Annahme in Frage zu stellen. Ein Ausgangspunkt ist die Tatsache, dass man sich als Zuhörer mitunter dabei beobachtet, dass man einen Satz, in dem der Sprecher eben eine Pause gemacht hat, an seiner Stelle hätte zu Ende formulieren können. Diesem Phänomen kann man unter kontrollierten Bedingungen nachgehen. Einer Versuchsperson wird bildlich eine Szene gezeigt, in der mehrere Personen und Objekte vorhanden sind. Mit dem Bild-Onset hört die Versuchsperson einen Satz und ihre Blickbewegungen beim gleichzeitigen Betrachten der Szene werden aufgezeichnet. So lässt sich beobachten, dass sich die Aufmerksamkeit des Hörers je nach verarbeitetem Stück des Satzanfangs auf Bildinhalte richtet, die als nächste in der Äußerung geäußert werden. An der Treffsicherheit der ›Vorahnung‹ (*prediction*) wird geschlossen, dass Verstehen und Sprechen eng miteinander verbundene kognitive Prozesse sind (Altmann/Kamide 1999).

Dass die Vorhersage von Fortsetzungen bei hoher genereller Übergangswahrscheinlichkeit im Langzeitgedächtnis bis aufs Wort genau sein kann, zeigen ERP-Beobachtungen. Die Versuchsperson hört einen Satz, der ein allgemein bekanntes Szenario beschreibt, z. B. *Herbst, Wind, Junge, Rausgehen zum ...* (Englisch *The day was breezy so the boy went outside to fly ...* (vgl. DeLong/Urbach/Kutas 2005). Die Wahrscheinlichkeit, dass das Objekt ein Drachen (*kite*) ist, ist größer als dass es ein Flugzeug ist (*airplane*). Da *kite* die unbestimmte Artikelform *a*, *airplane* aber *an* verlangt, lässt sich prüfen, wie spezifisch die Vorhersage beim Hören ist. Bei Fortsetzung mit *an* wird ein größeres N400-Potential (Indikator für Mismatch) gemessen als bei Fortsetzung mit *a*. Das zeigt an, dass die *pre-*

diction-Prozesse bis auf die Ebene der phonologischen Enkodierung hinunter das erwartete Segment spezifizieren.

Fazit: Sprachverstehen ist wie die Sprachproduktion ein komplizierter kognitiver Prozess, der zwei große Schritte umfasst, die lexikalische Erkennung und das Verstehen der Äußerung. Jeder dieser Schritte umfasst wiederum mehrere Aktivitäten, zwischen verschiedenen Repräsentationsebenen. Wie weit das Sprachverstehen- und das Sprachproduktionssystem sich überlappen, ist unklar und Gegenstand aktueller Forschung. Mit dem derzeitigen Kenntnisstand vereinbar ist, dass hinsichtlich der Überlappung zu unterscheiden ist zwischen:

- sprachlichem Wissen und Sprachverarbeitung
- den Sprachebenen, also phonologisch/graphematische, syntaktische und semantische Verarbeitung
- der lautlichen gegenüber der schriftlichen Modalität