

Hagen Peukert

Kindliche Kalkulationen

Eine Computersimulation über den Einfluss stochastischer Informationen auf die Wortsegmentierung beim Erstspracherwerb

Die vorliegende Arbeit wurde vom Fachbereich Sprach- und Literaturwissenschaften der Universität Kassel als Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie (Dr. phil.) angenommen.

Erster Gutachter: Prof. Dr. Josef Wallmannsberger

Zweiter Gutachter: Prof. Dr. Winfried Nöth

Tag der mündlichen Prüfung

5. November 2008

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

Zugl.: Kassel, Univ., Diss. 2008

ISBN print: 978-3-89958-704-3

ISBN online: 978-3-89958-705-0

URN: urn:nbn:de:0002-7058

© 2009, kassel university press GmbH, Kassel

www.upress.uni-kassel.de

Druck und Verarbeitung: docupoint GmbH, Magdeburg
Printed in Germany

für Rita

Dank

Mein größter und besonderer Dank gilt Professor Josef Wallmannsberger, der mir die Erforschung des Themas ermöglichte und die beste Betreuung zum kritischen Denken angedeihen ließ. Unzählige Stunden seines Sparrings begleiteten mich über die Jahre der Promotion und vielleicht auch darüber hinaus. Aus diesen Gesprächen sind Ideen entstanden, die mich von nun an begleiten; einige davon haben bereits ihren Niederschlag in der vorliegenden Arbeit gefunden. Eine wertvollere Unterstützung kann ich mir kaum mehr vorstellen.

Andreas Köthe danke ich für die programmiertechnische Unterstützung und seinen trockenen Humor gegenüber der linguistischen Denkweise. Ohne ihn wäre die Simulation nicht in dieser Form lauffähig geworden.

Enrique Garcia Sobolevski hat mich fachlich bei der Einarbeitung in Matlab unterstützt. Als Freund fand ich bei ihm immer ein offenes Ohr für die zahlreichen kleineren Probleme bei der Forschungsarbeit. Ich werde die alltäglichen Spaziergänge und die Diskussionen über den Sinn der Wissenschaft vermissen. Oliver Bracht danke ich für seine fokussierten Erklärungen zur logistischen Regression in SPSS.

Uta und Daniel Wolf haben mit viel Ausdauer verschiedene Varianten des Manuskripts durchgesehen und auf allgemeine Verständlichkeit überprüft. Ihnen möchte ich meinen besonderen Dank aussprechen. Professor Winfried Nöth bin ich für die wertvollen Anmerkungen und kritischen Hinweise zu tiefen Dank verpflichtet. Für alle noch verbleibenden Fehler trage allein ich die Verantwortung.

Klaus Raymund gebührt für unsere gemeinsamen sportlichen Aktivitäten, die in allzu stressvollen Situationen für die Erfrischung des Geistes sorgten, ein sportliches Dankeschön.

Mein herzlichster Dank richtet sich an meine Frau Margarita Manakova, die all ihre Liebe und Hingabe dem größten Glück unseres Lebens, Alina und Konstantin, widmete und mir für die gesamte Zeit den Rücken von allen Problemen des Alltags freihielt. Nur so hatte ich die Zeit, mich voll auf mein Forschungsvorhaben zu konzentrieren.

Abstract

At first glance, word segmentation seems trivial – a simple mapping between an incoming stream of sounds and an existing representation thereof in the mental lexicon. When realizing that 6 to 8-months-old babies have no lexicon available (Jusczyk und Aslin (1995, 2); Morgan (1994); Swingle und Aslin (2000); (Aslin et al., 1999, 364)), word segmentation reaches the status of extraneous profundity. Noticing moreover that unlike adults listening to a conversation in an unknown language, babies have little access to non-verbal communication, word segmentation receives once again some more attention. If one acknowledges furthermore that pauses in the speech stream do not collide with the beginnings or endings of words as in written language, word segmentation is finally recognized as a real research problem.

This description reflects the proceedings in research on infant word segmentation for the last fifty years. Only in the last quarter, word segmentation received the deserved attention. Yet the literature documents four strategies for infant word segmentation: allophonic, phonotactic, prosodic, and statistical (Mattys et al. (1999, 467); Gambell und Yang (2005); Weijer (1999, 15ff); Jusczyk et al. (1999c, 161-163); Batchelder (2002, 168); Johnson et al. (2003, 87); Aslin et al. (1998, 322); Bortfeld et al. (2005, 298); Perruchet und Vinter (1998, 246)).

Allophonic varieties are in part predictable and happen to mark word boundaries. Some allophones do only occur in the beginning or ending of words, but in no other environment. Knowing which allophones appear in which position of the word, enables the listener to define the respective boundary. Experimental findings have shown that this ability is present in 10.5-months-old English infants (Jusczyk et al., 1999b).

By the same token, phonotactic sensitivity starts to develop from nine months on (Friederici und Wessels, 1993). Phonotactic knowledge comprises which combinations of sounds and which position within a word are acceptable in a given language. Since some sounds never or only show up in the beginnings or endings of words (e.g. /ŋ/, /mb/ do not occur word initially in English) or some combinations never appear at all (e.g. /ptsk/ in English), it is clear that their occurrence must define word boundaries.

Prosodic cues play a crucial role in word segmentation. In English, they happen to be a reliable indicator. 90% of all words are stress initial. Using this information, it is most efficient to extract words out of a speech stream. 8-months-old English infants are able to use stress cues to segment words (Jusczyk et al. (1993a); Morgan (1996b); (Echols et al., 1997); Johnson und Jusczyk (2001)).

Finally, statistical information considers the relation from sounds or syllables to each other. Statistical learning in word segmentation suggests that the distribution of sounds

within words differs significantly from sounds present in larger units than the word. 8-months-old babies are aware of these differences (Saffran et al., 1996a) and may use them for processing language (Saffran, 2001b). In addition, Thiessen und Saffran (2003) showed that 7-months-olds prefer statistical over prosodic information.

All of the aforementioned segmentation strategies but statistical learning need some representative set of words to derive its rule set, that is, the marking of a word boundary has to be aligned with the language specific property of prosodic, allophonic, or phonotactic cues. To illustrate, Maori infants must learn that velar nasals /ŋ/ may occur word initially and may coincide with a word boundary as opposed to English toddlers who should never assume that. By the same token, English babies as opposed to French, should internalize a rule that defines a word boundary at the primary word stress. Lastly, the aspirated stop [t^h], for example, can only be detected word initially in English but not in Hindi. So the English child maps the word boundary to a phonetic clue, i.e. the aspirated plosive. In contrast to that, transitional probabilities as a general learning principle (Saffran, 2002) do not require an initial set of lexical items since the mechanism is based on a frequency distribution. Items as sounds or syllables that often occur together are simply categorized as a function of their occurrence relative to all others. These units then define the beginning and end of words.

It is still unclear, how these strategies interact, which strategy is prevalent or used first. Moreover, the question how the initial set of prototypes can be extracted is raised in the literature. Since 10% of all words occur in isolation (Brent und Siskind (2001); Aslin et al. (1996); Weijer (1999)), babies would have a sufficient list of words at their disposal from which word segmentation rules could be exploited. However, this suggestion should be dismissed considering that the language learner cannot know whether a word appears in isolation or not. How should one know that *isolate* and *I'm so late* comprise one or three words respectively? Where would this knowledge come from? Because of this contradiction, the statistical approach should be used as the preferred bootstrapping strategy, so it is argued, to start building up a first lexical structure.

On this note, Gambell und Yang (2005) showed that the statistical approach does not hold for representative data of child directed speech. No more than a third of the CHILDES (MacWhinney, 1995) collection could be segmented correctly. This result dispels the statistical approach as an alternative mechanism for word segmentation. The argument goes again: how can infants know which segmentations are right or wrong. Since none of the suggested strategies really holds consistently, the appropriate solution is deemed best to be found within the generative paradigm. It is argued that some information of the structural components of language has to be innate if word segmentation were to be successful. In fact, one could only explain how the segmentation process gets started if specific information on language is admitted (Gambell und Yang, 2005).

The model developed here suggests yet another possibility. Following the logic remarked above, that is, only transitional probabilities can be used as a basis for word segmentation for they circumvent the circular argument of isolated words, the model at hand starts at the point where Gambell und Yang (2005) stopped. In short, it is argued

that the wrongly segmented corpus contains enough information enabling infants to segment lexical units from the speech stream considering infant's cognitive abilities around 7 month of age. Thus, it is shown that a mechanism for word segmentation exists without recourse to components of universal grammar.

The algorithmic design of a model is supposed to reflect the cognitive mechanisms as closely as possible. Of course, the model must abstract away so that the underlying processes can well be understood. The model discussed here will only consider auditory input in the form of transcribed data files from which all prosodic information is deleted. This is not to deny the relevance of rhythmic input, but keeping them out of the model ensures that the mere influence of the statistical information in the speech input is measured.

The last implementation of the model takes representative samples of a controlled size from CHILDES (MacWhinney, 1995), converts them into an IPA-format, deletes all stress information and white spaces. Now the transitional probabilities are calculated for each combination of phoneme chains of the size between 1 and 10. White spaces are marked when a limit is reached. The limit is a variable manipulating the sensitivity of the child and is defined between 0 and 1. Thus, a second loop runs through ten values of limit and outputs its maximum for each phoneme chain combination.

The value at the optimum of both variables (partial derivative), length of phoneme chain and sensitivity of the child, is now assumed to be fix for a given corpus size. From the by then segmented sample corpus, the most frequent items (differing between 5 and 30 words when a constant function is chosen) are selected and saved in a list ordered by length. Eventually the next corpus is input and processed as described above. Depending on a fifth variable encoding the delay of the mapping, the word list is either mapped onto the corpus or delayed to the next specified input.

The constraints defined in the model are all based on general cognitive abilities. Attention was paid not to hide any language specific information in the underlying architecture. Even though all of the constraints were proven experimentally, their interaction is only a suggestion that the simulation can prove realistic. As mentioned above the most important constraint is the inclusion of transitional probabilities (Saffran et al., 1996a) in the model. Second, it is important to pin down the exact form of the unit of perception. From this unit, calculations of transitional probabilities will be accomplished. After thoroughly investigating the psycholinguistic experiments carried out for the last 50 years, the phoneme seemed best suited (Brent (1999b); Jacobs (2003); Pisoni und Luce (1987); Swingley (2005)). Using phonemes allows for a more elaborated inquiry by controlling both possibilities: syllables as well as all other existing combinations of phonemes. Third, it is clear that babies have to memorize words for a longer period of time (Jusczyk und Aslin (1995); Jusczyk und Hohne (1997)), so that, fourth, the most frequent ones (Shi et al. (2006); Jusczyk et al. (1994)) can be mapped top-down (Bortfeld et al., 2005) onto an unknown speech input.

The results of the simulation are as follows. First, the simulation of the proposed model replicates the findings of Gambell und Yang (2005) or others (Brent, 1999b). In addition

to the proposed models so far, the simulation here specifies a number of variables that can be manipulated. This allows to measure the exact size of the corpus that is necessary to make a maximum of correct segmentations as a function of the child's sensitivity and a corresponding unit of perception. The later, so the surprising result, does not matter as long as the sound chain varies somewhere between one and five phonemes. Therefore, for word segmentation, the exact form of the unit of perception provides no additional information. The infant's sensitivity, however, should be minimized. That means phoneme chains need to occur only a few times in the same environment, so that they are recognized as a lexical unit. It also means that even if a unit appears in different sound environments here or there, the infant would chose the most frequent one relative to all others as an lexical entry.

The most important contribution of the simulation is nonetheless an important property of the segmented corpus. It is indeed true that only about one third (if parameters are optimized one half) of all segmented words are correct and that this number cannot possibly account for a starting point in the segmentation process. At first sight this is a convincing argument, but searching the corpus for more detail, it becomes obvious that the wrongly segmented corpora all encode some more important information which is not at all obvious looking at it superficially. The most frequent segmentations of the corpus happen to be lexical items.

Only that finding allows to solve the segmentation problem since out of each representative corpus a certain number of words can now be extracted reliably. This number is described by a function and encodes the number of words an infant should be able to memorize. The simulation showed here that the best results are attained when the function is kept constant, that is, in each period (or for each new corpus) the same amount of words is memorized. Furthermore the number of words memorized from each corpus should not exceed 30 entries since the error function raises rather drastically from there on. Assuming that the function is set to 7, the measured errors are zero. This finding happened to confirm the research done by Miller (1956).

As Pinker (1984, 28) already assumed some time ago, it is now possible to use the extracted words as a source of additional information to define undetected word boundaries. In particular, adding a recursive structure that would map the existing words into the sound stream after the transitional probabilities are computed, is key to compensate the shortcomings of the probability calculations.

Using the cohort model (Marslen-Wilson und Welsh, 1978) to align the lexical entries with the new sound stream, first intermediate data of the simulation revealed notwithstanding a problem that resulted from embeddedness. The problem was overcome by delaying the starting point of recursively entering the word list to a subsequent period in which a sufficient number of corpora have been processed. This ensured that the word list comprises enough words, so that a competition can be initiated among the items of the word list as required in the cohort model. Since the longest though appropriate of all entries would win the competition, it is of great advantage that larger units are included in the list. The statistical relevance between the length of a word and its frequency occurrence is used here as a mechanism (Best, 2003).

The set of functions delineating the increase of extracted lexical entries for each cycle are all precisely described by one master function. When specifying the number of words that the child memorizes during each period, the limit of the master function gives the maximum words a child could extract from an indefinite number of corpora. To be specific, while a constant function of 10 words for each cycle still has an error rate that is statistically insignificant, one could obtain a maximum of 107 lexical entries. This number should be sufficient to serve as an initial word list for deriving other regularities, such as prosodic, allophonic, or phonotactic rules specifically for each language.

To illustrate a concrete example for English CHILDES data, a typical word list extracted from the statistically segmented corpora contains around 80 entries. While an average between 2 to 3 entries do not correspond to words, one of the false segmentations is always the grammatical morpheme /*ɪŋ*/. All of the entries but one carry their primary stress on the first syllable, which makes it easy to deduce the prosodic rule for English. Still, some phrases, such as /*ðerjugəv*/, /*wʌtərjuduɪŋ*/, and /*kənjuseɪ*/ are also included. Although this result is perfectly in line with research done by Peters (1983) or Pinker (1984), it would cause wrong segmentations if used recursively. However, their constituent parts are all listed and so it is reasonable to suppose that these phrases could be split up correctly.

In summary, the simulation carried out approves the correctness of applying the statistical approach to solving the word segmentation problem. As such it argues against the generative paradigm, in which innate knowledge accounts for the poverty from the stimulus. The simulation shows that innate knowledge is not necessarily required to locate word boundaries. The information in the input is sufficient.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung und Ziele	4
1.2	Vorüberlegungen und definitorische Abgrenzungen	7
1.3	Die zweifelhafte Prämisse des Arguments der Reizarmut	11
1.4	Lösungsansatz und Genese der Idee	15
1.5	Vorgehensweise und Aufbau	18
1.6	Einordnung der Arbeit	20
2	Nativismus und Empirismus als grundlegende Positionen	25
2.1	Einführende Bemerkungen	25
2.2	Philosophische Positionen	28
2.3	Symbolistische und konnektionistische Paradigmen	34
2.4	Positionierung der Arbeit in einem weiten Wissenschaftskontext	40
3	Theorien der Wortsegmentierung	43
3.1	Modelle der Wortsegmentierung	44
3.2	Wortsegmentierung ohne anfängliches Lexikon	55
3.2.1	Prosodie: Rhythmus und Betonung	55
3.2.2	Allophonie	62
3.2.3	Phonotaktik	64
3.2.4	Übergangswahrscheinlichkeiten und regelmäßige Verteilungen	67
3.3	Vereinbarung konträrer Standpunkte	71
3.4	Zusammenfassung	74
4	Randbedingungen der Simulation	77
4.1	Wahrnehmung von sprachlichen Einheiten	78
4.1.1	Wahrnehmung linguistischer Einheiten im Allgemeinen	80
4.1.2	Wahrnehmung von linguistischen Einheiten bei Säuglingen	89
4.1.3	Kritische Betrachtung der experimentellen Befunde	96
4.2	Statistische Berechnungen und logisches Schließen	106
4.3	Erinnerungsvermögen von Lautsequenzen	109
4.4	Verwendung und Form der segmentierten Einheiten	110
4.5	Zusammenfassung	111

5	Modellierung einer stochastischen Sprachsegmentierung	113
5.1	Einleitende Bemerkungen	113
5.2	Bestimmung der Parameter und Maßzahlen	115
5.2.1	Allgemeine Leistungsmessung in der Computerlinguistik	116
5.2.2	Anwendung auf den Bereich der Wortsegmentierung	118
5.2.3	Berechnung der Übergangswahrscheinlichkeiten und Lexikongröße	119
5.2.4	Spezifizierung der Parameter	120
5.3	Korpora und Repräsentativität	127
5.4	Beschreibung des Algorithmus	130
5.4.1	Randbedingungen und schematischer Ablauf	131
5.4.2	Zusammenhängende Darstellung	132
5.4.3	Der Sprachinput	134
5.4.4	Die Übergangswahrscheinlichkeiten	139
5.4.5	Häufigkeitsfunktion und Paralexikon	147
5.4.6	Programmanfang	149
5.4.7	Nachbetrachtung	151
5.5	Zusammenfassung	153
6	Simulationsergebnisse und Auswertung	155
6.1	Einleitende Bemerkungen	155
6.2	Bestimmung der genauen Parameterwerte	155
6.2.1	Beschreibung der Simulationsdaten	155
6.2.2	Dateninterpretation und Erklärungen	159
6.2.3	Zusammenfassung	169
6.3	Einfluss rekursiver Strukturen auf die Lexikongröße und F_1	170
6.3.1	Aufbereitung der Daten	170
6.3.2	Beschreibung der Simulationsdaten	184
6.3.3	Dateninterpretation und Erklärungen	185
6.3.4	Zusammenfassung	191
6.4	Entwicklung und Eigenschaften der Lexikonfunktion	192
6.5	Eigenschaften des Paralexikons	212
6.6	Zusammenfassung	217
7	Schlussbetrachtung	221
7.1	Beitrag zur Anlage-Umwelt-Debatte	221
7.2	Anknüpfungspunkte für weitergehende Forschung	228
7.3	Ausblick zur wirtschaftlichen Nutzung der Ergebnisse	231
8	Zusammenfassung	235
	Anhang: Quellcodes, Wortlisten und Kodierungen	239

Abbildungsverzeichnis

1.1	Psycholinguistische Computermodellierung	19
1.2	Das Wissenschaftsumfeld im Rahmen der Dissertation	21
2.1	Mögliche Typologie vorherrschender Forschungsparadigmen	26
3.1	Verarbeitungsebenen für die Erkennung gesprochener Wörter	45
3.2	Entwicklung prosodischer und statistischer Sensibilität	56
3.3	Mögliche Abfolge von Segmentierungsstrategien	75
4.1	Häufigkeitsabhängige Repräsentation von Wahrnehmungseinheiten	102
5.1	Abbildung der Simulation auf den Spracherwerbsprozess	113
5.2	Entwicklung eines repräsentativen Korpus	128
5.3	Allgemeine Funktionsweise des Algorithmus	132
5.4	Problematik des Stringabgleichs	142
5.5	Abzählen von Phonemketten	144
5.6	Informationsfluss der rekursiven Struktur	152
6.1	$F_1(\sigma, \lambda)$ für verschiedene ψ	156
6.2	$F_1(\sigma), \lambda = 1, \psi \leq 500$	160
6.3	$F_1(\sigma), \lambda = 1, \psi \leq 1000$	161
6.4	$F_1(\sigma), \lambda = 5$	162
6.5	$F_1(\sigma), \lambda = 9$	163
6.6	$F_1(\lambda), \sigma = 1, \psi \leq 20000$	164
6.7	$F_1(\lambda), \sigma = 0.6, \psi \leq 20000$	165
6.8	$F_1(\lambda), \sigma = 0.1, \psi \geq 20000$	166
6.9	$F_1(\lambda), \sigma = 0.1, \psi \leq 10000$	167
6.10	F_1 -Maximalwerte, $\operatorname{argmax}((F_1)(\omega)); 10 \leq \zeta \leq 30 \lambda, \sigma = \textit{konst}$	171
6.11	Lexikonentwicklung für große Korpora	172
6.12	Lexikonentwicklung für kleine Korpora	173
6.13	$\psi = 1000, \zeta = 10$ (Veränderung der Wortanzahl)	174
6.14	$\psi = 1000, \zeta = 20$ (Veränderung der Wortanzahl)	175
6.15	$\psi = 1000, \zeta = 30$ (Veränderung der Wortanzahl)	176
6.16	$\psi = 50000, \zeta = 10$	177
6.17	$\psi = 50000, \zeta = 20$	178
6.18	$\psi = 50000, \zeta = 30$	179
6.19	$\psi = 450000, \zeta = 20$	180
6.20	$\psi = 450000, \zeta = 10$	181

Abbildungsverzeichnis

6.21	$\psi = 450000, \zeta = 10$ (ohne Ausreißer)	182
6.22	$\psi = 50000, \zeta = 10$ (ohne Ausreißer)	183
6.23	Widersprüchliche Parameterkombinationen	188
6.24	$F_1(\omega)$ für $\zeta = 10 + \omega$	195
6.25	Entwicklung des Paralexikons	198
6.26	Trendlinien für $\zeta = 10$ und $\zeta = 30$	199
6.27	Lineare Interpolation der Koeffizienten	201
6.28	Entwicklung der Paralexika für $10 \leq \zeta \leq 30$	202
6.29	Vorhersage der Entwicklung des Paralexikons	204
6.30	Fehlerhafte Einträge im Paralexikon	205
6.31	Vergleich zwischen tatsächlichen Fehler und Vorhersage	207
6.32	Trends für falsche und richtige Wörter im Paralexikon	208
6.33	Zusammenhang zwischen F_1 und $\mu; \zeta = 10$ (schematisch)	210
6.34	Zusammenhang zwischen F_1 und $\mu; \zeta = 10$ (faktisch)	211

Tabellenverzeichnis

5.1	Zuordnung der CHILDES-Korpora zu untersuchten Subjekten	122
5.2	ψ , Anzahl der Stichproben und deren Zusammenhang	124
5.3	Form der Ausgabedatei ($maxLength = 5$, $raster = 10$)	143
6.1	Vergleich der Leistungsmaße von zwei Durchläufen	187
6.2	Lexikonentwicklung für $\zeta = 10$	194
6.3	Segmentierungswerte für $F_1(\omega) = 10 + \omega$	196
6.4	Vergleich der Leistungsmaße für $\zeta = 10$ und $\zeta = 30$	209
6.5	Lexikonentwicklung für $\omega = 6$, $\zeta = 10$	214
6.6	Lexikon nach 30 Durchläufen bei $\zeta = 10$, $\mu = 30$	215

Quellcodeverzeichnis

5.1	Variablenbelegung	134
5.2	Zuordnung des Aussprachelexikons	134
5.3	Die „Setzerklasse“ NextWord	135
5.4	Die statische Methode <i>findWord</i>	136
5.5	Umwandlung eines Textes in seine IPA-Form	136
5.6	Umwandlung in IPA Notation	137
5.7	Stringsortierung für beliebige ω und μ	139
5.8	Datenstrukturen für Berechnung der Leistungsmaße	141
5.9	Berechnung der Übergangswahrscheinlichkeiten	144
5.10	Einfügen der Leerzeichen am Wortende und Kohortenmodell	146
5.11	Einlesen der Variablen	149
5.12	Erstellen der Dateiarchitektur	150

1 Einleitung

Eines der faszinierendsten Phänomene der menschlichen Kognition ist der Erstspracherwerb. Jedes Kleinkind meistert diese Aufgabe innerhalb der ersten Lebensjahre. Trotz unterschiedlicher Stimuli beziehungsweise unterschiedlichem Sprachinput gelangt es genau zu jener Grammatik, die alle Angehörigen einer Sprachgemeinschaft verstehen. Dabei sind die kognitiven Prozesse dieses Phänomens weitestgehend ungeklärt. Dies liegt natürlich an der großen Komplexität höherer kognitiver Leistungen, bei denen noch unverstandene Mechanismen überlagernd ineinander greifen. So ist auch der Spracherwerb ein vielschichtiger kognitiver Prozess, bei dem auf neuronaler Ebene dynamische Entwicklungen und Interaktionen zwischen den Schichten stattfinden, die der Wissenschaft bislang keinerlei Einsicht erlaubten und damit mehr Hypothesen als gesichertes Wissen hervorbrachten. Dieser missliche Umstand verleitet den Wissenschaftler wieder einmal dazu, sich auf Input-Output-Relationen zu konzentrieren. Damit scheinen die Atlanten des Behaviorismus in die Kognitionswissenschaften hineingetragen worden zu sein. In der Tat lässt sich das Stimulus- und Response-Denken immer auch in kognitiven Erklärungsansätzen ausfindig machen, aber nunmehr stets im Bemühen das Ergebnis des Outputs auf eine dazwischen liegende Struktur zurückzuführen. Eine direkte Beziehung zwischen Input und Output wurde zugunsten interner Repräsentationsebenen aufgegeben. Problematisch bleibt trotz alledem der hypothetische Charakter der internen Repräsentationsmodelle.

Für die Psycholinguistik ist dabei schon die so wichtige Ausgangsfrage strittig: Kann die äußere Form der Sprache als ein ausreichender Informationsinput angesehen werden oder liegen konkrete sprachspezifische Informationen als mentale Repräsentationen vor? Welche Qualität und Quantität müssen diese Informationen haben? Insbesondere diese Frage nach der Rolle des unvollständigen und unterschiedlichen Sprachinput polarisiert sich in den Wissenschaften zu nativistischen und empiristischen Positionen, die gemäß ihrem Ansatz das beobachtete Phänomen auf verschiedene Weise erklären und interpretieren, ohne aber eindeutige Ergebnisse zu präsentieren (Matthews, 2006, 81). Das nativistische Lager bringt in zeitlich regelmäßigen Abständen das Argument der Reizarmut¹ hervor. Dabei gehen deren Verfechter von einem rationalen, aber unterbewussten

¹(*poverty of stimulus argument* (POSA), *argument from the poverty of the stimulus* (APS), Argument der Reizarmut). Chomsky (1980, 34) hat wohl die Begrifflichkeit als solche eingeführt. Grundzüge dieser Idee sind seit seinen frühen Schriften und anderer Wissenschaftler zu verfolgen (vgl. Thomas, 2002). In der Literatur finden sich teils leicht abweichende Definitionen dieses Arguments (siehe Kap. 1.3 oder Scholz und Pullum (2002)). Zur Vermeidung einer begrifflichen Verwirrung sei hier die Definition von Reali und Christiansen (2003, 2) gegeben:

1. Natürliche Sprache zeichnet sich durch eine komplexe Grammatik aus.
2. Die Informationen dieser Grammatik, welche dem Sprachlerner zur Verfügung stehen, ist der PLI (*Primary Linguistic Input*).

1 Einleitung

mentalen Prozess aus und stützen sich mittlerweile auch auf spezifische Beobachtungen und Untersuchungen unterschiedlicher Sprachen. Sie versuchen zu zeigen, dass es keine vollständige Abbildungsfunktion zwischen dem Sprachinput, den das Kind von der Umwelt erfährt, und dem Sprachoutput, dem internen Regelwerk der Sprache mit einer komplizierten hierarchischen syntaktischen Struktur, geben kann. Die Vollständigkeit des äußeren Sprachinputs für den Erwerb einer Sprache wird bestritten. Als Lösung schlagen die Nativisten das Sprachorgan vor, eine genetische Kodierung mit spezifischen Voreinstellungen für die Sprachverarbeitung.² Genau daran zweifeln die wissenschaftlichen Vertreter der empiristischen Generation. Der empiristische Standpunkt fordert, dass dem Sprachinput (Umwelt) eine größere Rolle zugeordnet wird. Das Lernen einer Sprache ist nicht spezifisch, sondern kann mit allgemeinen Lernmechanismen erklärt werden, die auch für andere kognitive Fähigkeiten verantwortlich sind. Die Reichhaltigkeit des Inputs kann nicht mit herkömmlichen Strukturvergleichen zwischen In- und Output verstanden werden. Deshalb versuchen die empiristischen Konnektionisten,³ die Beziehung zwischen Sprachinput und interner Repräsentation (Output) mit neuronalen Netzwerken nachzuweisen. Diese Computersimulationen geben Aufschluss über mögliche Gesetzmäßigkeiten von selbstorganisierenden Ordnungsstrukturen in der Sprache, welche äußerlich chaotisch anmuten. Dabei geht es nicht darum, die *tabula rasa* des menschlichen Geistes nachzuweisen.⁴ Es steht außer Frage, dass beispielsweise sowohl die Neurogenese als auch die Synaptogenese erblichen Gesetzmäßigkeiten folgt und dadurch auch die Sprachfähigkeit des Menschen als artenspezifische Eigenschaft an sich unzweifelhaft eine genetische Basis besitzt.⁵ Die wissenschaftliche Debatte entzündet sich an der Frage, wie viel und

3. Ein Sprachlerner kann die Grammatik nicht aus dem PLI erschließen.

4. Säuglinge erlernen ihre Erstsprache durch vererbte Konditionierung (*innately-primed learning*).

²Modularität (Fodor, 1983) und Mentalismus treten oft als Begleiter des Nativismus auf. Im Grunde muss man den Nativismus von Modularität und Mentalismus logisch trennen. Eine strikte Trennung (wie sie beispielsweise Cowie (1999, 25f, 153-159) oder Gardner (1989, 208) angeben) gestaltet sich schwierig und lässt sich konsequent nicht durchhalten. Das lässt sich schlüssig damit erklären, dass prominente Vertreter des Nativismus wie Chomsky oder Fodor sich eben auch zur Modularität und zum Mentalismus bekennen (vgl. Gardner, 1989, 197ff). So ist die Modularität „ein Bekenntnis zum Mentalismus“ mit einer Tendenz zum Nativismus (Gardner, 1989, 208). Zudem wird die Modularität als Argument für eine nativistische Sicht der Dinge geschildert (Matthews, 2006, 84).

³Mit der Wiederbelebung der parallel verteilten Verarbeitungsmechanismen in den 80er Jahren hat sich das Wort „Konnektionismus“ durchgesetzt und findet sich oft auch in Abgrenzung zu den symbolverarbeitenden Ansätzen. Aufgrund der ideologischen Nähe zu dem epigenetischen, empiristischen Paradigma (siehe Kap. 2.3), welches auch stets auf die Simulationen von Neuronalen Netzen zurückgreift, sei hier eine Substitution der Begriffe gestattet. Es sei jedoch darauf hingewiesen, dass der Konnektionismus nur Teilbereiche der Epigenese abdeckt und oft mit dem Begriff der empirischen Konnektionisten enger gefasst werden kann.

⁴vgl. Rumelhart und McClelland (1986a, 139ff).

⁵Mit den Worten von Bates und Elman (1996, 1849-50):

Even if we assume that a brain (real or artificial) contains no innate knowledge at all, we have to make crucial assumptions about the structure of the learning device, its rate and style of learning, and the kinds of input that it 'prefers' to receive. Siehe auch Pullum und Scholz (2002, 10).

welche Informationen genetisch kodiert sind.⁶ Dazu schreibt der Nativist Fodor (2001, 9):

On the other hand, empirical investigations of the conditions under which children learn to talk would estimate the information their linguistic environments provide, hence how much poverty of stimulus the language learning process tolerates. One then subtracts the information that is required for the child to achieve linguistic mastery. The remainder after the subtraction is what the child's innate knowledge contributes to the language acquisition process.

Die Forschung der zurückliegenden 50 Jahre hat für die Grundsatzfrage des vollständigen Sprachinput noch keine Klarheit schaffen können.⁷ Möglicherweise ist eine Lösung erst mit fortschreitenden Ergebnissen in den Neurowissenschaften zu erwarten (Müller und Rickheit, 2003, 4). Eine solche Entwicklung würde vom Input-Output-Denken fortkommen und das Interesse auf die Beschreibung der internen Mechanismen fokussieren,⁸ so wie es die Kognitionswissenschaften schon seit Anbeginn für sich deklarieren. Bedenkt man allerdings heute die Vielschichtigkeit der möglichen Betrachtungs- und Vorgehensweisen der unüberschaubaren kognitiven Prozesse, erkennt man schnell, dass ein solches Fortkommen vom Input-Output-Denken noch weit vor uns liegt. Also sollte es der Wissenschaft derzeit dienlicher sein, eine von zahlreichen Teilfragen des Input-Output-Schemas zu betrachten und zumindest für diese Teilfrage eine Erklärung für einen internen Mechanismus vorzuschlagen. Gegebenenfalls kann die Beantwortung der Teilfrage die Beantwortung der anfangs gestellten Grundsatzfrage erleichtern.⁹

Dieser Gedanke wird in der vorliegenden Arbeit aufgenommen. Eine interessante und noch ungeklärte Teilfrage des Spracherwerbsprozesses, in der Nativisten und Empiristen mit ihren unterschiedlichen Auffassungen aufeinandertreffen, ist das Problem der Sprachsegmentierung: Wie gelingt es, wenige Monate alten Babys, genau die Einheiten aus einem Lautstrom zu filtern, die sie als Entitäten zur Bedeutungszuordnung benötigen? Der Autor wird der Frage nachgehen, bis zu welchem Punkt statistische Informationen eine Herauslösung von relevanten Spracheinheiten aus dem Lautstrom gesprochener Sprache erlauben. Dabei werden bewusst andere Segmentierungsinstrumente wie die Intonation, die Satzmelodie und die sprachlich-funktionalen Anhaltspunkte isoliert, damit

⁶vgl. hierzu Szagun (2001, 14 ff u. 40ff) zur epigenetischen Sichtweise.

⁷vgl. z.B. Knobloch (2003, 29); Müller und Rickheit (2003, 2f); Matthews (2006, 82f) und für die gleiche Beobachtung vor 20 Jahren vgl. Bruner (1987, 24).

⁸Fodor verdeutlicht diesen Standpunkt wie folgt:

A psychology (rationalist, empiricist, or whatever) needs to do more than just enunciate the laws it claims that mental processes obey. It also needs to explain what kind of thing a mind could be such that those laws are true of it; which is once again to say that it needs to specify a mechanism. (Fodor, 2001, 17)

⁹Der Genetiker Robert Plomin merkt in seiner Einleitung sehr passend an:

I believe that the best behavioral genetics research of the future will not be conducted by behavioral geneticists. It will be done by researchers in other fields who use this approach to answer theory-driven questions that arise in other fields.“ (Plomin, 1990, vf)

1 Einleitung

der Beitrag der statistischen Berechnungen messbar wird. Das Ergebnis wird also zeigen, welchen maximalen Beitrag die verdeckten stochastischen Informationen eines begrenzten Teils der Sprache für die Wortsegmentierung leisten könnten. Die hier vorgelegte Simulation sagt aber nichts darüber aus, ob diese Informationen tatsächlich in dieser Quantität entdeckt oder genutzt werden. Ein stichhaltiger Beweis steht weiterhin aus. Dennoch gibt die hier vorgestellte Lösung Grund zur Annahme, dass die stochastischen Informationen, zumal sie nun einmal vorhanden sind und Kleinkinder die Fähigkeiten haben, diese zu induzieren, eine plausible Erklärung geben können.

Gelingt es, die Existenz der Informationen im Sprachstrom, deren Qualität sowie Quantität nachzuweisen, und ist es zudem plausibel, dass Kleinkinder die Fähigkeit haben, diese Informationen zu nutzen, kann man schließlich einen Beitrag zur Grundsatzzfrage des Arguments der Reizarmut leisten. Da dieses Argument nur die Aussage trifft, dass es keine eindeutige Beziehung zwischen Input und Output der Sprache gibt, ist es für den Bereich der Wortsegmentierung logisch widerlegt, wenn sich herausstellt, dass alle notwendigen Informationen für die Wortsegmentierung objektiv aus dem Input erfassbar sind. Es sei darauf hingewiesen, dass weder das Argument der Reizarmut noch der Gegenbeweis etwas über die Wirklichkeit der realen kognitiven Prozesse bei der Wortsegmentierung aussagen. Der hier zu führende Gegenbeweis soll nur zeigen, dass das rein hypothetische Argument des unzureichenden Input für die Wortsegmentierung nicht stimmen kann. Aus der Falschheit des zu widerlegenden Arguments, zumal auch hypothetischer Natur, darf man jedoch den Gegenbeweis nicht als korrekten Mechanismus unterstellen.

Stellt sich hingegen heraus, dass die verwendeten statistischen Methoden nicht ausreichend sind, müsste geprüft werden, ob das Zusammenspiel mit anderen Segmentierungsmethoden hinreicht. Erst wenn dies nicht der Fall ist, würde das Argument der Reizarmut bei der Wortsegmentierung mehr Gewicht gewinnen, aber weiterhin unter seinem hypothetischen Charakter zu leiden haben.¹⁰

1.1 Problemstellung und Ziele

Die Problemstellung der Dissertation lässt sich mit den folgenden Fragen spezifizieren.

1. Sind die notwendigen Informationen zur Wortsegmentierung im Sprachsignal vorhanden?
2. Welche Quantität haben diese Informationen?
3. Können Kleinkinder diese Informationen nutzen?

¹⁰Es sei noch angemerkt, dass kompliziertere statistische Verfahren als die hier verwendeten Methoden, theoretisch denkbar sind. Die Wahrscheinlichkeit der tatsächlichen Verwendung dafür ist aber gering, weil sehr viel dafür spricht, dass Kleinkinder im Alter von sechs Monaten die einfachsten Segmentierungsstrategien verwenden, die für eine Kodierungsaufgabe Erfolg versprechen. In einer Experimentserie haben Saffran (2003a), Saffran (2003c), Saffran und Griepentrog (2001) und Saffran et al. (2005) begründen können, dass bei einer tonalen Kodierung eines Inputs sechs Monate alte Babys absolute Tonhöhen als Segmentierungsinstrument bevorzugen. Erst bei Unkenntlichkeit der absoluten Tonhöhe griffen die Subjekte auf die komplizierteren relativen Tonhöhen zurück.

Worauf zielen nun die Fragestellungen im Einzelnen ab? Mit der Beantwortung von Frage 1 wird die qualitative Existenz der unterschiedlichen Übergangswahrscheinlichkeiten im und zwischen den Wörtern festgestellt. Sind die statistischen Informationen, so wie sie in den entwicklungspsychologischen Experimenten geschildert werden, überhaupt vorhanden? Lassen sich die experimentell zu Grunde gelegten Annahmen einer statistischen Verteilung von Lautgebilden in kindgerichteter Sprache auch als ein repräsentatives Phänomen begreifen, das nicht nur in in Einzelfällen, sondern in entsprechenden Korpora nachweisbar ist?

Frage 2 rückt die quantitativen Eigenschaften der mit dem Sprachsignal gegebenen Informationen in den Vordergrund. Also kann an dieser Stelle genauer formuliert werden: Sind die Informationen, die mit dem Sprachsignal dem Lerner als Input zur Verfügung stehen, ausreichend, um Wörter aus dem Lautstrom gesprochener Sprache herauszulösen? Die Quantität der Informationen muss schließlich die kritische Masse erreichen, sodass sie das Kleinkind nutzen könnte. Weiterhin werden schon sehr spezifisch die genauen Parameter herauszuarbeiten sein, auf welche sich das Kind denn zu konzentrieren habe, möchte es die maximale Effizienz in der Wortsegmentierung erhalten.

Schließlich stellt Frage 3 einen Zusammenhang mit den aktuellen Forschungsergebnissen her. Es wird geprüft, ob das Kleinkind mit den experimentell nachgewiesenen Fähigkeiten die Informationen denn extrahieren könnte. Damit lässt sich hier konkret konstatieren, ob die Informationen vom Kind überhaupt wahrnehmbar sind.

Die Beantwortung von (1.) und (2.) richtet sich gegen das Argument der Reizarmut aus dem nativistischen Lager. Es stellt sich die Frage, ob die Informationen des Sprachsignals ausreichend sind und wie in (3.) dann geklärt werden muss, ob das Kind diese Informationen überhaupt nutzen kann. Wird bereits die erste Frage negativ beantwortet, erübrigen sich die Folgefragen. Selbst im Falle einer positiven Erkenntnis in den ersten zwei Fragestellungen kann daraus zwar das Argument unzureichender Daten für die Wortsegmentierung im Generellen widerlegt werden, aber wie gerade erläutert, kann die Existenz von ausreichenden Informationen im Sprachsignal das Argument eines Sprachorgans nicht eindeutig widerlegen, weil unklar ist, ob die Informationen tatsächlich genutzt werden. Allerdings würde dann die Theorie des angeborenen Sprachorgans noch mehr hypothetisch anmuten, als sie es ohnehin schon impliziert. Schließlich stellt sich hier die Frage, warum nur der Mensch einen höchst komplizierten und langwierigen Entwicklungsprozess zu einem genetisch kodierten Sprachorgan auf sich nehmen musste, insofern er doch die einfache Möglichkeit besäße, die bereits vorhandenen allgemeinen kognitiven Heuristika effizient einzusetzen.¹¹

Um einen irreführenden Gedanken und seinen Schlussfolgerungen entgegen zu wirken, sei eine Randbemerkung wiederholt gestattet. Die vorliegende Untersuchung wird nicht die Frage beantworten können, ob ein Kleinkind die eventuell vorhandenen statistischen Informationen tatsächlich benutzt. Eine Erklärung dafür kann nur auf experimentellem Wege eindeutig erfolgen und bleibt somit den Entwicklungspsychologen überlassen.¹² Es

¹¹vgl. Lieberman (2006); Lieberman (2000); Waddington (1975, 304ff).

¹²Mit Saffran (2001b) kann bereits auf einen ersten Beitrag verwiesen werden.

1 Einleitung

besteht ein grundsätzlicher Unterschied zwischen der Existenz einer bestimmten Eigenschaft und deren tatsächlichem Gebrauch. Für die Wortsegmentierung bedeutet diese logische Trennung, dass alle Informationen im Sprachsignal zwar vorhanden sein können und auch mit den kognitiven Fähigkeiten des Kindes eindeutig erkennbar wären, dass sie aber, aus welchen Gründen auch immer, ungenutzt bleiben und das Kind auf andere Anhaltspunkte oder Informationen zurückgreift. Dementsprechend versucht die Arbeit nur zu zeigen, ob es denn theoretisch möglich wäre, lediglich das Sprachsignal als Informationsquelle zur Wortsegmentierung zu nutzen.

Mit der spezifizierten Problemstellung lassen sich nun die zwei Ziele der Arbeit verstehen. Das primäre Ziel der Arbeit ist, zur Diskussion des Arguments der Reizarmut beizutragen und zu zeigen, dass die Wortsegmentierung im Erstspracherwerb zwar universalen, aber eher allgemein kognitiven Gesetzen gehorchen könnte. Diese Gesetze sind nicht in einem separaten Sprachmodul mit noch unerforschten Phänomenen eingegliedert und genetisch vorprogrammiert, sondern sie beruhen auf in der Sprache versteckten Informationen. Diese Informationen werden dem Kleinkind durch allgemeine kognitive Fähigkeiten zugänglich gemacht. Die Schwächen des Arguments der Reizarmut werden für die anfängliche Wortsegmentierung nachfolgend benannt. Dabei soll impliziert werden, dass das tautologische Argument sich hauptsächlich auf das Fehlen geeigneter Methoden gründet, die versteckten Informationen der Sprache so, wie sie dem Kleinkind zugänglich sind, einer Sprache offen zu legen und zu erklären. Aus dieser Erklärungsnot heraus entsteht ja letztendlich das derzeit unwiderlegbare Argument eines genetisch kodierten Sprachmoduls.

Als sekundäres Ziel wird ein Beitrag innerhalb der Computerlinguistik im Bereich des datengetriebenen Lernens (*Data-Driven-Learning*) angestrebt. Dieses Teilgebiet ist nach anfänglichen Erfolgen (Harris, 1967) in der computerlinguistischen Forschung etwas vernachlässigt worden, erlebte aber zu Beginn des neuen Jahrtausends eine kleine Renaissance, einerseits mit der Weiterentwicklung alter, andererseits mit der Entwicklung neuer Denkansätze (z.B. Goldsmith (2001); de Marcken (1996); Baroni (2002) und Baroni (2005)). Prinzipiell setzt das *Data-Driven-Learning* an der Idee der maschinellen Informationsverarbeitung an. In seiner höchsten Erfüllung streben Anhänger dieses Ansatzes nach der Erschaffung von intelligenten, wissensbasierten Systemen, die, ähnlich dem menschlichen Informationsverarbeitungsprozess, einen Informationsinput zu Wissen selbständig verarbeiten können. Hier greift eine Vielzahl von Schnittmengen der Linguistik mit denen der KI-Forschung ineinander (vgl. Görz und Nebel, 2003, 45ff). Mit den Ergebnissen dieser Arbeit kann ein kleiner Schritt in Richtung selbständiges Lernen von Maschinen gegangen werden. Es wird gezeigt, dass Algorithmen existieren, die einen in Phoneme zerlegten Lautstrom in bedeutungstragende Einheiten kategorisieren können, ohne dass dabei sprachspezifische Informationen vorgegeben sein müssen. Diese beiden Ziele spiegeln so die idealistische und pragmatische Begründung für die Beschäftigung mit der Computerlinguistik im weiteren Sinne und der Korpuslinguistik im engeren Sinne wider.¹³

¹³vgl. Paprotté (2005, 376).

1.2 Vorüberlegungen und definitorische Abgrenzungen

In der Literatur finden sich teilweise feinere Aufschlüsselungen von konträren Wissenschaftsstandpunkten für die Erstspracherwerbsforschung, als für diese Arbeit angebracht erschien. Klann-Delius (1999) unterscheidet beispielsweise nativistische (z.B. Chomsky, 1965), kognitivistische (Piaget, 1923) und interaktionistische (Vygotsky, 1978) Erklärungsmodelle, welche vielversprechend erscheinen.¹⁴ Zimmer (2003) nimmt der Vollständigkeit wegen auch noch den Behaviorismus hinzu, wohl eher aufgrund seiner historischen Bedeutung in der Psychologie. Eine vollständige Auflistung findet sich in Pence und Justice (2008, 56-57). Neben den bereits genannten behavioristischen (Skinner, 1957), interaktionistischen und kognitivistischen Theorien, werden auch jüngere Entwicklungen vorgestellt. Darunter das Intentionalitätsmodell von Bloom (2001) und die *usage-based-theory* von Tomasello (2003); eventuell kann man dazu auch noch MacWhinneys Wettbewerbsmodell (MacWhinney, 1987) zählen. Die nativistischen Ansätze lassen sich in Fodors Modularitätstheorie (Fodor, 1983), die Bootstrappingtheorien (Pinker (1984) für die Semantik; Gleitman (1990) für die Syntax) und die bereits genannte, prominenteste Theorie einer universalen Grammatik von Chomsky (1965) spezifizieren. Die konnektionistische Bewegung (Rumelhart und McClelland (1986a), Rumelhart und McClelland (1986b)) kann man zwar beiden Seiten zuordnen, wie aber später argumentiert wird, stehen sie dem empiristischen Paradigma näher.

Derartigen Einteilungen wird hier ganz bewusst nicht gefolgt, und die Arbeit wird nicht in den Dienst einer spezifischen Theorie gestellt. Für das Erkenntnisziel der Arbeit ist die allgemeinere Zuordnung in nativistisch und empiristisch angebrachter, weil damit das Wesen der gegensätzlichen Standpunkte am besten erfasst werden kann. Das kognitivistische Modell scheint noch immer stark auf die Arbeit Piagets ausgerichtet zu sein. In der Tat liegt genau dort eine der Wurzeln dieses Ansatzes, aber gerade in jüngerer Weiterentwicklung dieser Ideen mussten einige der ursprünglichen Gedanken revidiert werden. Außerdem führt das Wort „Kognition“ oder auch „Kognitivismus“ oft zu terminologischen Verwechslungen. Schließlich können gerade die Nativisten im Bereich der Sprache auf ihren Verdienst der kognitiven Revolution verweisen, und so lassen sich in der Literatur immer wieder Überschneidungen erkennen. Der interaktionistische Einfluss ist nun wieder in nativistische und empiristische Modelle eingeflossen und wird dabei im Prinzip oft implizit vorausgesetzt, wie auch in Berk (2005, 223) beschrieben. Aus diesem Grund folgt die Arbeit einer Zweiteilung der Ansätze wie sie etwa von Szagun (2001) und Szagun (2006) – nativistisch vs. epigenetisch – oder Dittmann (2006, 59ff) – Vinat vs. Vanit¹⁵ – vorgeschlagen wird. Diese Einteilung gibt im Übrigen das Gegensatzpaar von Nativismus und Empirismus in der Erstspracherwerbsforschung wieder. In der Entwicklungspsychologie spricht man oft von Präformation vs. Epigenese wohl in Anlehnung an Piaget (1974, 15ff). Bei Elman et al. (1997) wird der Begriff *constructivism* verwendet, welcher zwar unter Zuhilfenahme von *constraints* die Idee des Nativismus in sehr stark abgeschwächter Form assimiliert, aber im Übrigen dem epigenetischen Gedanken-

¹⁴vgl. hierzu die Zusammenfassung von Bußmann (2002, 620).

¹⁵Vinat: von innen nach außen; Vanit: von außen nach innen.

1 Einleitung

gut treu bleibt, sodass auch diese Spielart empiristisch einzuordnen wäre. Des Weiteren finden sich noch Anlage und Umwelt als typische Begriffspaare. Da sich diese Arbeit interdisziplinär orientiert und die detaillierten Argumente der Positionen nicht Gegenstand der hiesigen Untersuchungen sind, sei auch eine quasi-synonyme Verwendung der Begriffe zugestanden. Demnach verweisen hier Rationalismus, Nativismus, Präformation und Anlage im Wesentlichen immer auf den gleichen Referenten. Auf gleiche Weise wird mit Empirismus, Epigenese, Konstruktivismus oder Umwelt verfahren.

Im universalgrammatischen Ansatz des Nativismus besteht Einigkeit darüber, dass bestimmte Sprachphänomene aus der Erfahrung gewonnen werden können, andere Phänomene sich aber nicht damit erklären lassen. Folglich müssen alle Besonderheiten einer Sprache, die nicht auf ein Erlernen aus dem praktischen Sprachgebrauch zurückzuführen sind, auf Sprachprinzipien beruhen (vgl. Culicover, 1997, 10). Setzt man diese Zweiteilung von erlernbarem und angeborenem Sprachwissen voraus, ist es für das universalgrammatische Paradigma von grundlegender Bedeutung, eindeutige Aussagen zu treffen, welche spezifischen Phänomene denn nun erlernbar sind und welche zwangsweise von einem empirischen Lernprozess ausgeschlossen werden müssen. Da vererbbares Sprachwissen zwingend unabhängig von einer Einzelsprache betrachtet werden muss, finden die Ergebnisse für alle natürlichen Sprachen Anwendung und können somit eindeutig verifiziert werden. Nur eine einzige natürliche Sprache ist ausreichend, um zu beweisen, dass ein vermeintlich universalgrammatisches Phänomen nicht als solches klassifiziert werden kann. Dies ist genau dann der Fall, wenn die universale Regel nicht widerspruchsfrei in dieser einen Sprache angewendet werden kann.

Dazu können neben der empirischen Sprachforschung, speziell dem Analysieren von Sprachphänomenen unterschiedlichster Sprachfamilien, die Ansätze in der theoretischen Linguistik fruchtbar sein, welche die logischen Voraussetzungen eines universalsprachlichen Prinzips überprüfen. Stellt sich bereits auf theoretischer Ebene ein Widerspruch heraus, ist es hinfällig, die zu Grunde liegenden Phänomene auch empirisch auf ihre Universalität hin zu überprüfen. Sie sind schlichtweg dem erlernbaren Sprachwissen zuzuordnen. In gleicher Weise wird in dieser Dissertation verfahren. Die Beantwortung der ersten Frage (in Kap. 1.1 auf Seite 4) wird zunächst nur Aufschluss darüber geben, ob unterschiedliche Übergangswahrscheinlichkeiten in und zwischen Wörtern überhaupt existent sind. Erst dann wird deren Repräsentativität geprüft. Schließlich werden die kognitiven Fähigkeiten eines acht Monate alten Babys darauf bezogen. Für Letzteres sind mögliche Überschneidungen mit anderen Teilbereichen der theoretischen Linguistik, wie der Semantik möglich, da das Erkenntnisobjekt der Arbeit insbesondere das Erkennen und Herauslösen von bedeutungstragenden Strukturen aus dem Gesamtsprachapparat umfasst. Die Berücksichtigung der *Bedeutung* im Modell führt oftmals zu Fehlschlüssen hinsichtlich der zeitlichen Abfolge der Bedeutungszuordnung. Und so wird das Segmentierungsproblem mit Lösungsansätzen der Semantik vermengt. Diese Argumente haben sicherlich ihre Berechtigung, wenn der Sprachapparat eine bestimmte Größe erreicht hat und man nun aus den Bedeutungsinhalten analytische Ableitungen treffen kann. Für die anfängliche Bildung eines Sprachapparates ist dies aber nicht möglich. Eine semantische Klassifikation und Bedeutungszuweisung kann ja erst dann erfolgen, wenn die zu

1.2 Vorüberlegungen und definitonische Abgrenzungen

bezeichnende Einheit auf irgendeine Weise grob umrissen, also bestimmt werden kann. Betrachtet man die einzelnen Morpheme als Zeichen im Sinne de Saussures, dann wird klar, dass die Beziehung zwischen *signifiant* (das Lautbild des Morphems) und *signifié* (die Vorstellung einer Bedeutung davon) erst dann zu Stande kommt, wenn das *signifiant* überhaupt vorhanden ist oder in seinen Grenzen wahrgenommen werden kann. Die Erstellung der semantischen Verbindung als solche ist demnach dem Erkennen der äußeren Form nachgelagert und hilft bei der anfänglichen Analyse des Segmentierungsproblems nicht. Diese Überlegung spiegelt sich auch in der wissenschaftlichen Literatur wider.¹⁶

Woher also wissen Kleinkinder, welche Einheiten der Sprache als morphemische Bausteine zu größeren Komplexen verbunden werden können und dass gegebenenfalls gerade auf diese Einheiten bestimmte Regeln anzuwenden sind? Im Zusammenhang mit der universalgrammatischen Idee stellt sich die Frage, ob diese Strukturen aus dem vorhandenen Sprachmaterial anhand von generellen kognitiven Fähigkeiten erlernbar sind oder ob sie auf einem im Sprachmodul vorhandenen Prinzip beruhen. In dieser Arbeit wird die erste Alternative favorisiert, wobei die generelle kognitive Fähigkeit das Vergleichen von Phonemkombinationen ist und mittels der Statistik errechnet werden kann. Diese Vorgehensweise wirft natürlich Komplexitätsprobleme auf. Es müssen Einschränkungen getroffen werden.

Insbesondere können statistische Informationen in einer Sprache vielfältig ausgelegt werden. Abhängig vom kryptographischen Detailwissen ließe sich ein Zeichensystem wie die natürliche Sprache in unendlich viele Ebenen untergliedern und mit stochastischen Rechenverfahren durchdringen. Derartige Verfahren werden in der vorliegenden Arbeit nicht angestrebt und der computerlinguistischen kryptographischen Forschung überlassen. Es sollen vielmehr einfache Häufigkeiten und Korrelationen von möglichen Phonemkombinationen einer Zielsprache auf einer Ebene untersucht werden. Das bedeutet, dass es theoretisch noch Korrelationen geben kann, welche erst zu Tage treten, wenn man mehrere Ebenen betrachtet. Dies ist beispielsweise bei Zirkumfixen der Fall.¹⁷ In diesem Zusammenhang sind auch Korrelationen denkbar, die in der Syntax kodiert sind und über viele Wortgrenzen hinwegreichen können. Diese Phänomene werden hier aus den folgenden Gründen ausgeklammert: Erstens tritt die kognitive Verarbeitung von wortübergreifenden morphosyntaktischen Zusammenhängen nicht in der Frühphase¹⁸ des Erstspracherwerbs auf (vgl. etwa Szagun, 2001, 48). Die Frühphase ist für den Ansatz des statistischen Lernens aber besonders interessant, weil hier die Fähigkeit des Säuglings eine *Top-down*-Lernstrategie anzuwenden, zwar vorhanden, aber noch sehr begrenzt ist (Bortfeld et al., 2005). Der Säugling sollte vorrangig mittels des *Bottom-up-Prinzips* ohne

¹⁶Connine (2004, 1084); Jusczyk und Aslin (1995, 2); Morgan (1994); Swingley und Aslin (2000); (Aslin et al., 1999, 364).

¹⁷So ist es bei der deutschen Partizipbildung von beispielsweise *ge-mach-t* nicht möglich die eingeschlossene Einheit *mach* in eine lineare statistische Abhängigkeit zu den umschließenden Präfix *ge* und Suffix *t* zu bringen, weil die dazwischenliegende Einheit höchst unterschiedliche Phonemkombinationen besitzt, die nicht vorhersagbar sind. Die Einführung einer zweiten Ebene ermöglicht es, die morphologische Konstruktion unabhängig von dem Verbstamm zu verarbeiten (vgl. hierzu den Ansatz von Sproat (1992) zur computerlinguistischen Umsetzung der Zweiebenenmorphologie).

¹⁸Zumindest nicht für den hier interessierenden Zeitraum, d.h. bis zum achten Monat.

1 Einleitung

weiteres Vorwissen der Sprachstruktur induzieren und die dadurch gewonnenen Informationen erst später im *Top-down*-Verfahren nutzen. Zweitens verschafft die Untersuchung von nur einer Ebene eine relative Übersichtlichkeit und Einfachheit. Komplizierte statistische Verfahren werfen große Zweifel auf, ob die menschlichen Fähigkeiten insbesondere innerhalb des ersten Lebensjahres für derartige Berechnungen ausreichend sind. Bereits mit empirischen Experimenten abgesichert sind die einfachen Häufigkeiten eine gute Alternative, die Bildung einer ersten Sprachstruktur zu erklären. Drittens können die grundlegenden Prinzipien, welche auf der ersten Ebene möglich sind, letztendlich auf alle nachfolgenden Ebenen übertragen werden, sollte sich diese Erforderlichkeit herausstellen. Hier ändern sich natürlich die Komplexität und der Rechenaufwand, nicht jedoch das grundlegende Prinzip.

Der Erstspracherwerb erstreckt sich über eine Zeitspanne von 10 bis 12 Jahren, die man in der Entwicklungs- und Psycholinguistik in verschiedene Phasen unterteilt.¹⁹ Im Hinblick auf das Erkenntnisziel dieser Arbeit ist es sinnvoll, ein Problembereich aus der Frühphase des Prozesses zu wählen, weil hier die Auswirkungen der Umwelteinflüsse, insofern überhaupt möglich, noch am ehesten kontrollierbar sind. Zwar bestimmen auch pränatale Umweltreize entscheidend die nachgeburtliche Entwicklung,²⁰ diese sind aber auf Grund der derzeitigen wissenschaftlichen Methodiken noch nicht umfassend in Sprachmodelle integrierbar und können folglich auch hier nicht berücksichtigt werden. Somit steht die Frühphase des postnatalen Spracherwerbs im Mittelpunkt der angestrebten Untersuchungen.

Bei der Einteilung der Spracherwerbsphasen ist man anfangs von Beobachtungen der Sprachproduktion ausgegangen und kam dadurch zu relativ weitläufigen Zeitspannen, wie Lall-, Einwort- oder Zweiwortphasen, die sich aber lediglich auf die Äußerungen der Sprachlerner beziehen und nicht allen kritischen Entwicklungen der Sprachaufnahme gerecht werden (vgl. Ingram (1989, 32ff) oder Curtin und Werker (2007, 579)).²¹ Wie Jusczyk (1997, 1) andeutete, hat die wichtige Rolle der Sprachperzeption im Spracherwerbsprozess erst später Anerkennung gefunden, obwohl die zeitliche Verschiebung von Sprachverstehen und der späteren Sprachproduktion zu keinem Zeitpunkt in Frage stand. Insbesondere ist es dem Verständnis der Problemstellung dieser Arbeit dienlich, die Frühphase des Spracherwerbs feiner zu gliedern, als in der Sprachproduktionsforschung dominierend impliziert wird. Somit liegt der Problembereich nicht in oder zwischen der Lall- und Babbelphase,²² die mit vier bis fünf beziehungsweise sechs bis

¹⁹Clark (2003, 16); McLean und McLean (1999, 137ff); Oerter und Montada (1998, 713ff)

²⁰vgl. z.B. Altmann (1997, 14); Lecanuet (1998); Golinkoff und Hirsh-Pasek (2000, 14ff); Karmiloff und Karmiloff-Smith (2001, 43f); Hüther und Krens (2005); Dehaene-Lambertz et al. (2006, 369); Curtin und Werker (2007, 579f).

²¹Mit dieser Sichtweise kam man auch zu Beschreibungen wie *prelinguistic infant* für das erste Jahr nach der Geburt, welcher ein sehr unpassender Ausdruck für die linguistischen Fähigkeiten eines Kleinkindes in dieser Zeit ist (vgl. Macken, 1995, 675).

²²vgl. z.B. Butzkamm und Butzkamm (2004, 58); Elsen (1991, 41ff); Golinkoff und Hirsh-Pasek (2000, 39ff); Pierce und Profio (2006, 111) oder auch kurz Anderson (2001, 373-374). Abhängig von der Quelle weichen die Zeitangaben der Phasen und deren Benennung leicht ab. Klann-Delius (1999, 23) beispielsweise benennt die erstgenannte Phase die „Phase der stimmlichen Expansion“, die bei

1.3 Die zweifelhafte Prämisse des Arguments der Reizarmut

zwölf Monaten angegeben wird, sondern zwischen dem sechsten und achten Lebensmonat, welcher als kritischer Zeitraum für die Segmentierung von bedeutungstragenden Einheiten in der Sprachperzeption zur Diskussion steht (Johnson und Jusczyk (2001); Thiessen und Saffran (2003)). Davor und parallel zu dieser Segmentierungsphase erfolgt die Unterscheidung von lautlichen Segmenten. Auch für diese Phase ließen sich Beiträge zur Grundsatzfrage des ausreichenden Sprachinput finden. Allerdings scheint die „Wortsegmentierungsphase“ insofern besser geeignet, als dass hier nicht nur die Fähigkeiten der lautlichen Wahrnehmung und Kategorisierung einfließen, sondern auch kognitive *Bottom-up*- und *Top-down*-Prozesse bereits deutlicher Wirkung zeigen. Somit beginnen die Untersuchungen nicht direkt am Anfang des postnatalen Spracherwerbs, sondern wenig zeitlich versetzt.

In dieser Zeit, also ab dem sechsten Lebensmonat, beginnen Kleinkinder aus einer schier endlos scheinenden Aneinanderreihung von Lautsegmenten zusammengehörige Einheiten zu extrahieren. Komplizierte Konkatenationen, Koartikulationen und ähnliche Assimilationsprozesse erschweren die Aufgabe erheblich. Die Einflüsse konnten allerdings nicht in der Simulation berücksichtigt werden.

In Kurzform lässt sich das Problem zur Veranschaulichung, wie folgt reduzieren. Bei gesprochener Sprache, im Gegensatz zu geschriebener Sprache, sind keine Hinweise in Form von Pausen beziehungsweise Lücken vorhanden. Das Kleinkind muss also auf andere Anhaltspunkte im Lautstrom zurückgreifen. Ein ähnliches Problem tritt auf, wenn Erwachsene eine fremde Sprache hören und nicht wissen, wo einzelne Wörter beginnen oder aufhören. Der Sprachfluss erscheint ihnen dann meist schneller als derjenige von Sprachen, mit denen sie vertraut sind. Es stellt sich die Frage, wie es die kleinen Sprachkünstler schaffen, diese Aufgabe zu bewältigen.

1.3 Die zweifelhafte Prämisse des Arguments der Reizarmut

Unter Bezugnahme auf Platons Menon und Descartes Auffassung der angeborenen Ideen prägt Chomsky (1980, 34-35) den Begriff des *argument from the poverty of the stimulus*,²³

drei bis vier Monaten beginnt. Golinkoff und Hirsh-Pasek (2000, 39) andererseits geben zwischen vier und acht Monaten an und nehmen keine feineren Unterteilungen vor. Eine recht übersichtliche Darstellung unterschiedlicher Untersuchungen zur Entwicklung des Lautrepertoires während des ersten Lebensjahres gibt Papoušek (1994, 71-73).

²³Das Argument der Stimulusarmut wird hauptsächlich auf die Syntax bezogen (Clark (2003, 18); Pullum und Scholz (2002, 19); Ritter (2002)). Dennoch wird die *heated theoretical argumentation* (Bernstein Ratner, 1996, 135) oft auf vorgelagerte Bereiche des Syntaxerwerbs angewandt. Darunter zählt auch die Lautsegmentierung (vgl. Bernstein Ratner (1996, 135); Collins (2003, 176); Yang (2004)). Liberman und Mattingly (1989, 492) postulieren in diesem Sinne ein *phonetic module*, um die kategorische Sprachwahrnehmung erklären zu können (vgl. hierzu auch Massaro (1994, 220f); Aslin und Pisoni (1980)). Stampe (2004) nimmt auch einen nativistischen Standpunkt ein, wenn er davon ausgeht, dass alle phonologischen Eigenschaften schon vorhanden sind und sich dann nur jene herausbilden, welche in der umgebenden Sprache existieren. Jusczyk (1982) geht der Frage nach, ob Babys mit einem speziellen Prozessmechanismus speziell für die Sprachaufnahme ausgestattet sind. Dittmann (2006, 74) identifiziert vier angeborene Fähigkeiten, welche der chomskyschen

1 Einleitung

den er definitorisch aber nicht präzise abzugrenzen weiß und deshalb viele verschiedene Spielarten, wenn auch zugegebenermaßen in der Idee sehr ähnlich, provozierte.²⁴ Pullum und Scholz (2002, 17) spezifizieren den Kern des Arguments in deduktiver Form wie folgt:

- a Human infants learn their first languages either by data-driven learning or by innately-primed learning. [Disjunctive premise; by assumption]
- b If human infants acquire their first language via data-driven learning, then they can never learn anything for which they lack crucial evidence. [By definition of data-driven learning]
- c But infants do in fact learn things for which they lack crucial evidence. [Empirical premise]
- d Thus human infants do not learn their first languages by means of data-driven learning. [From (b) and (c), modus tollens]
- e Conclusion: human infants learn their first languages by means of innately-primed learning. [From (a) and (d), disjunctive syllogism]

Wie die Autoren richtig bemerken, liegt der entscheidende Punkt bei der empirischen Prämisse (c), darin, ob überhaupt die „Beweise“ fehlen oder ob wir sie einfach nicht explizit beschreiben können. Bislang scheitert der Versuch, beispielsweise bei Legate und Yang (2002) als Antwort auf Pullum und Scholz (2002), eindeutig darzulegen, dass die kleinen Sprachlerner tatsächlich etwas lernen, wofür ihnen eindeutige Anhaltspunkte fehlen. Dies mag in der Tat oberflächlich so aussehen, insbesondere, wenn man den Spracherwerbsprozess aus der Sicht der kognitiven Fähigkeiten und den bereits veränderten Gehirnstrukturen von Erwachsenen betrachtet.²⁵ Der Umstand, dass die notwendigen Informationen im Sprachinput nicht offensichtlich erkennbar sind, bedeutet doch nicht, dass sie nicht existieren. Zudem ist es nicht zielführend, Sprachbeispiele isoliert vom Umfeld des gesamten Sprachsystems zu betrachten. In diesem Glauben merken auch Bates und Elman (1996, 1849) an: „Chomsky’s belief in the impoverished nature of linguistic input holds only if we look ‘locally’ at relatively short segments of speech.“ Betrachtet man dagegen den Gesamtsprachapparat, erscheinen gerade auch die von Chomsky angeführten Ausnahmen in einem anderen Licht. „Such imperfections wash out with a large enough sample“ (Bates und Elman, 1996, 1849).

Universalgrammatik zu Grunde liegen: „die Fähigkeit, den Lautstrom zu segmentieren, Wortkategorien . . . und syntaktische Kategorien zu bilden und Phrasenstrukturanalysen durchzuführen . . .“. In der Tat können diese vier Fähigkeiten in Chomsky (1988) gefunden werden, wenn man Chomskys Ausdruck der *form* (S. 133) einer entsprechenden Interpretation unterzieht. Am deutlichsten wird Chomskys Haltung gegenüber der Segmentierung, wenn er sie als „prinzipiell verfehlt“ (*mistaken in principle*) Chomsky (1977, 181) beschreibt.

²⁴Eine Zusammenfassung der unterschiedlichen Definitionen und Prämissen gibt Pullum und Scholz (2002, 11-17).

²⁵Zum Beispiel bemerken Erwachsene oft nicht, wenn beim Hören von zwei ihnen unbekanntem Sprachen hin und her gewechselt wird (Mehler et al., 2000). Säuglinge aber verfügen noch über diese Fähigkeit (Mehler et al., 1988).

1.3 Die zweifelhafte Prämisse des Arguments der Reizarmut

Der Beweis, den Gold (1967) anführt und auf den gern als Rechtfertigung gegen die empirische Prämisse verwiesen wird, sagt lediglich aus, dass eine reine Form des empirischen Sprachlernens nicht möglich ist, und selbst diese Beweisführung ist äußerst strittig, da hier Annahmen gemacht werden, welche für natürliche Sprachen unzutreffend sind.²⁶ So werden die statistischen Verteilungen außen vor gelassen und dies führt natürlich zu dem Ergebnis, dass der *negative evidence* unverzichtbar wird (vgl. Rohde und Plaut, 1999). Unter dieser Annahme kann man nun zeigen, dass es nur unzureichende Beispiele in der Sprache gibt, die dem Kind ein eindeutiges Urteil über die Verwendung falscher oder richtiger Konstruktionen erlaubt.

Entscheidend bei der Diskussion des Arguments der Reizarmut ist jedoch die Darstellung des logischen Schlusses. Weil man in der Anfangsprämisse (a) eben nur zwei mögliche Szenarien annimmt und diese dann auf die Logik überträgt, gelangt man automatisch zu falschen Erkenntnissen über den Charakter des Erstspracherwerbs, obwohl die Schlussfolgerung im Bereich der Logik völlig korrekt ist. In formaler Notation kann die Prämisse (a) aufgeschlüsselt werden in A1 und A2, welches dann für die oben zitierte Schlussfolgerung als

$$\{A1 \vee A2\} \vdash D \quad (1.1)$$

bezeichnet werden kann und wobei A1 den Spracherwerb aus der Umwelt, A2 den Spracherwerb durch ein Sprachorgan und D die erlernte Sprache (Sprachbeherrschung) symbolisieren. Wenn

$$\neg A1 \rightarrow \neg D, \quad (1.2)$$

wird modus tollendo tollens

$$A2 \rightarrow D \quad (1.3)$$

geschlossen. Mit einer analogen Argumentation ließe sich aber auch

$$\neg A2 \rightarrow \neg D \quad (1.4)$$

und

$$A1 \rightarrow D \quad (1.5)$$

herleiten, indem man zeigt, dass nicht alles Erlernbare der Sprache genetisch kodiert ist. Obwohl auch hier in der Logik ein zulässiger Schluss vollzogen wurde, entbehrt die Beweisführung für (1.4) jegliche Nähe zur Realität.²⁷

²⁶Scholz und Pullum (vgl. 2002, 194-196); Elman et al. (1997, 385 und 45ff; 107ff).

²⁷Wenn man soweit gehen will und Gold (1967) Aussagen akzeptiert, dann muss zumindest die Plausibilität der ersten Prämisse a in Zweifel gestellt werden beziehungsweise einer Modifizierung unterzogen werden: Säuglinge lernen ihre erste Sprache durch die rein äußere Form oder durch ein Sprachorgan oder aus einer Mischform zwischen beiden. Die Mengen A1 und A2 sind also nicht grundsätzlich disjunkt ($A1 \cap A2 \neq \emptyset$), sondern es gilt

$$A1 \vee A2 \vee (A1 \wedge A2) \vee (a1 \wedge a2) \vdash D, \text{ wobei } a1 \supset A1; a2 \supset A2$$

In der Tat geht der Nativismus auch von einer derartigen Konstellation aus. Die *nature-nurture*-Debatte handelt vielmehr von den Mächtigkeiten der zwei Teilmengen A1, A2 und deren Schnitt.

1 Einleitung

Die zweifelhafte empirische Prämisse c gründet sich auf dem fehlendem Wissen über die Fähigkeiten von Kleinkindern, über deren ontogenetische Entwicklung und die Funktionsweise des Gehirns, wenn es höhere kognitive Prozesse ausführt. Die Gesamtheit der Möglichkeiten, welche den *crucial evidence* geben könnten (entspricht #A1), kann noch nicht einmal klar umrissen werden, und trotzdem postuliert der nativistisch geprägte Forscher die Unzulänglichkeiten der noch unerforschten Einflussvariablen. Bevor A1 in (1.2) negiert werden kann, muss doch zunächst klar sein, was A1 umfasst. Das ist aber bislang noch nicht geschehen und deshalb ist auch der Folgeschritt in (1.3) $A2 \rightarrow D$ zu früh vollzogen worden.

Chomsky (1965, 30) geht von einer „class of possible hypotheses about language structures“ aus. Die möglichen Hypothesen sind dabei „presumably, infinitely many“ Chomsky (1965, 30), sodass das Kleinkind einen Mechanismus benötigt, um aus der Gesamtmenge möglicher Hypothesen genau die auszuwählen, welche für seine Sprache zutreffend sind. Ansonsten könnten weder die Kongruenz zwischen zeitlicher Abfolge und der Sprachkompetenz, noch der Umstand, dass alle Sprachlerner überhaupt eine Sprache erwerben, erklärt werden. Es besteht nun weitestgehend Einigkeit darüber, dass die Zahl an möglichen Hypothesen zu groß ist.²⁸ Das lässt sich mit mathematischen Überlegungen der Kombinatorik leicht beweisen.²⁹ Weiterhin besteht Konsens darin, dass ein Sprachlerner nicht die Zeit hat, alle Hypothesen mit dem Sprachmaterial abzugleichen. Außerdem legen die Beobachtungen aus der Entwicklungspsychologie nahe, dass die Aneignung von sprachlichen Fähigkeiten im Zeitablauf ähnlich ist (Brown, 1973). Wenn nun unterschiedliche Hypothesen gewählt werden würden, könnte das Ergebnis zwar eventuell gleich sein, aber die sprachlichen Fähigkeiten bis zur vollendeten Entwicklung müssten ebenso unterschiedlich sein wie die verwendeten Hypothesen. Und dies ist mit den Beobachtungen der ontogenetischen Entwicklungen der Kleinkinder unvereinbar.

Will man das Phänomen erklären, ohne auf die angeborenen Sprachmechanismen zurückzugreifen, stellt sich automatisch die Frage, ob das Kleinkind überhaupt Hypothesen verwendet oder ob die Strukturen der Sprachen nicht doch andere Anhaltspunkte beinhalten, welche mit allgemeinen kognitiven Fähigkeiten extrahiert werden könnten und zumindest die möglichen Hypothesen stark eingrenzen würden. Akzeptiert man die „Hypothesentheorie“ ohne weitere Anpassungen, ist der Schluss zwingend, dass ohne vorgegebene Grenzen und *negative evidence*, das Problem des Spracherwerbs nicht zu bewältigen ist. Würde aber der Sprachlerner überhaupt keine Hypothesen bilden, sondern die Regelmäßigkeiten des Inputs mittels genereller kognitiver Prinzipien filtern, hätte man eine plausible Erklärung gefunden, welche die angeborenen sprachspezifischen Einstellungen vermeidet. Allerdings ist dies nicht weniger kompliziert, weil nun die Regelmäßigkeiten des Inputs mit einer konkreten kognitiven Fähigkeit in einen direkten Zusammenhang gebracht werden müssen.

²⁸vgl. Pinker (1989, 5).

²⁹vgl. Niyogi (2006, 53f).

1.4 Lösungsansatz und Genese der Idee

Wie noch in Kapitel 3.2.4 ausführlich darzulegen sein wird, muss der statistische Ansatz Vorrang vor allen anderen bekannten Segmentierungsansätzen genießen, weil er als einziger kein anfängliches einzelsprachspezifisches Datenmaterial unterstellt (Thiessen und Saffran (2003); Yang (2004)) und damit die Gefahr eines *circulus vitiosus* minimiert wird. Da kein anfängliches Datenmaterial in den Köpfen der Säuglinge vorhanden sein kann – schließlich ist es unstrittig, dass die genauen Inhalte des Lexikons nicht genetisch kodiert sind – können allophonische, metrische und phonotaktische Prozesse nur zeitlich verzögert in die Segmentierung eingreifen, da sie alle eine anfängliche Wortliste zur Induzierung der unterschiedlichen Muster von spezifischen Sprachen benötigen. Sie müssen also der statistischen Verarbeitung nachgelagert sein. Vermengt man die zeitliche Abfolge und stellt die prosodische Segmentierung beispielsweise vorab, entsteht automatisch das *Bootstrapping*-Problem in Form eines Zirkelschlusses, weil sich sofort die Frage stellt, woher das Neugeborene denn wissen kann, wie Wortende oder -anfang und die spezielle rhythmische Figur zusammen passen.

Der statistische Ansatz wurde für die Wortsegmentierung mit den Übergangswahrscheinlichkeiten formuliert (Saffran et al., 1996b). Man geht davon aus, dass sich die Repräsentationsentitäten (Silben, Diphone, Segmente, phonetische Eigenschaften), aus denen sich bedeutungstragende Einheiten (Morpheme) zusammensetzen, durch ihre Stellung zueinander definieren lassen, weil sie über Wortgrenzen in einem anderen Umfeld auftreten. Durch das allgemeine kognitive Prinzip des Vergleichs und im weiteren Sinne der Analogie können selbst Säuglinge erste sprachmentale Strukturen mit den gegebenen Häufigkeiten bilden. In dieser Arbeit wurden Phoneme als grundlegende Bausteine gewählt. Hinsichtlich dieser Anpassung lässt sich der statistische Ansatz nun auf die Problemstellung übertragen und mit folgender These formulieren. Alle Phonemkombinationen innerhalb eines Wortes haben auf Dauer und im Durchschnitt eine andere Vorkommenshäufigkeit als Phonemkombinationen, die über Wortgrenzen hinweg reichen. Damit erfüllen die Phonemkombinationen, die am häufigsten in Relation zu allen anderen möglichen Phonemkombinationen einer Sprache vorkommen, unterstützende Segmentierungsfunktionen. Hiermit könnte ein Zusammenhang zwischen der Vorkommenshäufigkeit von bestimmten Phonemmustern und deren sprachlichen Funktionen hergestellt werden.³⁰

Die Intuition der These lässt sich anschaulich aus dem Blickwinkel der Kybernetik betrachten.³¹ Danach wäre die natürliche Sprache ein geordnetes System von Zeichen. Da die Sprache ein geordnetes System ist, muss es ein Ordnungsprinzip geben, nach welchem

³⁰Diese Häufigkeitseffekte wurden bereits für unterschiedliche lexikalische Repräsentationen aufgrund einzelner Phoneme (flaps und /t/, /d/) untersucht und bestätigt (Connine, 2004).

³¹So wie sie von Wiener (1949) in direkter Beziehung zu Kommunikation von Tieren und Maschinen vorgeschlagen wurde. Die Anwendung der dynamischen Systemtheorie auf die natürliche Sprache klingt in Ansätzen bei Wildgen (1982) und Wildgen (1985) an. In den Vordergrund rückt er aber die Frage der Semantik und versucht sie mit den Fragestellungen der Katastrophentheorie und der Selbstorganisation (Wildgen und Laurent, 1987) zu erklären.

1 Einleitung

sich Strukturen ausrichten. Das Ordnungsprinzip zur Bildung von sprachlichen Strukturen muss auf bestehende Regelmäßigkeiten in den Sprachbestandteilen zurückgreifen. Gibt es keine Regelmäßigkeiten in den Sprachbestandteilen, so kann keine Ordnungsstruktur gebildet werden. Eine Ordnung entsteht nur durch die Bildung von Einheiten, die mindestens eine gemeinsame Eigenschaft besitzen (vgl. Bischof, 1995, 12ff).³² Ohne gemeinsame Eigenschaften existieren alle Bestandteile eines Systems im Chaos. Die Einheiten mit gemeinsamen Eigenschaften sind also die Regelmäßigkeiten von Sprachbestandteilen, auf denen sich eine Ordnung gründet. In der Linguistik selbst werden solche Ordnungsstrukturen mit Mustern (*patterns*) umschrieben.³³

Die Ausprägungen dieser Ordnungsstrukturen sind sicherlich vielfältig. Sprachliche Kategorien können als Ordnungsstrukturen verstanden werden, die sich im Weltwissen und in den sinnlichen Wahrnehmungen abbilden, wobei zum Beispiel Abstraktes und Konkretes durch die sprachliche Kategorie der Substantive zum Ausdruck gebracht werden. Die Gesetzmäßigkeiten innerhalb einer solchen Kategorie werden größtenteils auf alle Mitglieder gleichermaßen angewendet. So werden Deklinationen bei Substantiven nach dem gleichen Muster in einem Sprachsystem vollzogen. Auftretende Ausnahmen kann man meist mit einer verfeinerten abgeleiteten Gesetzmäßigkeit beschreiben. Ebenso verhalten sich die Spracheinheiten, welche die Beziehungen unter den Gegenstandsbereichen regeln – die Verben beziehungsweise in vielen Sprachsystemen auch Einheiten, welche andere Einheiten näher beschreiben – die Adjektive und die Adverbien.

Eine zusammenhängende Theorie, die die Sprache hinsichtlich ihrer stochastischen Gesetzmäßigkeiten beschreibt, steht nach Kenntnis des Autors nicht zur Verfügung.³⁴ So erscheint es sinnvoll, bei den kleinsten bedeutungsunterscheidenden Sprachbestandteilen zu beginnen, den Phonemen. Im Lichte des systemtheoretischen Hintergrundes und in Bezug auf die Fähigkeiten menschlicher Kognition tritt zuerst die Frage auf, nach welchen Prinzipien sich die kleinsten relevanten Sprachbestandteile zu höheren Strukturen zusammenfinden und wie man diese Prinzipien äußerlich erkennen kann. Welche Erkenntnisse lassen sich also aus den Mustern zusammengesetzter Phoneme entnehmen? Systemtheoretisch muss ein Muster immer eine Regelmäßigkeit zu Grunde liegen. Regelmäßigkeiten in einer konstant gehaltenen Umgebung lassen sich durch einfaches Abzählen quantifizieren. In Abhängigkeit von der Umgebung erhält man so eine Häufigkeitsverteilung von unterschiedlichen Phonemverbindungen.

³²Wittgensteins Sprachspiele zeigen uns allerdings, dass diese Behauptung auf semantischer Ebene zu Problemen führen könnte.

³³Sprachen enthalten immer Muster, auch wenn diese aufgrund der Vielfältigkeit der Sprachen äußerlich nicht leicht zu erkennen sind. Gleichmaßen dient Sprache der Kommunikation und erfüllt somit immer einen Zweck (vgl. Householder, 1971, 25).

³⁴Das Zipfsche Gesetz (Zipf (1929) und Zipf (1935)) aus der quantitativen Linguistik könnte man der Phonologie und teilweise dem Lautwandel in der diachronischen Sprachwissenschaft zuordnen. Eventuell wäre im weiteren Sinne auch das Bradfordsche Gesetz in die Diskursanalyse integrierbar. Derzeit versucht auch Haspelmath (2008) eine zusammenhängende Theorie über syntaktische Strukturen und deren Verwendungshäufigkeit aufzustellen. Definiert man die Mathematik als die Wissenschaft der Mustererkennung (nach Devlin, 2004), dann könnte man in der Mathematik möglicherweise eine Theorie in abstrakter Form finden, welche aber sicher nicht auf die spezifischen Eigenschaften der Sprache ungeändert übertragbar wäre und deshalb für die Linguistik keine zusammenhängende Theorie ersetzen kann.

Die These der Arbeit hat ihren Ursprung in den vorangestellten Überlegungen der Systemtheorie, den Ansätzen des datengetriebenen Lernens der Computerlinguistik³⁵, der Arbeit von Harris (1954), Harris (1955), Harris (1967), der Entwicklungspsychologie mit den Arbeiten von Saffran et al. (1996a), Saffran et al. (1996b), Newport und Aslin (2000), Saffran (2002), Saffran (2003b), Saffran (2003d), Jusczyk et al. (1993a), Jusczyk et al. (1993b) und Jusczyk et al. (1994) und der Linguistik (Bybee und Hopper (2001); Langacker (1990); Langacker (1999)). Aus der Vermengung dieser Ideen hat sich eine Reihe von vorgängigen Fragestellungen ergeben, die im späteren Verlauf ihren Niederschlag zu den bereits vorgestellten Problemstellungen gefunden haben. Welche Erkenntnisse lassen sich aus häufigen und inhärent versteckten Phonemverbindungen einer Sprache entnehmen? Welche Rolle spielen sie beim Erstspracherwerb? Wie wirkt sich die Häufigkeit von bestimmten Phonemverbindungen auf die Sprachkompetenz aus? Wie können diese Informationen von Maschinen genutzt werden?

Im Folgenden wird die schrittweise Entwicklung dieser Gedanken bis zur Idee für die Forschungsthese kurz aufgezeigt. Die Forschung von Bortfeld et al. (2005) zeigt, dass bekannte Namen³⁶ von sechs Monate alten Kleinkindern zu einer ersten Strukturierung des Sprachstroms *top-down* genutzt werden. Saffran et al. (1996b) haben Untersuchungen durchgeführt, die eine statistische Verteilung von silbenähnlichen Konstrukten zur Erkennung von Wortgrenzen (*bottom-up*) nahe legen. Kent und Miolo (2004) betonen die hohe Bedeutung der Häufigkeitsverteilung bei der Bildung von phonologischem Wissen beim Erstspracherwerb. Spezifischer gehen Meschyan und Hernandez (2002) auf den Zusammenhang zwischen dem Alter im Verlauf des Spracherwerbs und der Vorkommenshäufigkeit bestimmter Wörter ein, indem sie eine Korrelation zu der Geschwindigkeit und der Genauigkeit beim mentalen Abruf der Wörter nachweisen können. Dieses Resultat ist schon bei Carroll (1976) anzutreffen, der die Einflussgröße des Alters dem der Vorkommenshäufigkeit eines Wortes gegenübergestellt hatte und ebenso zu dem eindeutigen Schluss kam, dass zwar das Alter beim Erlernen der Lexik entscheidender sei als deren Häufigkeit, trotzdem müsse aber der Häufigkeit großer Einfluss auf die Latenzzeit zugestanden werden. Saint-Aubin und Poirier (2005) argumentieren, dass für den erwachsenen Lerner die Bekanntheit und die Häufigkeit eines Wortes korrelieren. Sendlmeier und Sendlmeier (1991) können dies für die Lautproduktion ab dem achten Monat belegen. Auch hier sind ähnliche Ergebnisse in der Literatur schon früher dokumentiert (Mutter und Hashtroudi, 1987).

Hieraus entsteht zumindest die Vermutung, dass diese Eigenschaft nicht nur auf Wörter und Silben begrenzt ist, sondern auch auf andere Phonemkombinationen Anwendung finden könnte. Bereits weit davor hat Harris (1954) und Harris (1955) auf die distributionalen Strukturen der Sprache aufmerksam gemacht und später mittels einer Computersimulation konkrete Beweise für eine Wortsegmentierung mittels der Phonemverteilung geliefert (Harris, 1967).

³⁵innerhalb der künstlichen Intelligenzforschung, Kap. 2.3.

³⁶Diese Namen kann man als Phonemkombinationen ansehen.

1 Einleitung

Diesen Untersuchungen ist gemein, dass sie Phonemverbindungen beziehungsweise auch deren Häufigkeiten als erste Orientierung für eine Strukturierung der Sprache anerkennen. Da in einem System Strukturen höherer Ordnung auf Strukturen niedriger Ordnung aufbauen, kommt der anfänglichen Strukturierung der Sprache große Bedeutung für die weitere Sprachentwicklung zu. Damit ist zumindest die Möglichkeit denkbar, dass genau die Phonemverbindungen, welche die anfängliche Ordnung herstellen, weitere Funktionen in der Sprache erfüllen. Wenn sie als wichtige Eckpfeiler in der Sprachentwicklung dienen, und sei es nur versteckt, so müssen sie auch der Sprache funktional erhalten bleiben, damit sie diese Orientierungsfunktion auch zukünftig ausüben können.

Von dieser Betrachtungsebene aus scheint die letztgenannte Fragestellung von den in Kapitel 1.1 beschriebenen abzuweichen. Es handelt sich jedoch um genau denselben Untersuchungsgegenstand, nämlich um vorhersagbare Regelmäßigkeiten in der Sprache. Mit einer bewiesenen These wird damit auch plausibel gemacht, dass durch die Häufigkeit ganz bestimmter Phonemverbindungen eine wichtige Funktion (genauer eine Segmentierungsfunktion) in der Sprache selbst kodiert ist.

Beim Versuch, Maschinen mit natürlicher Sprache auszustatten, kann derzeit ein rein regelbasierter Ansatz nur teilweise das Wesen einer natürlichen Sprache erfassen, da für keine natürliche Sprache eine vollständige Grammatik formuliert werden kann (vgl. z.B. Pinker (1994); Arbib (2003, 42f); Klabunde (1998, 36)). Eine explizite Programmierung von Regeln ist somit nicht vollständig möglich. Es bleibt abzuwarten, ob die viel gepriesenen unifikationsbasierten Grammatiken (vgl. Klenk, 2002, 181) den hohen Anforderungen natürlicher Sprache mit regelbasierten Algorithmen gerecht werden. Auf jeden Fall ist bei einer Maschine unstrittig, dass sie in ihrem Chip kein kognitives Modul für eine Sprache oder einen *language device* versteckt hält. So eignen sich Computersimulationen bestens, um Fragestellungen nach dem Einfluss der Stimulusarmut im Spracherwerb zu untersuchen.

1.5 Vorgehensweise und Aufbau

Die Vorgehensweise dieser Arbeit ist eingebettet in den Kreislauf einer psycholinguistischen Computermodellierung, wie sie der allgemeinen Vorstellung dieses Wissenschaftszweiges entspricht (Abbildung 1.1). Eine erste noch recht unspezifische Theorie, oft auch nur eine etwas genauer formulierte Hypothese oder gar plausible Vermutung, spezifiziert die Rahmenbedingungen und Annahmen eines Modells, welche dann algorithmisch in einer Simulation programmiert werden können.³⁷ Deren Ergebnisse bestätigen dann die eingangs angestellten theoretischen Vorhersagen oder erzwingen weitere Annahmen und Änderungen des Modells. Möglicherweise muss die erste Theorie vollends verworfen werden. Bestätigt die Simulation aber die Grundannahmen der Theorie, kann sie mit recht ressourcenintensiven Experimenten direkt bewiesen beziehungsweise widerlegt werden. Stimmen diese Befunde nun abermals mit den Simulationsergebnissen überein, dienen

³⁷ vgl. auch Brent (1997, 7).

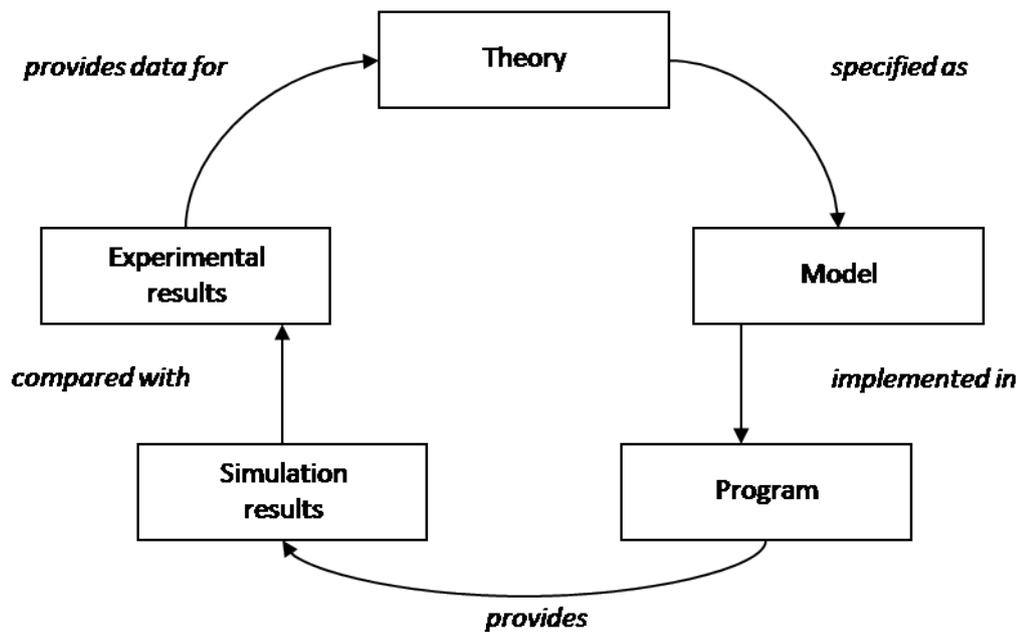


Abbildung 1.1: Psycholinguistische Computermodellierung,
(Quelle: Dijkstra, 1996, 5)

beide Resultate – Experiment und Simulation – einer synergetischen Spezifizierung der Theorie.

Die Problemstellungen (Kap. 1.1) können nun auf dieses Grundgerüst übertragen werden. Dies kann natürlich nicht ohne Anpassungen geschehen. Die Ergebnisse der Fragen (1)-(3) im Kapitel 1.1 bauen hinsichtlich der Durchführung der Simulation aufeinander auf und wurden deshalb bewusst in dieser Reihenfolge vorgestellt. Bei der schriftlichen Zusammenstellung der Arbeit und in Bezug auf den Kreislauf einer psycholinguistischen Computermodellierung (Abbildung 1.1) zeigte sich aber, dass die Befolgung dieser Reihenfolge einer verständlichen und nachvollziehbaren Beschreibung abträglich ist und eher Verwirrung stiften kann. Die logisch richtige Reihenfolge der Simulationsdurchführung lässt sich nicht vollständig auf den Aufbau der Arbeit übertragen. Es hätte zu einer unübersichtlichen Vermengung von theoretischen Erkenntnissen und simulierten Ergebnissen geführt. So steht die theoretische Abhandlung über die Fähigkeiten der Sprachlerner zur Beantwortung der dritten Problemstellung (Kap. 1.1) bereits am Anfang, weil man mit dem theoretischen Hintergrund des aktuellen Forschungsstandes die Simulation besser verstehen kann.

Die experimentellen Ergebnisse und die daraus abgeleiteten Theorien (Kap. 3 und 4) sind aus der Entwicklungspsychologie sowie Psycholinguistik bereits dokumentiert und gehen teilweise als Randbedingungen in die Simulation ein. Folglich kann die Arbeit an einem Modell für die Wortsegmentierung beginnen (Kap. 5). Die zentrale Entscheidung, die in dieser Entwicklungsphase des Modells ansteht, ist, welcher Abstraktionsgrad angemessen wäre. Einerseits dürfen keine Tautologien und andererseits keine ebenso komple-

1 Einleitung

nen Vorgänge wie die Realität selbst modelliert werden. Bei beiden Extremen konvergiert der Erklärungsgehalt gegen Null: Bei ersterem, weil es allgemein alles, also auch das, was die Theorie widerlegen würde, beweist und bei letzterem, weil die modellierten Vorgänge aufgrund der hohen Komplexität immer noch nicht verstanden werden können. Ein gutes Modell muss die Theorie also genau so spezifizieren und die relevanten Faktoren aus der Realität identifizieren, die zur Erklärung des beobachteten Phänomens minimal hinreichen.

Im fünften Kapitel wird der Algorithmus unter dieser Prämisse beschrieben und deren programmiertechnische Umsetzung dargelegt. Dadurch wird der Bezug zu den experimentellen Befunden hergestellt, um einer erklärenden Theorie näher zu kommen.³⁸ Das anschließende Kapitel 6 stellt schließlich die Ergebnisse vor und diskutiert sie kritisch hinsichtlich ihres Erkenntnisgewinns unter Rückgriff auf die entwicklungspsychologischen Experimente (der Kap. 3 und 4) und die gesetzten Ziele in Kapitel 1.1. Das Schlusskapitel greift die Argumentation zur Stimulusarmut auf, diskutiert sie unter Berücksichtigung der erbrachten Simulationsergebnisse und eröffnet schließlich Perspektiven für weitergehende Forschung und deren gesellschaftlich-wirtschaftliche Nutzung.

Auf diese Weise wird die Theorie der Wortsegmentierung mittels der Übergangswahrscheinlichkeiten detaillierter formuliert und in ihrer derzeitigen Ausprägung bestätigt. Für eine mögliche Erweiterung dieser Theorie kann vorgeschlagen werden, dass die Existenz einer quasi existenten Lexikonstruktur, die sich durch die stochastische Wortsegmentierung herausbildet, eine Ausgangsbasis für weitere Regelableitungen schaffen würde.

1.6 Einordnung der Arbeit in das interdisziplinäre Wissenschaftsumfeld

Der interdisziplinäre Charakter der vorliegenden Arbeit lässt sich am einfachsten durch die Schnittmenge der drei Wissenschaftsdisziplinen Linguistik, Psychologie und Computerwissenschaften, veranschaulichen (Abbildung 1.2).³⁹ Sie entspringt damit einem

³⁸Für die Simulation werden die typischen Komponenten von Lernmodellen übernommen, die Pinker (1984, 13) leicht verständlich wie folgt zusammenfasst:

IN THE MATHEMATICAL study of language learnability, a learning model is said to have four parts: a characterization of the class of languages within the scope of the learner's acquisition powers, one of which is the „target language“ spoken by the adult community; an input sample containing the information that the learner uses to learn; a learning strategy, describing what computations the learner performs on the input data; and a success criterion that the learner has to meet in order for us to conclude that his or her learning strategy works.

In der vorliegenden Arbeit entspricht die Zielsprache (Output, *target*) dem Englischen oder spezifischer englischen Wörtern. Der Input ist kindgerichtete Sprache. Die Lernstrategie lässt sich mit Übergangswahrscheinlichkeiten und Rekursion umschreiben. Letztlich übernimmt die Erfolgsmessung das F_1 -Maß.

³⁹Ähnlich wie die Darstellungen des *Sloan Reports* in Harnish (2002, 6f) oder Gardner (1989, 47ff).

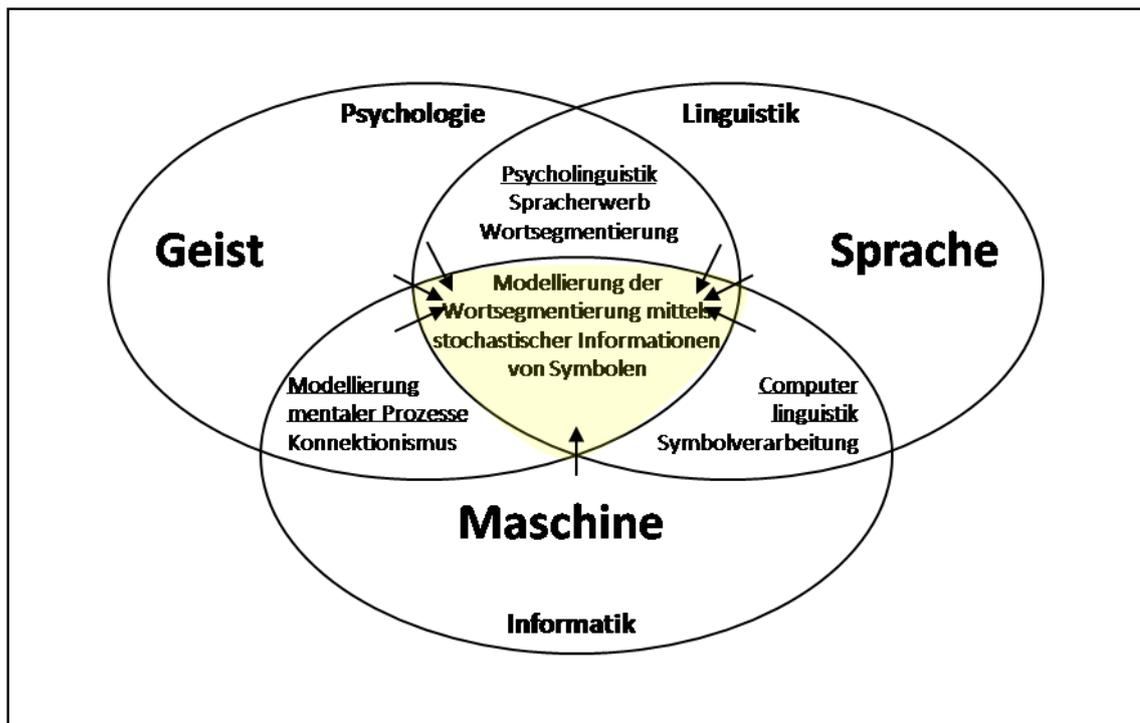


Abbildung 1.2: Das Wissenschaftsumfeld im Rahmen der Dissertation, Quelle: Eigene Darstellung

Mixtum Compositum der traditionellen Einteilung von Geistes-, Sozial- und Ingenieurwissenschaften. Die wohlbekannte Schnittmenge aus Linguistik und Psychologie, die Psycholinguistik, bildet den Schwerpunkt der Arbeit, innerhalb dessen die Forschungsthese für die Wortsegmentierung im Erstspracherwerb entstanden ist und welchem der Erkenntnisgewinn der Arbeit letztendlich zugeordnet werden muss. Die Computerwissenschaften ermöglichen jedoch die Umsetzung der Forschungsthese, weil erst aufgrund ihrer Werkzeuge und des operativen Wissens die Simulation durchgeführt werden kann. Speziell gehen aus der Überlappung von Psychologie und Informatik Grundannahmen des Konnektionismus zur Modellierung von mentalen Prozessen hervor, und mengentheoretische Überlegungen der Symbolmanipulation fließen aus der Linguistik-Informatik-Vereinigung ein. Die Linguistik trägt direkt mit umfangreichen Konzepten und sprachwissenschaftlichem Wissen über Laut- und Wortstrukturen bei. Aus der Psychologie entstammt die Idee des statistischen Lernens, welches von der Informatik aus mit dem datengetriebenen Lernen weiter gefasst wird und in die Algorithmik Einzug hielt. Im Übrigen tangieren darüber hinaus Bereiche der ingenieurstechnischen Signalverarbeitung und der Neurowissenschaften leicht die Argumentation in einigen Teilen der Arbeit. Da diese aber nur Randbereiche umfassen, sind sie nicht explizit als Disziplinen aufgeführt. Somit steht die Arbeit im Zentrum der drei Einflüsse aus Sprache, Geist und Maschine.

Die computerlinguistischen Ansätze erklären meist nicht direkt das Phänomen des Spracherwerbs, dessen Prozesse und mögliche Repräsentationen, wie sie tatsächlich be-

1 Einleitung

obachtbar sind, sondern stellen die effiziente informationstechnische Verarbeitung von beobachtbaren Spracheigenschaften für Maschinen in den Vordergrund (vgl. Rissanen, 1989).⁴⁰ Dennoch können auch diese Modelle als gesonderte Erklärungsansätze für die Spracherwerbsforschung herangezogen werden. Schließlich ähneln sich die Voraussetzungen und die Prozesshaftigkeit. Nur das primäre Erkenntnisobjekt unterscheidet sich. So steht das datengetriebene Lernen der Informatik im Ansatz dem empiristischen Paradigma nahe, weil auch hier aus dem Datenmaterial selbst Ableitungen getroffen werden. Wie im empiristischen Paradigma geht man davon aus, dass der Input der Zielsprache so viel Informationen enthalten muss, dass daraus eine nahezu vollständige Grammatik entwickelt werden kann. Der Begriff spielt auf die Fähigkeit an, selbstständig Informationen aus einem gegebenen Input zu verarbeiten. Hier ist eine offensichtliche Parallele zum statistischen Lernen der Psychologie gegeben, nur mit dem Unterschied, dass die Verhaltenswissenschaft den Menschen in den Mittelpunkt stellt und sich daran ausrichtet.

Damit wird klar, dass sich die vorliegende Arbeit in den Dienst des statistischen und des datengetriebenen Lernens stellt, bei denen die mentalen Prozesse des Spracherwerbs durch das Verarbeiten statistischen Wissens erklärt und maschinell simuliert werden. Es wird hier nicht der herkömmliche Ansatz des Experiments gewählt, so wie er für die psychologisch geprägten Wissenschaften typisch ist, sondern es wird die computerlinguistische Herangehensweise übernommen, ähnlich wie sie schon von Harris (1954), Harris (1967), Hayes und Clark (1970) und Wolff (1975b)⁴¹ Anwendung fand, später dann von Goldsmith (2001), de Marcken (1996) oder Swingley (2005) aufgenommen wurde. Bei dieser Methodik wird eher danach gefragt, was theoretisch unter bestimmten Voraussetzungen möglich ist, ohne den Beweis experimentell zu erbringen.⁴² Ausgehend vom empiristischen Ansatz und dem datengetriebenen Lernen lässt sich also festhalten, dass die notwendigen Informationen aus der äußeren Form der Sprache gewonnen werden.

Im Paradigma des datengetriebenen Lernens laufen ebenso Schwerpunkte aus den oben genannten Wissensdisziplinen zusammen. Erstens geschieht der Informationsinput durch textlich-visuelle oder auditive Zeichensysteme. Damit bestätigt die erste Wissensdisziplin, die Linguistik, nochmals ihre Existenzberechtigung. Prinzipiell erklärt und beschreibt sie die zeichentheoretischen und möglicherweise sprachlichen Phänomene, welche dann informationstechnisch verarbeitet werden. Somit tut sich die Informatik als zweite Disziplin als ein Helfer auf operativer Ebene hervor. Während die Linguistik Sprachstrukturen aufdecken muss, sichert die Informatik die Übersetzung dieser linguistischen Strukturen für Apparate. Wird natürliche Sprache verarbeitet, grenzt die Informatik

⁴⁰Ausnahmen sind die Arbeiten von de Marcken (1996), der zwar auch im Geiste der MDL höchste Priorität der effizienten Datenverarbeitung schenkt, aber dennoch seinen Lernalgorithmus direkt auf den Spracherwerb zuschneidet oder Swingley (2005).

⁴¹Die Dissertation von Wolff (1975a) war leider nicht beschaffbar. Der Titel *The acquisition of linguistic structures* weist auf einen sehr hohen Bezug zum Thema dieser Arbeit hin.

⁴²In computerlinguistischen Ansätzen beschränkt die Abbildungsgenauigkeit zwischen dem zu simulierenden Sprachphänomen und dem tatsächlichen Algorithmus die Aussagekraft des Ergebnisses. Sind die Voraussetzungen eindeutig vorgegeben und können eindeutig in einem Algorithmus formuliert werden, lässt sich die Richtigkeit dieses Algorithmus meist mathematisch beweisen.

ihren Zuständigkeitsbereich an der Schnittstelle zwischen Mensch und Maschine ab. In diesem Fall müssen die kognitiven Wissenschaften als dritte Disziplin hinzutreten und die menschlichen Verarbeitungsmechanismen von Informationen zu Wissen erklären. Die Informationswissenschaften treten dann nicht nur auf rein operativer Ebene zu Tage, sondern müssen prüfen, in wie weit sich diese Mechanismen auch informationstechnisch verarbeiten lassen.

Die Schaffung von intelligenten Sprachsystemen innerhalb der Computerlinguistik⁴³ ist ein komplexer Prozess, der mindestens durch die hier genannten Forschungsdisziplinen abgedeckt wird. Die Analyse und Erkennung der Besonderheiten des Informationsinput werden von Teilgebieten der Linguistik dokumentiert. Deren informationstechnische Repräsentation und systemhafte Verarbeitung gehört zu den Informationswissenschaften. Schließlich zeigt die Psychologie noch Erklärungsansätze menschlicher Lernprozesse auf, die maschinell simuliert werden können. Handelt es sich beim Informationsinput also um natürliche Sprache, sind die kognitiven Wissenschaften zwingend eingebunden, da natürliche Sprache eine kognitive Fähigkeit ist und somit auch in den Erklärungsbereich der Psychologie fällt.

Menschliches Lernen als Vorbild für maschinelles Lernen zu nehmen, ist umstritten – bieten doch die Maschinen und Apparate derzeit nur sehr begrenzte Möglichkeiten, die Vielschichtigkeit und das Ineinandergreifen menschlicher kognitiver Prozesse ausreichend abzubilden.⁴⁴ Deshalb werden stets nur modellhafte, also vereinfachte Annahmen getroffen, die dem Zusammenwirken des Systems als Ganzem nicht gerecht werden. Andererseits ist der Mensch das einzige wissensverarbeitende Lebewesen höherer kognitiver Leistungen und so macht es Sinn, die ihm zu Grunde liegenden Prozesse nachzubilden und auf die Möglichkeiten der Maschinen zuzuschneiden.⁴⁵

⁴³Weil es hierbei um natürliche Sprache geht, kann man im engeren Sinne von der Computerlinguistik sprechen.

⁴⁴vgl. aber hierzu kritisch Lappin und Shieber (2007).

⁴⁵vgl. Wallmannsberger (1994, 37).

2 Nativismus und Empirismus als grundlegende Positionen in philosophiegeschichtlicher Betrachtung und angewandter linguistischer Forschung

2.1 Einführende Bemerkungen

In den folgenden Abschnitten soll zunächst eine Hinführung zum Thema in einem sehr breiten wissenschaftlichen Kontext gesehen werden. Eine konzise Einordnung der Arbeit erfolgt in mehrfacher Hinsicht. Zum einen wird der Versuch unternommen, die Nähe zwischen Nativismus und Symbolverarbeitung sowie zwischen Empirismus und Konnektionismus vor einem philosophiegeschichtlichen Hintergrund zu beleuchten; stets in dem Bemühen, einen sprachwissenschaftlichen Bezug herzustellen (Kap. 2.3). Diese Bezugnahme geschieht im Bereich einer kognitiven Wissenschaft nunmehr auch verstärkt in Hinsicht auf deren theoretische Hintergründe einer *representational theory of mind* (RTM), welche in jüngerer Zeit unter dem schon spezifischeren Begriff der *computational theory of mind* (CTM) verstanden wird und sich in den oben benannten Dualismus von *digital computational theory of mind* (DCTM) und *connectionist computational theory of mind* (CCTM) unterteilt (Harnish, 2002, 105 und 275). Dieser Hintergrund dient dann, zum anderen, als Struktur für die Positionierung der eigenen Arbeit in diesem Gefüge, dessen Bedeutung gerade durch das so verdichtete Gesamtbild in einem epistemologischen Kontext entsteht. Im Kapitel 1.6 ist bereits eine Einordnung in das engere wissenschaftliche Umfeld geschehen. Abbildung 2.1 ist eine Visualisierung der Tendenzen von Forschungsschwerpunkten, wie sie der historischen Zusammenfassung von Harnish (2002) oder in Teilen aus Newell (1990), Bechtel und Abrahamsen (1991, 3ff) entnommen werden können, und einigen neueren Ansätzen (vgl. Pospeschill, 2004).

Das große Interesse an der kindlichen Sprachverarbeitung lässt sich auf die Bedeutung von nachgelagerten Fragestellungen zurückführen. Man hofft, aus der Erklärung des Erstspracherwerbs eindeutige Hinweise auf die Repräsentation und Verarbeitung von Sprache im Allgemeinen zu finden (vgl. Mehler et al., 1998) und damit den anhaltenden Disput zwischen symbolischer (Nativismus) und assoziativer Repräsentation (Konstruktivismus/Epigenese beziehungsweise Konnektionismus) der Sprache beizulegen.¹ Diese

¹Die Position des Nativismus zu vertreten, bedeutet natürlich nicht zwingend, eine symbolische Verarbeitung von Sprache zu akzeptieren oder umgekehrt, wie es in der *Optimality Theory* (Smolensky und Legendre, 2006b, 65ff) als dem vielleicht derzeit bekanntesten Beispiel deutlich wird. In der sprachwissenschaftlichen Debatte unterstellen die nativistischen Ansätze aber in Mehrheit den symbolischen Charakter von Sprache und deren direkter Übertragbarkeit als mentale Repräsentationen.

2 Nativismus und Empirismus als grundlegende Positionen

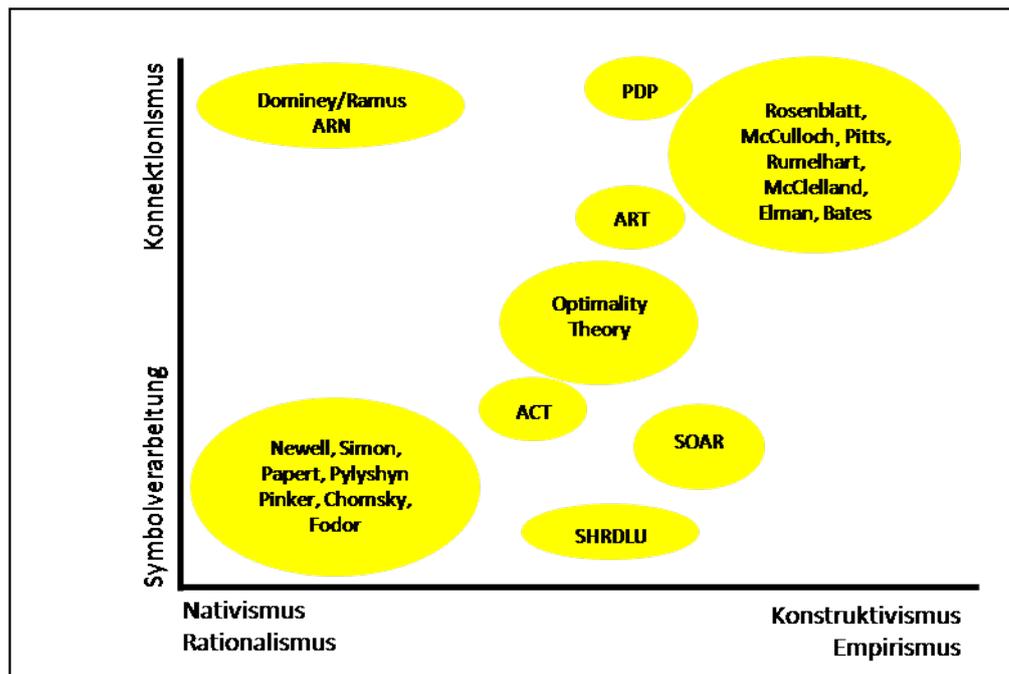


Abbildung 2.1: Mögliche Typologie vorherrschender Forschungsparadigmen innerhalb der CTM, Quelle: Eigene Darstellung

Denktraditionen sind in der aktuellen Debatte des Spracherwerbs führend, wenn auch nicht die einzigen Forschungsparadigmen.² Sie lassen sich auf ein philosophisches Grund-

Es ist natürlich ebenso denkbar, dass bestimmte Netzwerkarchitekturen, welche verteilte Informationen repräsentieren, in ihrem Wachstum genetisch kodiert, also angeboren sein können (vgl. z.B. Rumelhart und McClelland (1986a, 139); Karmiloff-Smith (1995, 165ff u. 179f); Keil (2001, 584) oder als konkrete Anwendung Dominey (2000)). Auf gleiche Weise lassen sich auch Mischformen im konstruktivistischen Paradigma finden, in dem verteilte Repräsentationen auch noch als komplexe Symbole anerkannt werden (Hybride Modelle oder vgl. Elman et al. (1997, 35ff)). Um einer begrifflichen Verwirrung entgegen zu wirken, werden hier die Teilmengen aus Nativismus und Konstruktivismus verstanden, welche sich eindeutig auch als symbolisch beziehungsweise subsymbolisch auffassen lassen und nicht in die eben benannten Zwischenbereiche fallen. Für den Empirismus ist mit dem Begriff des Konnektionismus bereits eine definitorische Abgrenzung mit nur sehr wenigen Ausnahmen gegeben, sodass im Folgenden dieser Terminus auch vorrangig benutzt wird und damit auch den Standards in der Literatur folgt (vgl. Jacobs, 2003, 127ff).

²Neben dem in der Philosophie verwurzelten Nativismus und Empirismus, lassen sich zwei weitere Erklärungsansätze ausmachen (vgl. z.B. Strömquist, 2003, 791f), welche in der Empirismus-Rationalismus-Dualität nur schlecht fassbar sind beziehungsweise nicht immer eindeutig zuzuordnen und deshalb in der Literatur etwas verwirrend erscheinen: Funktionalismus (siehe z.B. Karmiloff-Smith, 1979) und crosslinguistische/intratypologische Ansätze (Slobin, 1985-1997). Der Funktionalismus zielt insbesondere auf die Rolle der semantischen und pragmatischen Informationen ab. Kommunikation ist eingebettet in das soziale Umfeld und kann nicht auf phonologische und syntaktische Informationen beschränkt werden. Hier ist die Entwicklung von linguistischer Form und dessen Inhalt (die Funktion) innerhalb des soziokulturellen Einflusses von internen und externen Faktoren Gegenstand der wissenschaftlichen Auseinandersetzung. Aus dieser Sichtweise lässt er sich sehr gut mit dem empiristisch/konstruktivistisch geprägten Konnektionismus als Ausgangspunkt

satzproblem zurückführen, welches in seinen Ansätzen bereits in der griechischen Antike in Platons Menon (angeborene Ideen), Phaidon (Ideenlehre und Geist/Seele-Körper) und Phaidros (Herkunft der Ideen) anzutreffen ist,³ spätestens aber im 17. und 18. Jahrhundert von Descartes und Leibniz einerseits und Locke und Hume andererseits formal als Rationalismus-Empirismus-Debatte diskutiert wurde.⁴

Der philosophiegeschichtliche Hintergrund dieser Arbeit lässt sich im weiteren Sinne am besten anhand zweier entgegensetzter Extrempositionen innerhalb eines Kontinuums betrachten und manifestiert sich in einer Jahrtausende alten Grundsatzdebatte der abendländischen Philosophie über die Natur des menschlichen Geistes, die sich ungelöst in der neuzeitlichen Erkenntnistheorie an den beiden Polen des Rationalismus und Empirismus wieder findet. So sieht auch Gardner (1989, 55) den „logischen Ausgangspunkt“ einer jeden kognitionswissenschaftlichen Forschung in den „philosophischen Fragestellungen der griechischen Antike und der Aufklärung“. Obwohl der Einfluss der Philosophiegeschichte auf diesem Gebiet umstritten ist, macht Gardner weiterhin deutlich, dass das bereits vorgedachte Wissen speziell in den Kognitionswissenschaften als Grundlage weiteren Denkens gewürdigt werden sollte. So wird in den folgenden Abschnitten ein kurzer Abriss über die bestimmenden philosophischen Theorien gegeben. Natürlich ist es nicht Sinn dieses Abrisses, sämtliche Theorien, die in irgendeiner Weise mit dem Thema im engeren oder weiteren Sinne beschäftigt sind, abzuhandeln. Ebenso wenig sollen die angeführten Theorien in ihrer Vollständigkeit tiefgründig diskutiert werden. Der Zweck der Ausführungen besteht lediglich darin, den philosophischen Ursprung der Arbeit zu skizzieren und deutlich zu machen, dass die grundsätzlichen Problemstellungen bereits längerer menschlicher Reflexion unterliegen; somit als Ausgangspunkt der hier dargelegten Überlegungen und der Einordnung der Arbeit ins weitere Wissenschaftsumfeld dienen dürfen.

Im engeren Sinne, d.h. bezogen auf die Problemstellung der Arbeit, besagt die empiris-

begründen (vgl. die Beiträge in Dirven (1987); darin insbesondere Bates und MacWhinney (1987, 209ff)). Ebenso deklariert nun auch die Symbolverarbeitungstheorie den Funktionalismus als dessen „philosophischen Überbau“, weil „mentale Zustände ... durch ihre kausale Rolle gegenüber anderen mentalen Zuständen und gegenüber dem Verhalten definiert“ sind (Scheerer, 1990, 2). Man kann den Funktionalismus also eher als gemeinsamen Vorfahren beider Theorien betrachten, der sowohl in der konnektionistischen als auch in der nativistischen Forschungstradition weitergeführt wird, zumal beide Paradigmen Modelle anstreben, welche algorithmisch beschreibbar sind. Hier verhält es sich wohl analog zu den interaktionistischen Modellen im Erstspracherwerb (siehe Kap. 1.2). Zudem kann man den Funktionalismus auch als Ausgangspunkt für die crosslinguistischen Ansätze, wie dem intratypologischen Ansatz von Slobin (1985-1997) ansehen. Crosslinguistische Ansätze untersuchen Gemeinsamkeiten und Unterschiede des Erstspracherwerbs in unterschiedlichen Sprachen. Entwicklungsstufen werden in allen Teilbereichen untersucht und verglichen. Gemeinsamkeiten über alle Sprachfamilien hinweg können Aufschluss über eine allgemeine kognitive Fähigkeit des Sprachlernprozesses geben. Die frühen Arbeiten favorisierten eine nativistische Erklärungsweise, während die späteren Arbeiten eher an die Sapir-Whorf-Hypothese erinnern, den Aspekt des Lernens sehr viel deutlicher betonen und deshalb schon eher den empiristischen Modellen zuzurechnen wären.

³Ein weiterer Zusammenhang zwischen dem abgeleiteten Formalismus/Funktionalismus und den antiken Analogisten/Anomalisten wird in Bates und MacWhinney (1987) gegeben.

⁴Diese Liste ließe sich sehr leicht mit Spinoza, Wolff und Kant sowie Bacon, Gassendi, Hobbes, Thomasius und Berkeley erweitern, was aber über einen einleitenden Gedanken hinausgehen würde.

2 Nativismus und Empirismus als grundlegende Positionen

tische Extremposition, dass statistische Methoden alle Phänomene der Sprache erklären können,⁵ die andere bestreitet jeglichen Erkenntnisgewinn verteilungsorientierter Methoden und bietet als Lösung das angeborene Wissen feil. Analog zu der philosophischen Debatte des Rationalismus und des Empirismus sind keine dieser absoluten Extreme tatsächlich in der Wissenschaft vertreten (vgl. Gordon und Slater (1998, 100); Ridley (2003, 196)). Forschungsergebnisse, die sich eindeutig innerhalb des Kontinuums wieder finden, werden häufig einem dieser Extreme im Nachhinein und teils von fachfremden Autoren zugeordnet (Scholz und Pullum (vgl. 2006); Pereira (2000, 1240)). Offenbar erfolgt diese Kategorisierung abermals nach tendenziellen Neigungen, also lediglich nach der Position im Kontinuum. Lässt sich die Mehrzahl der Argumente eher mit dem einen Extrem vereinen, spiegelt sich diese Tendenz in der wissenschaftlichen Debatte oft als eine Extremposition wider. Diesem Umstand verdanken wir eine gedankliche Vereinfachung der kognitionswissenschaftlichen Debatte, welche jedoch die Vielfalt individueller Ideen nicht wiedergeben kann und welche dann folglich in der Grundsatzdebatte untergeht.⁶ Gerade aus wissenschaftlicher Sicht ist hier der Zustand der mangelnden Differenzierung problematisch, weil sich die Diskussionsbeiträge eben vorrangig an den Extremen ausrichten, nicht aber die speziellen Argumente in ihrer genauen Konstellation berücksichtigen und in einer neuen Abwägung zwischen den Hauptargumenten entdecken lassen. Andererseits ist die vorherrschende Kategorisierung zur Erklärung des theoretischen Hintergrunds vorteilhaft, weil sie in vereinfachter Weise den gedanklichen Ursprung veranschaulicht, die Einflussgrößen und Grundvoraussetzungen übersichtlich darstellt sowie den Rahmen durch die Extrempositionen absteckt. Daher erscheint es sinnvoll, zunächst den theoretischen Hintergrund aus der Perspektive der Extrempositionen zu beschreiben und erst dann den eigenen Beitrag dazwischen zu positionieren.

2.2 Philosophische Positionen

Platons Menon (Platon, 2004, 411ff) wird in der kognitionswissenschaftlichen Debatte häufig zur Frage des angeborenen Wissens herbei zitiert (z.B. bei Gardner (1989, 15ff); Cowie (1999, 3); Pospeschill (2004, 9)).⁷ Er soll auch hier als Einstieg dienen, weil bei ihm die Problematik sehr einfach und deutlich hervortritt. Beim Dialog zwischen Sokrates und Menon über die Tugend wendet sich das Gespräch der Frage des Nicht-Wissens und Wissens zu. Für Letzteres erbringt Sokrates einen Beweis, indem er einen Sklaven von Menon über geometrische Eigenschaften eines Quadrats befragt. Es zeigt sich, dass

⁵Hierzu sei angemerkt, dass die subsymbolische Merkmalskodierung der Netze immer einer Verteilung unterliegt, welche man mit Hilfe der Statistik zugänglich machen kann.

⁶Nicht zuletzt aus diesem Denken sind so assoziationsreiche Namen wie die Ost-West-Debatte der Kognitionswissenschaften hervor gegangen. Die Namensgebung „East Pole – West Pole“ geht wohl auf Fodor zurück, der während einer Konferenz am MIT in einer Diskussion diese Namen prägte (Dennett in Pinker (2002, 35)).

⁷Selbst Leibniz hat einen Vergleich zwischen Platon als Rationalist und Aristoteles als Empirist in seinen *Nouveaux Essais* gleich zu Anfang angeführt (Leibniz (2000, 321) und darauf in den *Metaphysischen Abhandlungen* Bezug genommen in Leibniz (2000, 178)).

der Sklave über ein unbewusstes mathematisches Wissen verfügt, ohne dass dieses in den Geist des Sklaven von außen hinein gelangt sein kann. „Es waren also wohl diese Vorstellungen schon in ihm? Oder nicht?“ (Platon, 2004, 436) fragt Sokrates und Menon bejaht. Im weiteren Fortgang dieser Passage wird Sokrates vom geometrisch-mathematischen Wissen auf andere Wissensbereiche induzieren.

... wenn er sie [die Vorstellung über das mathematische Wissen] aber einmal bekommen hat, hat er sie doch wohl nicht erst im jetzigen Leben bekommen? Oder hat jemand ihm Geometrie gelehrt? Denn er wird das gleiche in der ganzen Geometrie leisten und in sämtlichen anderen Kenntnissen. (Platon, 2004, 436)

Demnach gibt es also Wissen, das der Mensch schon immer besessen hat. Wenn man aber dieses Wissen nicht erlernt hat, wo kommt es dann her? Wie ist es in den „Geist“ gelangt? Diese Frage ist seitdem im abendländischen philosophischen Denken präsent.⁸

Bei seinem Schüler Aristoteles ist wenig später schon zu lesen:

Es muss sich so verhalten wie bei einer Schreibtafel, auf der noch nichts in Wirklichkeit geschrieben steht, was bei der Vernunft zutrifft. (Aristoteles, 1995, 171)

Diese Aussage ist zunächst nur ein möglicher Vorschlag für das Erkennungsproblem der Vernunft, die von Aristoteles als leidensunfähig definiert wurde, deren Erkennung aber erleidend ist. Wie bringt man nun die immaterielle Vernunft mit einem Objekt in Verbindung? Aristoteles' Lösung ist die Seele, welche ihrer Natur nach zwar ohne Materie existiert, aber eine bestimmte Form des Erleidens kennt, weil sie das Objekt zuerst als Möglichkeit und dann in Wirklichkeit annimmt.⁹ Dieser Prozess ist vergleichbar mit einer Schreibtafel (*grammateion*), die zunächst leer ist und dann beschrieben wird.¹⁰

Später hat Descartes diese Fragestellungen wieder aufgegriffen. Er wird typischerweise als der Urvater der nativistischen Idee in der philosophischen Tradition des Rationalismus anerkannt.¹¹ Im dritten Teil seiner Meditationen beginnt der Mathematiker Descartes den für diese Arbeit auch interessanten Gedanken der eingeborenen Ideen explizit zu formulieren: „... aber dass solche Vorstellungen in mir seien, leugne ich auch jetzt nicht.“ (Descartes, 2006, 161) und bezieht sich in der Fortführung seiner Argumentation auch auf mathematische Vergleiche, wie es auch in Platons Menon schon geschehen ist:

Wie war es aber, wenn ich in der Arithmetik oder Geometrie irgend etwas ganz Einfaches und Leichtes betrachte, wie z.B. dass zwei und drei zusammen fünf ergeben, und was dergleichen mehr ist. (Descartes, 2006, 161)

Nur wenig später fasst er den Gedanken begründend zusammen:

Von diesen Vorstellungen sind nun dem Anscheine nach die einen mir angeboren, andere dagegen sind mir von außen gekommen, wieder andere von mir selbst gebildet. Denn wenn

⁸Sie kann durchaus schon sehr viel früher gestellt worden sein, konnte aber in dieser Klarheit nicht schriftlich fest gehalten werden, sodass Platons Menon auch hier als Ursprung dieses Gedankens gelten soll.

⁹vgl. Seidl (1995, 265).

¹⁰Hierzu muss man jedoch anmerken, dass die Positionen von Platon und Aristoteles nur sehr oberflächlich mit der Rationalismus-Empirismus-Debatte des 17. Jahrhunderts vergleichbar sind (vgl. Engfer, 1996, 237).

¹¹vgl. Engfer (1996, 55) für eine detaillierte Auflistung der philosophiegeschichtlichen Quellen.

2 Nativismus und Empirismus als grundlegende Positionen

ich erkenne, was das Ding ist, was die Wahrheit ist, was der Gedanke ist, so kann ich dies nur aus meinen eigenen Wesen geschöpft haben, nicht aber woandersher, wenn ich dagegen ein Geräusch höre, die Sonne sehe, das Feuer fühle, so habe ich bis jetzt angenommen, diese Vorstellungen kämen von gewissen Dingen außerhalb meiner, Vorstellungen wie Sirenen, Hippogryphen und ähnliches bilde ich mir selber. (Descartes, 2006, 164)

Descartes trifft hier schon Unterscheidungen, die sich dann später als zentral herausstellen sollten. Er sieht einmal die angeborenen Vorstellungen, weiß aber dazu noch die äußeren von den selbst gebildeten zu unterscheiden. Letzteres erinnert sehr stark an Lockes duale Aufschlüsselung von *sensation* und *reflection*, der inneren und äußeren Erfahrung. Offensichtlich geht diese Übereinstimmung nicht rein zufällig auf die Beobachtung zurück, dass aus der Erfahrung gewonnene, komplexere Sachverhalte einem weiteren internen Verarbeitungsprozess unterliegen, dessen Output dann nicht mehr direkt auf die einfachen Beobachtungen der Umwelt bezogen werden kann, sondern mehrere äußere Reize derart neu verknüpft und kombiniert, dass daraus etwas Neues entsteht. Sollte man diesen Mechanismus nun als eine innere Idee ansehen oder die internen Umwandlungen nach noch unerforschten Gesetzmäßigkeiten von unbewussten Erfahrungseinflüssen anerkennen? Wie differenziert Descartes die Problematik bereits erkennt, zeigt sich in der fünften Meditation. Er schreibt:

Am bemerkenswertesten ist dabei, daß ich unzählige Vorstellungen von Dingen in mir finde, die, auch wenn sie vielleicht nirgends außer mir existierten, doch nichts als bloßes Nichts bezeichnet werden könnten, und obwohl ich sie mir gleichsam nach Belieben denken kann, so sind sie doch nicht rein von mir erdacht, sondern haben ihre wahre und unveränderliche Natur. Wenn ich mir z.B. ein Dreieck vorstelle, so mag vielleicht eine solche Figur nirgends in der Welt außer in meinem Denken existieren oder je existiert haben, aber dennoch ist seine Natur durchaus bestimmt, seine Form unveränderlich und ewig und nicht von mir erdacht, auch nicht abhängig von meiner Seele, was schon daraus hervorgeht, daß von diesem Dreieck verschiedene Eigenschaften bewiesen werden können, wie daß seine drei Winkel gleich zwei Rechten sind, daß seinem größten Winkel die größte Seite gegenüberliegt, und was dergleichen mehr ist. Ob ich nun will oder nicht, ich muß das anerkennen, auch wenn ich früher bei der Vorstellung eines Dreiecks nicht daran gedacht habe und es mithin nicht von mir erdacht sein kann. (Descartes, 2006, 193)

Natürlich sei ihm die Flucht in die Mathematik gewährt. Die geometrischen Eigenschaften und Zusammenhänge von Dreiecken existieren also im Denken offensichtlich, ohne dass diese aus der äußeren Umwelt gelernt wurden. Sie sind nach Descartes aber auch nicht in jedem einzelnen Geiste aus sich selbst entstanden, obwohl sie möglicherweise nur dort existieren. Andererseits waren sie nicht schon von Anfang an da, weil die Vorstellungen nicht bewusst wahrgenommen werden konnten. Descartes bleibt nun nichts anderes übrig, als dieses Dilemma im Übersinnlichen zu lösen und genau dafür braucht er den Gottesbeweis. Damit beginnt Descartes einen Gedanken zu entwickeln, der womöglich für Leibniz größere Bedeutung für seine Monadentheorie gehabt haben könnte.

Im weiteren Verlauf der Meditationen lässt Descartes den Versuch nicht aus, die sinnliche Wahrnehmung wieder auf die Grundlage der inneren oder äußeren (wohl göttlichen) Vorstellungen zu stellen.

Nun besteht aber in mir ein passives Vermögen wahrzunehmen oder die Vorstellung sinnlicher Dinge zu empfangen und zu erfassen. Allein ich könnte gar keinen Gebrauch davon

machen, wenn nicht auch ein aktives Vermögen in mir oder außer mir vorhanden wäre, das jene Vorstellungen hervorruft oder bewirkt. (Descartes, 2006, 208)

Er sieht das passive Vermögen wohl in dem Automatismus unserer Sinnesorgane, Reize unabhängig vom Bewusstsein aufzunehmen und zu verarbeiten. Das aktive Vermögen muss dann die Fähigkeit sein, die Aufmerksamkeit zu steuern oder bewusst zu lenken. Der Versuch, die *Ratio* als die Grundlage für die Aufnahme von Sinnesreizen herzuleiten und ihr damit das Vermögen der geistigen Verarbeitung zuzuschreiben, folgt natürlich aus der Idee des Angeborens, welche nun wieder in der Wahrhaftigkeit Gottes begründet werden muss.

Leibniz sieht sich in der rationalistischen Tradition Descartes. In den *Nouveaux Essais* (Leibniz, 2000, 320ff) lässt sich sogar ein möglicher Anhaltspunkt für die Polarisierung der Debatte finden. Dort ordnet Leibniz gleich zu Beginn die Positionen von Rationalismus und Empirismus recht radikal sich selbst und Locke zu (Leibniz, 2000, 325ff). Dabei bekennt sich Leibniz auch zur „Unhintergebarkeit der Erfahrung“, weil er natürlich sonst einfache Wahrnehmungsphänomene, dessen Verarbeitung, Erinnerung und einfache Verknüpfung nicht erklären könnte. Für Leibniz existieren streng genommen aber immer nur innere Erfahrungen; also interpretierte Sinneseindrücke. Und er hält an der Erkennbarkeit der Vernunftswahrheiten jenseits der Erfahrung fest (Engfer, 1996, 174). Man muss jedoch einen klaren Unterschied zwischen den phänomenologischen Schriften und den philosophischen Darstellungen zur Monadologie erkennen, weil beide Ideen unterschiedliche Auffassungen des Erfahrungsbegriffs unterstellen (Engfer, 1996, 167). Im Mittelpunkt der philosophischen Ausführungen steht die Theorie von Monaden, deren Fensterlosigkeit, Individualität und Unteilbarkeit (Leibniz, 2000, 406ff). Die Fensterlosigkeit löst offensichtlich einen Widerspruch aus, weil eine Interaktion zwischen der Umwelt und dem Inneren eines Menschen in irgendeiner Form passieren muss. Dies gilt auch dann, wenn man annimmt, dass die Monade alles schon in sich hat. Dieses Alles ist aber „verworren“, also ungenau und nicht bewusst reflektierbar. Die Spezifizierung des Inneren übernimmt die Erfahrung. Ihr wird dafür eine Art vermittelnde Rolle zugestanden, die sodann den Widerspruch der Verkapselung der Monaden auflöst.

Zwar ist die Monade ohne Fenster und unterliegt keinen äußeren Einflüssen . . . , aber weil sie das meiste in sich selbst nur verworren erkennt, erkennt sie es dennoch bloß im Modus der Erfahrung. (Engfer, 1996, 167)

In den phänomenologischen Schriften, die er an John Locke adressiert, scheint Leibniz den Erfahrungsbegriff gemäß der damaligen Konvention zu gebrauchen. Allerdings resonieren die *Neuen Abhandlungen* auf einer religiösen Grundfrequenz und zunächst garnicht auf der erkenntnistheoretischen Fragestellung, die man als Leser dieser Schrift erwarten würde. Der entscheidende Streitpunkt sollten doch die eingeborenen Ideen und Prinzipien sein, auf welche sich die Monadentheorie gründet, und Lockes Auffassung über die Natur des menschlichen Verstandes. Die posthum erschienene Schrift führt kein schlagendes Argument an, das dem Argument der angeborenen Ideen einen Vorsprung in der Diskussion verschaffen würde.

2 Nativismus und Empirismus als grundlegende Positionen

John Locke selbst hatte zunächst versucht, das Argument Descartes der eingeborenen Idee zu entkräften, wurde aber gleichzeitig stark von Descartes Denken geprägt, sodass er sogar den Gedanken der Intuition übernimmt (Locke, 1987, 84ff und 530ff). In seinem *Essay Concerning Human Understanding* (Locke, 1987) argumentiert Locke, dass der Verstand bei der Geburt einem weißen Blatt Papier gleiche. Er stellt hier auch gleich zu Beginn klar, dass er nicht die Dinge, die dem menschlichen Verstand unzugänglich sind, diskutieren kann (Locke, 1987, 44-45) und proklamiert zunächst eine Gegenthese zu den angeborenen Prinzipien und Ideen. Er postuliert einfach deren Nichtexistenz. Also schreibt Locke zu Beginn seines zweiten Buches die Wurzel aller Erkenntnis der Erfahrung zu:

To this I answer, in one word, From Experience: In that, all our Knowledge is founded; and from that it ultimately derives itself. (Locke, 1987, 104)

Die Erfahrung geht dabei auf die reine Sinneswahrnehmung (*sensation*) und auf die innere Selbstwahrnehmung (*reflection*) zurück, die sich nun wiederum durch die Natur des Materiellen und der Bewusstseinsvorgänge selbst bestimmen.

These two, I say, viz. External, Material things, as the Objects of SENSATION; and the Operations of our own Minds within, as the Objects of REFLECTION, are, to me, the only Originals, from whence all our Ideas take their beginnings. (Locke, 1987, 105)

Im Fortgang des zweiten Buches spezifiziert er den Begriff der Idee als Vorstellungen des Bewusstseins, geht im dritten Buch auf die Wörter als Repräsentanten der Ideen ein und gelangt dann im letzten und vierten Buch zur Ausformulierung seiner Erkenntnistheorie. Dort kristallisiert sich der Wahrheitsbegriff als zentral heraus. Es geht im Grunde immer um einen Abgleich zwischen einzelnen Ideen. Wahrheit bedeutet, dass zwei Ideen übereinstimmen. Für Locke bezieht sich die Erkenntnis nicht direkt auf die Erfahrung, sondern auf die Wahrheit von Ideen, also deren Verhältnis zueinander. Diese Verhältnisse finden dann ihren Niederschlag in den drei verschiedenen Formen: intuitives, demonstratives und sensitives Wissen (Locke, 1987, 531f und 536). Diese Wissensformen wirken erst durch die Wahrnehmung des Zusammenhangs von Ideen als Wissen. Die Intuition als höchste Form der Erkenntnis ist die Un- oder Gleichheit von Ideen, die unmittelbar im Selbst geschehen. Womöglich kann man hier eine Parallele zu Descartes sehen, als er die Grundeigenschaften des Dreiecks erkannte. Bei einer demonstrativen Erkenntnis lässt sich das Verhältnis von zwei Ideen erst durch eine dritte Idee herstellen. Bei der Bildung von sensitivem Wissen lässt sich ein direkter Zusammenhang zwischen der Existenz und deren Vorstellung nachvollziehen, der wohl nur für „the particular existence of finite Beings without us“ (Locke, 1987, 537) bestimmt zu sein scheint.

Um diesen Gedanken zu verstehen, muss man Lockes Ideenlehre des zweiten Buches¹² heranziehen. Er führt hier die Entstehung komplexer Ideen aus vier einfachen Ideen vor (Locke, 1987, 166f). Diese einfachen Ideen gelangen einmal durch die rein äußere Wahrnehmung eines oder mehrerer Sinne, durch die gemeinsame Aufnahme äußerer Wahrnehmung und innerer Selbstwahrnehmung sowie allein durch die innere Selbstwahrnehmung zur aktiven Tätigkeit des Geistes. Diese Tätigkeit kann der Vergleich, die Abstraktion

¹²Eventuell auch das Schlusskapitel des ersten Buches.

oder das Bilden von neuen Zusammenhängen sein. Daraus entstehen schließlich Ursache-Wirkungs-Zusammenhänge (*Relationen*), Substanzen wie Mensch und Tier und deren Bestandteile (*Modi*).

Ein wichtiger Punkt der Lockeschen Erkenntnistheorie, gerade in Hinsicht auf die vorliegende Arbeit, ist die Lockesche Handhabung der Wahrscheinlichkeit (Locke, 1987, 654-657). Da sich Locke im Klaren gewesen sein muss, dass erstens das Wissen des Menschen immer begrenzt ist, zweitens Sinnestäuschungen die Erfahrung als Quelle zur Wissensverarbeitung angreifbar machen und daher drittens eine absolute Gewissheit nach seiner Theorie nicht möglich sein kann, zieht er die Wahrscheinlichkeit zur Überbrückung des Wissensmangels heran. Damit nimmt die Zuteilung von Wahrscheinlichkeiten zu bestimmten Wissensphänomenen, welche dann nur noch als Glauben oder Meinung bezeichnet werden können und dadurch den zu Grunde liegenden einfachen Ideen, ihren Platz in der empiristischen Epistemologie ein und zeigt, wenn auch unabsichtlich, eine Tendenz zum Skeptizismus.

Mit der Brille des Skeptizismus lässt sich sicherlich auch David Hume lesen (Engfer, 1996, 312-313). Hier soll Hume aber für die empiristische Idee herhalten. Locke hat zumindest in weiten Teilen die Humesche Erkenntnistheorie mitgeprägt. Dies zeigt sich schon in einer ähnlichen Einleitung des *Treatise of Human Nature*, in welcher auch Hume die Prinzipien der menschlichen Erkenntnis als Ziel ankündigt, aber

... in effect propose[s] a complete system of the sciences, built on a foundation almost entirely new, and only one upon which they can stand with any security. (Hume, 1968, 5)

Dieses „fast gänzlich neue Fundament“ ist wieder der empiristische Grundgedanke der Untersuchung von Phänomenen mittels der Erfahrung und Beobachtung und nicht der der hypothetischen Annahmen rationalistischen Gedankenguts. Hume hält auch an Lockes Einteilung von äußeren und inneren Sinneserfahrungen fest. Allerdings nimmt er eine weitere Differenzierung des Ideenbegriffs vor. Hier unterscheidet Hume die *ideas* von den *impressions*, welche sich in ihrer emotionalen und zeitlichen Intensität graduell voneinander unterscheiden, während beide zu den Sinneswahrnehmungen (*perceptions*) gehören (Hume, 1968, 11f). Hume behält auch die Aufteilung der Ideen in einfache und komplexe bei. Allerdings bezieht er diese, in der Logik seiner ersten Unterscheidung, nur auf die Eindrücke (*impressions*). Für die einfachen Eindrücke gilt auch Lockes Vorstellung des Abgleichs von Ideen.

... all our simple ideas in their first appearance, are derived from simple impressions, which are correspondent to them, and which they exactly represent. (Hume, 1968, 13f)

Damit zwingt sich auch Hume zu dem Schluss, dass nur das unmittelbar Wahrzunehmende in der Vorstellung präsent sein kann. Er führt dann aber, sicher ganz im Sinne Lockes, weiter aus (Hume, 1968, 17ff), dass sich durch die Vorstellungskraft (*imagination*) einfache Ideen zu komplexen Ideen verbinden können. Letztere haben schließlich die Eigenschaft, nicht unmittelbar beobachtbar zu sein und können auch durch den inneren Wahrnehmungssinn entstehen.

Dieser Prozess der Neuorganisation einfacher Ideen zu Konglomeraten höherer Ordnung – und das ist nun neu – geschieht durch bestimmte Prinzipien der Assoziation (Hu-

2 Nativismus und Empirismus als grundlegende Positionen

me, 1968, 19ff). Möglicherweise hat Locke diese Prinzipien einfach als selbstverständlich vorausgesetzt, Hume formuliert sie nun explizit. Er identifiziert „three, viz. resemblance, contiguity in time or place, and cause and effect“ (Hume, 1968, 19).

Nun übernimmt Hume aber die Leibnizsche Vorstellung von Vernunftwahrheiten und Tatsachenwahrheiten und ordnet sie den Begriffen *Relation of Ideas* und *Matter of Fact* zu (Hume, 2007, 25). Die Vernunftwahrheiten liegen wie bei Leibniz im Bereich der Logik und Mathematik, und man kann sich deren gewiss sein, weil sich das Gegenteil eines Beweises logisch ausschließen lässt. Bei Tatsachenwahrheiten ist das nicht möglich. Sie beruhen auf Erfahrung und über diese kann es keine absolute Gewissheit geben.¹³ Damit wird auch hier die Wahrscheinlichkeit eines Ereignisses offenbar (Hume, 1968, 76ff). Für die Tatsachen ist nun eines der Prinzipien der Assoziation wichtig: Ursache und Wirkung.

Mehrfache Beobachtung einer Erfahrung verleitet den Menschen, eine Kausalität zu unterstellen. Hume schildert in den *Enquiries* das Billardspiel zur Veranschaulichung (Hume, 2007, 28f). Dadurch, dass der Mensch immer die gleiche Beobachtung bei dem Kräftewirkungsverhalten der einen Kugel auf die andere beobachtet, schließt er aus der Erinnerung auf das zukünftige Verhalten. Beide Vorstellungen, Vergangenheit und Zukunft, müssen gleich sein, damit eine Kausalität zwischen Ursache und Wirkung der Kugeln entdeckt werden kann. Im Grunde erlaubt uns diese Assoziation aber nur eine Aussage über die Gewohnheit zu treffen, nicht jedoch über eine von der Erfahrung unabhängige Gesetzmäßigkeit.

2.3 Symbolistische und konnektionistische Paradigmen

Ähnlich wie Engfer (1996) für die Philosophiegeschichte überzeugend dargestellt hat, dass die Aufteilung in Empirismus und Rationalismus ein philosophiehistorisches Schema ist, in welches sich nicht einmal die Urväter beider Theorien einordnen lassen, ist auch das Pendant in der KI, das bis in die Linguistik hineinreicht, Symbolverarbeitung und Konnektionismus beziehungsweise Nativismus und Empirismus, lediglich ein Denkschema. Man ist sicherlich verleitet, die Wurzeln der Symbolverarbeitung im Rationalismus zu suchen und auf gleiche Weise den Empirismus als gedanklichen Vorreiter des Konnektionismus anzunehmen. Zahlreiche Parallelen in den Grundannahmen weisen darauf hin.¹⁴ Allerdings muss man dann auch konsequenterweise zugestehen, dass beide Ansätze ineinander übergreifen und nur in gedanklichen Extremen an den Enden eines Kontinuums angesiedelt werden können, so wie es auch in der Rationalismus-Empirismus-Debatte der Fall ist.¹⁵ Kurzweil stellt einen Zusammenhang zwischen beiden Positionen her.

Die Fähigkeit zu lernen, sich Wissen anzueignen und dieses deduktiv und folgerichtig zu handhaben, wird häufig als symbolisches Denken bezeichnet, als die Fähigkeit, Symbole

¹³Möglicherweise wird Hume aus diesem Grund auch als Skeptiker bezeichnet.

¹⁴Wie beispielsweise von Pinker (1999, 40), Marcus (1998) oder Newell (1990, 76) beschrieben.

¹⁵vgl. auch Cowie (1999, 44).

2.3 Symbolistische und konnektionistische Paradigmen

zu manipulieren. Ein Symbol ist eine Benennung oder ein Zeichen, das stellvertretend für einen anderen Sachverhalt, im allgemeinen für eine Struktur oder für ein Netzwerk von Fakten, steht. Symbole sind für gewöhnlich in komplizierten Mustern angeordnet . . . (Kurzweil, 1993, 17)

Mithin macht er deutlich, dass beide Ansätze miteinander vereinbar sind, wenn man unter dem symbolischen Charakter des Denkens auch verteilte Informationen eines Musters oder Netzwerkes fasst. Auch Plunkett räumt für die linguistisch orientierten Ansätze ein, dass

Even the staunchest of modern-day empiricists must admit that guarantees of a given cognitive architecture cannot be fully determined by the environment. (Plunkett, 1998, 98)

Andererseits steht sie der nativistischen Auffassung kritisch gegenüber.

. . . it quickly became apparent that some language acquisition researchers had grossly underestimated the structural regularities that could be extracted from the input. (Plunkett, 1998, 98)

Auf gleiche Weise trifft Marcus (2001, xii) eine klare Aussage zu seinem Standpunkt.

I am not an anti-connectionist, I am opposed only to a particular subset of the possible connectionist models.

Dabei ist er deutlich vom Nativismus inspiriert:

Without some innately given learning device, there could be no learning at all. (Marcus, 2001, 143)

Dieser Ansatz ist ebenso in den Neurowissenschaften präsent.

Die Anatomie des neuronalen Schaltkreises ist ein simples Beispiel für die apriorische Erkenntnis Kants, während die Veränderungen der Stärke bestimmter Verbindungen im neuronalen Schaltkreis den Einfluss der Erfahrung widerspiegeln. (Kandel, 2006, 225)

Scheerer (1990, 23) stellt fest:

Keine der von uns diskutierten theoretischen Dimensionen hat eine wirklich trennscharfe, absolut unüberbrückbare Diskrepanz zwischen den beiden Paradigmen ergeben.

Den Unterschied sieht Scheerer darin, wie man das Explanandum letztendlich betrachtet.

Symboltheoretiker . . . unterstellen, daß das Denken produktiv und systematisch ist; und sie halten es für notwendig, diese Eigenschaft mit Hilfe syntaktischer Konstituentenstruktur, semantischer Kompositionalität und kontextfreien semantischen Konstituenten zu erklären. Konnektionisten unterstellen, daß das Denken wesentlich durch Plastizität, spontane Generalisation und Bedeutungsholismus gekennzeichnet ist; und sie halten es für notwendig, solche Eigenschaften durch assoziatives Lernen, simultane Erfüllung von Randbedingungen und Einschwingen auf Zustände minimaler Strukturenergie zu erklären. (Scheerer, 1990, 23)

Für die Entwicklungspsychologie bestätigt Karmiloff-Smith (1995) das hier gegebene Gesamtbild.

I do not choose between these two epistemological stands [nativism and constructivism], one arguing for predominantly built-in-knowledge and the other for a minimum innate underpinning to subsequent domain-general learning. (Karmiloff-Smith, 1995, xiii)

2 Nativismus und Empirismus als grundlegende Positionen

Stattdessen bestätigt sie, dass die endgültige Theorie des Geistes Aspekte aus beiden Paradigmen umfassen wird und dass sich beide Paradigmen in wichtigen Punkten ergänzen (Karmiloff-Smith, 1995, xiii). Diese kurzen Beispiele illustrieren bereits, dass die Nativismus-Empirismus- und Symbolismus-Konnektionismus-Dualitäten eher gemeinsam und ergänzend betrachtet werden sollten.

Die nativistische Strömung in der Linguistik ist dennoch unbestreitbar eng mit der Idee der Symbolverarbeitung der KI-Forschung verbunden (vgl. Marcus, 1998). Bei Pinker wird diese Denkweise besonders deutlich, wenn er schreibt:

The theory of symbol processing seems better suited to explaining the brain ability to handle complex ideas and the aspects of language that communicate them. People are not slaves to similarity. (Pinker, 1999, 40)

oder

It is a cliché of neuroscience that the brain works differently from a digital computer . . . one of the mechanisms that makes computers intelligent – manipulating symbols according to rules – may be a basic mechanism of the human brain as well. (Pinker, 1999, 40)

Da

advocates of symbol manipulation assume that our minds have both a set of internally represented primitive elements and a way of internally representing structured combinations of those elements (vgl. Marcus, 2001, 85),

wäre ein gemeinsamer Nenner von Linguistik und der Symbolverarbeitung der künstlichen Intelligenz formal schon hergestellt. Im symbolischen Ansatz der KI-Forschung dienen der Repräsentation und der Verarbeitung von Wissen Symbole, die auf verschiedenste Weise manipuliert werden können. (Newell (1990, 88ff); Bechtel und Abrahamsen (1991, 8f)). So werden kognitive Phänomene als „Symbolmanipulationsaufgaben“ verstanden (Görz und Nebel, 2003, 39). Dabei spielt die formale Logik zur Beschreibung der Repräsentationen eine bedeutende Rolle wie die Arbeiten von McCarthy in den 50er und 60er Jahren belegen. Im Prinzip bringt hier die Mathematik als hintergründige Wissenschaft ihre bereits weit entwickelten Verfahren an. In der Weiterentwicklung dieses Ansatzes sind Multi-Agenten-Systeme entstanden, bei denen die zentrale Problemstellung in viele kleinere Teilaufgaben herunter gebrochen wird und durch das Zusammenspiel der einzelnen Komponenten Synergien erzielt werden. Der Hauptkritikpunkt des symbolischen Ansatzes ist das Komplexitätsproblem, das bei der Repräsentation des Weltwissens auftritt. Das komplexe Ineinanderwirken von Umweltkomponenten ist nur äußerst begrenzt mathematisch modellierbar und ist deshalb nur in abgegrenzten Teilbereichen sinnvoll.

Da nun Sprache ihrer Natur nach symbolisch und zudem unstrittig eine kognitive Fähigkeit ist, lag hier eine Ausweitung beziehungsweise gegenseitige Vermengung der vorhandenen Konzepte aus der KI-Forschung und Linguistik auf der Hand (vgl. Pylyshyn, 1984, 49ff). Die Grundlagen des nativistischen Ansatzes passen nun insbesondere zum Paradigma der Symbolhypothese, dass hier fast eine begriffliche Verschmelzung in der Literatur zu beobachten ist (z.B. vgl. Haase (2005, 126); Hövel (2002, 51)). Schon allein der Umstand, dass Chomsky, Miller, Newell und Simon zusammen auf dem MIT-Symposium

2.3 Symbolistische und konnektionistische Paradigmen

von 1956 ihre Ergebnisse vorstellten und die kognitive Revolution einleiteten, lässt eine gewisse gemeinsame Grundhaltung vermuten. Chomsky macht zwar unmissverständlich deutlich, dass ein Modell der Sprachproduktion, wie es in Shannons informationstheoretischem Ansatz (Shannon und Weaver, 1971) dargestellt wurde, nicht auf natürliche Sprachen angewendet werden könne (Gardner (vgl. 1989, 40);¹⁶ Bechtel und Abrahamsen (1991, 13)) aber in der Tat basiert Chomskys Transformationsgrammatik auf Regeln, die sich symboltheoretisch gut simulieren lassen (vgl. Pereira, 2000). Andererseits bereitet natürlich die hohe Abstraktionsebene der Chomskyschen Theorien Probleme, und diese lassen sich überhaupt nicht algorithmisch darstellen, sondern müssen immer spezifiziert werden.¹⁷ So bemerkt Gardner diesbezüglich:

Obwohl Chomskys Ansatz mit der künstlichen Intelligenz viele Wurzeln teilt, sind einige seiner grundlegenden Ideen nicht ohne weiteres auf Computerformat übertragbar (Gardner, 1989, 231).

Der *Logic Theorist* von Simon und Newell (1958) beruht dagegen schon im Entwurf auf Symbolmanipulationen und greift auf einen explizit formulierten mathematischen Regelapparat zurück. Nur wenige Monate vor dem MIT-Symposium hatten sich Minsky, McCarthy und wiederum Simon und Newell in Dartmouth getroffen, um sich über ihre jüngsten Ideen zu Problemlösungsstrategien bei Maschinen auszutauschen. Diese Ideen werden heute unter dem Begriff der Künstlichen Intelligenz (KI) subsumiert und finden ihren Niederschlag im symbolorientierten Ansatz der starken KI.

Chomsky (1957) hat die Idee des symbolorientierten Ansatzes schließlich in die Sprachwissenschaft eingebracht, um der behavioristischen Tendenz des amerikanischen Strukturalismus eine andere Richtung zu geben. Der wohl damals prominenteste Vertreter des Behaviorismus, Burrhus F. Skinner, vertrat in seinem Buch *Verbal Behavior* im Einklang mit der behavioristischen Tradition die Auffassung, dass Sprache ein Verhalten sei (Skinner, 1957). Chomskys Kritik (Chomsky, 1959) an dieser Sichtweise ebnete den Kognitionswissenschaften den Weg, sich zur vorherrschenden Wissenschaft des menschlichen Geistes zu entwickeln. Damit gewann der symbolische Ansatz in der Linguistik großen Einfluss. Zeitgleich ist dieser Gedanke auch in die sich parallel entwickelnde Disziplin der künstlichen-Intelligenz-Forschung als grundlegendes Paradigma eingeflossen.

Der Kern des symbolischen Ansatzes ist, was Smolensky und Legendre (2006a, 18) die „kombinatorische Strategie“ (*combinatorial strategy*) nennen und Chomsky (1957) für die natürlichen Sprachen ausführte. Auf abstrakter Ebene versteht man darunter, dass durch eine Produktionsregel aus einer endlichen Menge eine unendliche Menge entstehen kann. Dieses einfache mathematische Prinzip, das sich seit Jahrhunderten in der Zahlentheorie der Folgen und Reihen wieder findet, wird im symbolischen Ansatz auf das System der Sprache übertragen und hier mittels neuer kombinatorischer Möglichkeiten eingeführt. Hier werden die Phoneme oder das Lexikon einer Sprache als endliche

¹⁶Gardner bezieht sich hier offensichtlich auf Shannons Diplomarbeit aus dem Jahr 1938 „A Symbolic Analysis of Relay and Switching Circuits“, in: Transactions of the American Institute of Electrical Engineering, 57, 1-11 erschienen. Der informationstheoretische Ansatz findet sich aber in Shannon und Weaver (1971) (beziehungsweise 1949) in zugänglicher Form.

¹⁷vgl. Lappin und Shieber (2007) als ein praktisches Beispiel.

2 Nativismus und Empirismus als grundlegende Positionen

Menge definiert. Den Produktionsregeln der Sprache liegt ebenso ein endliches Wissen (*finite knowledge*) zu Grunde. Mit diesen Regeln werden aus den Wörtern (beziehungsweise Konstituenten) einer Sprache unendlich viele Sätze hergeleitet, welche wiederum unterschiedliche Komplexitätsstufen besitzen können und über rekursive Mechanismen verfügen.¹⁸ Dies bedeutet, dass Fakten durch unterschiedlich strukturierte Kombinationen gespeichert werden. Derartige Kombinationen lassen sich nun aber auch statistisch auffassen, gleichwohl man dadurch in weiten Teilen des Nativismus leider auf eine unvereinbare Haltung trifft, der es jedoch in Hinblick auf die Prämissen des Nativismus an einer stichhaltigen Begründung mangelt.¹⁹

Statistische Ansätze werden im Nativismus grundsätzlich abgelehnt.²⁰ Chomsky äußert sich bereits im Anfangsstadium der *Innateness*-Debatte zu quantitativen Ansätzen. Im Grunde genommen sieht er keinen Zusammenhang zwischen Sprachstatistik und grammatikalischem Wissen. Er schreibt:

I think that we are forced to conclude that . . . probabilistic models give no particular insight into some of the basic problems of syntactic structure. (Chomsky, 1957, 17)

In diesem Zusammenhang formulierte er seinen berühmten Satz der grünen Ideen. Der Standpunkt, Sprachwissen unabhängig vom quantitativen Sprachinput zu untersuchen, wird zumindest in einigen Teilen der nativistischen Positionen weiter propagiert. Beispielsweise bemerkt Fodor über die Theorie des Geistes erst kürzlich:

This theory can't of course, be associationistic, since associative relations among mental states are supposed to hold not in logical form, but rather in virtue of statistical facts about (e.g.) how often they have occurred together, or how often their occurring together has led to reinforcement, etc. (Fodor, 2001, 18)

Aus der ablehnenden Haltung gegenüber quantitativer Mustererkennung in der Sprache in nativistischen Theorien sowie in weiten Teilen der Symbolverarbeitung erklärt sich zumindest auch bruchstückhaft die Einstellung der Nativisten gegenüber dem Konnektionismus.

Die Geburtsstunde des Konnektionismus kann zwischen dem McCulloch-Pitts-Netzwerk (McCulloch und Pitts, 1943) und dem Perceptron (Rosenblatt, 1958) angesetzt werden²¹ und damit etwa zeitgleich mit den symbolverarbeitenden Ansätzen der KI. In einer vernichtenden Kritik zeigten Minsky und Papert (1969), dass Netzwerke mit direkten Bezie-

¹⁸Marcus (2001, 85):

complex units can themselves serve as input for building still more complex units – combinations that are recursively defined.

¹⁹vgl. auch Abney (1996, 22).

²⁰Eine Ausnahme ist sicherlich Niyogis Vorschlag eines *Triggering Learning Algorithm* (Niyogi, 2006, 103). Er beschreibt mögliche Ausprägungen der *Principles-and-Parameter*-Theorie mit Markoffketten als einen probabilistischen Verarbeitungsmechanismus.

²¹Sicherlich ließe sich auch für das McCulloch-Pitts-Netzwerk (McCulloch und Pitts, 1943) als alleiniger Ausgangspunkt argumentieren. Der Grund, warum dies hier nicht geschieht, ist, dass es eine doch recht wesentliche Eigenschaft auslöst, welche man heute allgemein mit neuronalen Netzwerken verbindet: das Hebbsche Prinzip (Hebb (2002) beziehungsweise als erste Auflage 1949). Diese Regel erlaubt die synaptische Plastizität zu simulieren.

2.3 Symbolistische und konnektionistische Paradigmen

hungen zwischen Input- und Output-Ebenen nicht alle logischen Strukturen berechnen können und deshalb für eine Theorie des Geistes unbrauchbar seien. Der Kern dieser Kritik, eine Spekulation, sollte sich erst später als falsch herausstellen. Bis dahin jedoch blieben die konnektionistischen Modelle unbeachtet. Erst die Hopfield- (Hopfield, 1982) und Kohonennetzwerke (Kohonen, 1982) verschafften dem Konnektionismus wieder Auftrieb. Wenig später sind dann auch Boltzmannmaschinen und natürlich die bekannten PDP-Modelle von Rumelhart und McClelland (1986a) und Rumelhart und McClelland (1986b) aktuell geworden.

Womöglich hat das in der Vorgängerzeit des Konnektionismus der 1960er und 70er Jahre vorherrschende universalgrammatische Paradigma Chomskys seinen Einfluss auf deren erbitterte Gegner ausgeübt. Diese fanden nun auf der Suche nach einer besseren Erklärung recht plötzlich eine fundierte Basis mit den Netzwerkmodellen. Damit lässt sich der große Zulauf ad hoc und die anschließende Polarisierung erklären. Allerdings ist wohl dadurch umso stärker der Eindruck entstanden, dass die Gegner des Nativismus und des Konnektionismus zusammengehören, sowie auch Nativismus und Symbolismus nun symbiotisch erscheinen. Hier lässt sich schon recht deutlich eine Parallele zum philosophiegeschichtlichen Dogma des Rationalismus und Empirismus entdecken.

Der Konnektionismus überträgt die Vorstellung natürlicher Nervensysteme auf künstliche Rechenmaschinen.²² Er unterstellt dabei verteilte Informationen in sogenannten *Units*. Die mittlerweile als raum-zeitliche Vektoren kodierten Kombinationen von solchen Einheiten erlauben in einer Verschaltung auf mehreren Ebenen höchst komplizierte Musterabgleiche. Die Aktivierung der *Units* ist meistens inputabhängig.²³ Durch das gemeinsame Feuern der *Units* entstehen Muster, welche in Abhängigkeit von Lernregeln oder/und Schwellenwerten erkannt werden. Verschiedenste Architekturen beeinflussen dabei die Erkennungsleistung des Musters.

Der wichtigste Vorteil des Konnektionismus lässt sich am besten mit dem Schlagwort der Plastizität erfassen. Dadurch, dass sich hoch komplexe Muster nicht nur in einer Eigenschaft voneinander unterscheiden, ist es möglich, bei unvollständigem Input noch den richtigen Output zu erzielen. Dies trifft genau dann zu, wenn die Einheiten aktiviert werden, die nur in dem Zielmuster vorkommen. Die restliche Information ist dann nicht mehr zwingend erforderlich, sodass hier eine Art logischer Schluss vorliegt. Aus diesem Grund können Lernprozesse besonders gut mit subsymbolischen Architekturen simuliert werden. Schließlich basiert das Argument der Reizarmut im Spracherwerb eben darauf (Empirische Prämisse unter c. in Kap. 1.3), dass Sprache trotz unvollständigem Input recht einheitlich erlernt wird. Die natürliche Sprache kristallisierte sich schon allein deshalb als der perfekte Schauplatz eines Schlagabtauschs zwischen Symbolismus und Konnektionismus heraus: Sprache ist unstrittig eine kognitive Fähigkeit des Geistes und verbirgt komplexeste Muster, deren Ordnung entweder mit verteilten Informationen in

²²Das Gegenteil ist bei der Symbolmanipulation geschehen. Hier sind Turing-Maschinen und von Neumann-Architekturen die geistigen Väter, welche erst später in einer Hardware-Software-Analogie versucht wurden, dem Körper und Geist des Menschen aufzupressen.

²³Eine solche Architektur muss aber nicht zwingend erforderlich sein. Durch *backpropagation* können die *Units* auch *Top-down* erregt werden.

2 Nativismus und Empirismus als grundlegende Positionen

Netzwerken oder durch Symbolmanipulation hergestellt wird. Kann man nun mit neuronalen Architekturen genaue Vorhersagen über den Spracherwerb oder die Sprache im Allgemeinen treffen, und gelingt dies nicht mit Symbolmanipulationen, so müsste dies für eine Reihe von Wissenschaftlern (vgl. Abb. 1.2) auch die Unbrauchbarkeit der nativistischen Idee bedeuten. Gleiches gilt natürlich für den umgekehrten Fall. Die Lösung des Rätsels der Sprache, eine stichhaltige Erklärung ihrer Verarbeitung, ihres Erwerbs und ihrer Natur, wäre somit der Schlüssel zur Bestätigung eines der Paradigmen.

Zusammenfassend lässt sich festhalten, dass die symbolverarbeitenden Modelle dem Nativismus schon sehr viel näher entgegen kommen (vgl. Pospeschill (2004, 146); Fetzer (2001, 131)). In ihrer Kritik an den konnektionistischen Modellen bemerken Pinker und Prince nicht ohne Grund:

From its inception, the study of language within the framework of generative grammar has been a prototypical example of how fundamental properties of a cognitive domain can be explained within the symbolic paradigm. (Pinker und Prince, 1988, 78)

Das Lernen spielt bei symbolverarbeitenden Ansätzen im Vergleich zu den konnektionistischen Modellen eine eher untergeordnete Rolle und erfolgt in der Symbolverarbeitung nach explizit vorgegebenen Regeln. Im Kern der Symbolverarbeitung und des Nativismus gleichermaßen steht die Manipulation von Symbolen. Die Repräsentationen von sprachlichen Einheiten werden quasi als Einzelsymbole angenommen, die im Geiste als solche weiterverarbeitet werden. Dies gilt auch für klassische AI-Programme. Sie arbeiten wie eine Turing-Maschine regelbasiert, oft auch unter Verwendung von rekursiven Funktionen (Newell, 1990, 68). Beim Konnektionismus liegt die Analogie zur Turing-Maschine nicht auf der Hand, weil eine verteilte Verarbeitung von Informationen nicht sequentiell wie bei der klassischen Turing-Maschine funktioniert. Dennoch lassen sich mit Netzwerken alle logischen Funktionen eindeutig lösen und damit ist das Ergebnis zwischen neuronaler und symbolmanipulierender Berechnung zumindest im Ergebnis gleich. Da der Konnektionismus oft mit Assoziationismus gleichgesetzt wird, folgt meist auch eine Zuordnung zum empiristischen Standpunkt (Fodor und Pylyshyn (1988); Fodor (2001, 18)). Dabei basiert der Konnektionismus nur auf einigen Annahmen der Assoziationstheorie; andere Annahmen, wie verteilte Repräsentationen und *hidden layers*, sind ihr fremd (Medler, 1998), was auch schon damals bekannt gewesen sein musste (Bechtel, 1985, 60).

2.4 Positionierung der Arbeit in einem weiten Wissenschaftskontext

Am Anfang des Kapitels 2.1 wurde ein gedankliches Kontinuum gezeichnet, indem die Arbeit nun platziert werden kann. Die Reflexion über die philosophischen Grundlagen und die kognitionswissenschaftlichen Paradigmen der angewandten Linguistik verleitet zu einer eher neutralen Positionierung. Im Hinblick auf die Rationalismus-Empirismus-Dimension der Abbildung 2.1 tendiert die Arbeit aber schon deutlicher zum empiristischen Standpunkt, weil hier dem Sprachinput eine größere Rolle zugeordnet wird. Diese

Einordnung soll jedoch keine Extremposition implizieren.²⁴ Die Tendenz wird nun auch durch die zweite Dimension leicht zur Mitte abgelenkt. Im Spannungsfeld der symbolischen und konnektionistischen Denkweisen kann die Arbeit in der Tat nicht polarisieren, weil der Algorithmus (Kap. 5) in symbolischen als auch in subsymbolischen Architekturen gleichermaßen funktionieren würde. Prinzipiell wird in der später vorzustellenden Simulation ein Algorithmus verwendet, der, durch Häufigkeiten bedingt, Strukturen in der Sprache herstellt. Das Resultat einer solchen Simulation wird auch beim *Unsupervised Learning* in neuronalen Netzen erzielt. Die Prozesse sind jedoch beim neuronalen Netz meist auf mehreren Ebenen und parallel organisiert. Außerdem existiert dort eine Funktion als Lernregel, mit der das Feedback erfolgt und die interessierenden Strukturen entdeckt werden können. Der hier entwickelte Algorithmus hat a priori keine Lernregel. Sie wird erst im Nachhinein durch Vergleich und Analyse des Outputs als Muster abgeleitet. Mithin darf man behaupten, dass der hier vorliegende Algorithmus eine vereinfachte Simulation eines neuronalen Netzes ist, somit auch die gleichen mathematisch kodierten Informationen nutzt, nicht jedoch in parallel organisierter Verarbeitung. In der Literatur operiert man mit dem Begriff des genetischen Algorithmus (vgl. Pospeschill, 2004, 86). Der hier verwendete Algorithmus ist insofern eine Sonderform davon, weil er im Grunde ein endlicher Automat in Form einer sichtbaren Markoffkette ist, welcher interne Strukturen einer Sprache verarbeiten kann, ohne dass der Apparat spezielle Anweisungen über diese Sprache erhält. In diesem speziellen Fall lernt er über die Sprachverarbeitung selbst. Der Apparat verdichtet selbständig den Informationsinput und gibt Gesetzmäßigkeiten in Form von Mustern aus. Mit zunehmendem Input erhöht sich die Sicherheit der Richtigkeit der abgeleiteten Gesetzmäßigkeiten.

Mit diesen Informationen sollen sich auch zwei Netzwerkarchitekturen miteinander verbinden lassen. Wie im fünften Kapitel verständlich werden wird, sollte der Teilbereich des genetischen Algorithmus, welcher das schrittweise Zusammenstellen und Einfügen des Paralexikons spezifiziert, mit einem kompetitiven Netzwerk erreicht werden. Dieses Netzwerk selbst würde nun der Teil eines rekurrenten Netzwerkes sein, der normalerweise durch Kontextknoten eingenommen wird. Das rekurrente Netzwerk ändert dadurch seinen Charakter der *Backpropagation*, da die neue Information durch den Output des kompetitiven Netzes gesteuert wird, es lernt also unüberwacht. Die optimalen Werte der Übergangswahrscheinlichkeiten werden dann durch die normale Architektur mit einer versteckten Schicht erlernt, wie es in anderen *Feed-Forward*-Netzen der Fall ist.

Mit dem Ergebnis dieser Arbeit könnte also recht problemlos ein neuronales Netz programmiert werden.²⁵ Neuronale Netze mit einer oder nur wenigen Ebenen, wie das bekannte *single layer connectionist network* von Rumelhart und McClelland (1986b) zum

²⁴Sicherlich ganz im Sinne Riddleys, der da schreibt:

Let me at once play my cards face up. I believe human behaviour has to be explained by both nature and nurture. I am not backing one side over the other. But that does not mean I am taking a 'middle of the road' compromise. (Ridley, 2003, 3)

²⁵Für eine ausführlichere Darstellung dessen sei auf die Schlussbetrachtung verwiesen.

2 Nativismus und Empirismus als grundlegende Positionen

Erlernen des englischen regelmäßigen und unregelmäßigen Präteritums, lassen sich sehr einfach auf die distributionalen Informationen ihres Inputs zurückführen.²⁶ Schließlich sind es genau die quantitativ verteilungsbedingten Vorkommen bestimmter Lautstrukturen, welche im Netzwerk als Gewichtungen zwischen ihren Verbindungen Verwendung finden oder in Kombination mit einer Lernregel abgestimmt werden können. Mittels der Multiplen Linearen Regression würde sich Rumelharts und McClellands Netzwerk als einfache Häufigkeitsverteilung darstellen lassen. Auch komplizierte konnektionistische Lernalgorithmen legen generelle statistische Gesetzmäßigkeiten, wie den Satz von Bayes oder MDL,²⁷ zu Grunde, basieren also auf einer statistischen Verteilung. Diese Prinzipien finden ebenso in den Algorithmen dieser Arbeit Anwendung. Allerdings bringen nichtlineare dynamische Netzwerke Strukturen eines Textes zu Tage, die man aus der äußeren Form zunächst nicht erraten kann (vgl. z.B. Elman, 1999, 97ff). Die genaue Abgrenzung zur nativistischen Idee und dem durch den *poverty of the stimulus* begründeten *language device* ist hier damit wiederum die gleiche wie im konnektionistischen Paradigma. Sie besteht in dem Bestreben, das Argument, dass der Sprachinput als *Primary Linguistic Data* (PLD) einen Großteil der Informationen beinhaltet, die das Kind zur Bildung der Grammatik beziehungsweise zur fehlerfreien Wortsegmentierung benötigt, zu erhärten.

²⁶ vgl. hierzu auch Redington und Chater (1997, 277).

²⁷ *Minimum Description Length*

3 Theorien der Wortsegmentierung

Segmentationsstrategien der lexikalischen Wortsegmentierung werden von der vorlexikalischen Wortsegmentierung (wie in Kap. 4) in der Regel recht streng auseinander gehalten. Dazu merkt Gaskell (2007) an

When it comes to learning models, the lexical/pre-lexical distinction becomes somewhat murky. (Gaskell, 2007, 61)

Wie der Autor dann fortführt, hängt dies damit zusammen, dass sich durch die Wortgrenzenerkennung natürlich auch das gesamte Wort als Einheit definieren lässt. Diese Vermischung spiegelt sich zwangsweise auch in der Beschreibung der Worterkennungsmodelle im Kapitel 3.1 wider. Das Hauptaugenmerk liegt jedoch darin, ein passendes Modell für den lexikalischen Zugriff in der Simulation zu finden, und unter Rücksichtnahme auf die Entscheidung der linguistischen Perzeptionseinheit (Kap. 4.1) soll kritisch geprüft werden, ob das Phonem auch für die Modellierung lexikalischer Modelle geeignet ist. Aus den im Kapitel 4.1 angestellten Überlegungen wäre zu vermuten, dass das Phonem auch zur lexikalischen Erschließung eine zentrale Rolle einnehmen müsste. Allerdings sind die Modelle des lexikalischen Abgleichs hinsichtlich des Alters des Kindes un spezifiziert. Vorbild sind kompetente, erwachsene Sprecher. Spezielle Modelle für Kleinkinder, die im Begriff sind, eine detaillierte lexikalische Struktur aufzubauen, stehen noch nicht zur Verfügung.

Kapitel 3.2 behandelt die möglichen Segmentierungsstrategien für Wortgrenzen bei Kleinkindern. In der Erstspracherwerbsforschung steht eine Reihe von Ansätzen bereit, die Segmentierung von Worteinheiten aus dem Lautstrom gesprochener Sprache heraus zu erklären. In der Literatur werden vier Segmentierungsstrategien übereinstimmend angeführt.¹ Danach werden Wortgrenzen durch prosodische Markierung (*prosodic markers*), phonotaktische Randbedingungen (*phonotactic constraints*), allophonische Variationen (*context sensitive allophones*) oder statistische Verteilungen (*statistical regularities*) im Sprachsignal erkannt. Im Gegensatz zu Kapitel 3.1 handelt es sich hierbei um nicht-lexikalische oder prälexikalische Prozesse. Das Kind verfügt noch über kein Lexikon und sucht aus diesem Grund nach anderen Anhaltspunkten im Lautstrom. Im Kapitel 3.3 werden schließlich Überlegungen angestellt, wie die unterschiedlichen Positionen in Einklang gebracht werden können.

¹Beispielsweise bei Mattys et al. (1999, 467); Gambell und Yang (2005); Weijer (1999, 15ff); Jusczyk et al. (1999c, 161-163); Batchelder (2002, 168); Johnson et al. (2003, 87); Aslin et al. (1998, 322); Bortfeld et al. (2005, 298); Perruchet und Vinter (1998, 246).

3.1 Modelle der Wortsegmentierung und lexikalische Erschließung

Erstens werden zwei Arten von Worterkennungsmodellen vorgestellt: Modelle, die eine prälexikalische Repräsentation unterstellen und Modelle, die eine direkte Abbildung des Signals zum mentalen Lexikon annehmen. Dabei werden in erster Linie die Modelle vorgestellt, welche hinsichtlich der Repräsentationseinheiten nutzbringend erscheinen. Eine direkte Abbildung erscheint aus psycholinguistischer Sicht nicht gerade plausibel, weil das Eingangssignal höchst unterschiedliche Variationen annehmen kann. Bei direkter Abbildung stellt sich das Problem, dass man ohne Abstraktionsmechanismus alle leicht divergenten Signalausschnitte im Langzeitgedächtnis speichern müsste. Diese dürften dann so zahlreich sein, dass ein effizienter Abgleich schwer vorstellbar wird.

Die Diskussion über die Wahrnehmung und Verarbeitung von linguistischen Einheiten ist eine Teilmenge der Forschung über die Worterkennung gesprochener Sprache. Deshalb ist die weitläufige Literatur zu diesem Themengebiet auch die Grundlage für das Verstehen der Repräsentation kleinster linguistischer Einheiten, die den Zugriff auf das mentale Lexikon ermöglichen. Für eine erste Orientierung sei auf Abbildung 3.1 verwiesen. Wie Frauenfelder und Floccia (1999, 1) darlegen, lässt sich die Diskussion auf drei Problembereiche zurückführen: Repräsentation, Identifikation und Integration. Unter Repräsentation wird dabei die innere Struktur der grundlegenden lexikalischen Einheit und die Organisation dieser Einheiten im mentalen Lexikon verstanden. Identifikation umschreibt, wie ein Eintrag gefunden und verarbeitet wird. Letztlich spricht man von Integration, wenn auf das Zusammensetzen der einzelnen Repräsentationen zu einer bedeutungstragenden Interpretation auch auf syntaktisch höheren Ebenen der Phrasen oder Sätze abgezielt wird.

In dem, was folgt, wird die Rolle der Repräsentation in psycholinguistischer Forschung eingehender beleuchtet. Als Übersicht für den Fortgang dient Abbildung 3.1. Darin ist im oberen Drittel ein einfaches Modell angedeutet, das aus drei Komponenten besteht: Sprachsignal, vorlexikalische Repräsentation und mentales Lexikon. Nach Kenntnis des Autors stimmen alle hier anwendbaren, psycholinguistischen Theorieansätze darin überein, dass ein mentales Lexikon existiert. An dieser Stelle hören die Gemeinsamkeiten allerdings schon auf.² Selbst eine genauere Spezifizierung des mentalen Lexikons, dessen Organisationsstruktur, Zugangsmöglichkeiten, interne Verarbeitungsprozesse und im Prinzip alles, was überhaupt im Lexikon enthalten ist und vor sich geht, kann nicht mehr im einheitlichen Kanon wiedergegeben werden.³ Auf abstrakter Ebene lässt sich aber unstrittig formulieren, dass das mentale Lexikon eine Relation zwischen einem Bedeutungsgehalt und einer dazu bestimmten Form, sei sie phonologisch-abstrakt oder signalspezifisch kodiert, herstellt. Da in dieser Arbeit das mentale Lexikon nicht Gegenstand des Erkenntnisziels ist, braucht eine Diskussion über dessen Spezifikationen nicht vertieft zu werden.

²vgl. Lively et al. (1994, 277ff)

³vgl. Moss et al. (2007, 217f); Clahsen und Rothweiler (1992) mit anschließender Diskussion (991-1060)

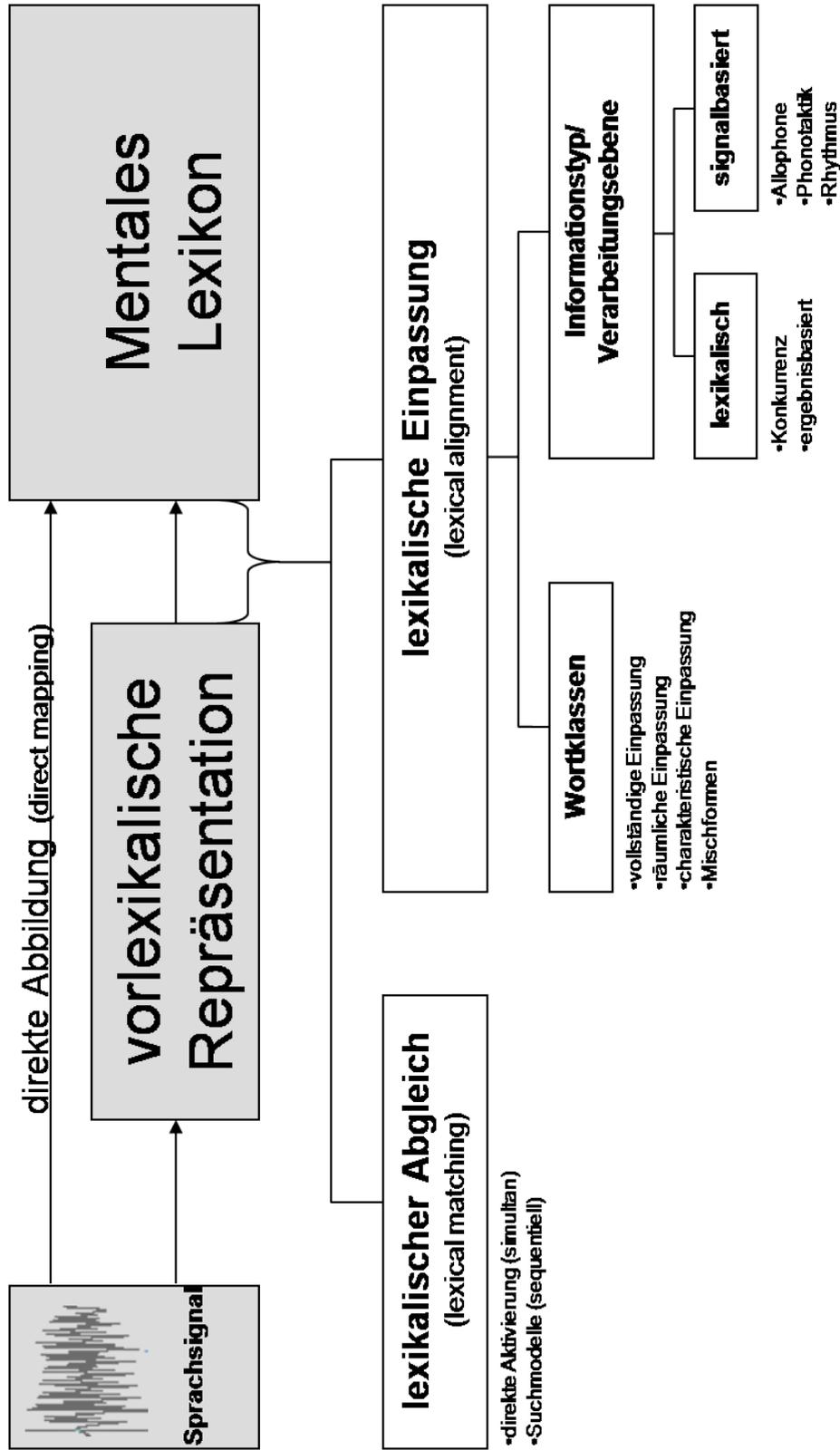


Abbildung 3.1: Verarbeitungsebenen für die Erkennung gesprochener Wörter und deren Zuordnung zu Theorien, Quelle: Eigene Darstellung (inhaltlich teilweise entnommen aus Frauenfelder und Floccia (1999))

3 Theorien der Wortsegmentierung

Als Informationsinput dient das Sprachsignal,⁴ das möglicherweise in eine vorlexikalische Repräsentation überführt oder direkt auf eine mentale Lexikonstruktur abgebildet werden muss. Dieser Prozess wird in der Tat kontrovers diskutiert. Zuerst sollte geklärt werden, warum eine vorlexikalische Einheit überhaupt existiert. Zweitens wäre dann wichtig, wie diese aussieht und organisiert ist. Schließlich müsste man noch spezifizieren, wie sie aus dem Signal extrahiert wird, ob sie als Repräsentationseinheit gespeichert wird und wie sie auf das mentale Lexikon zurückgreift. All diese dringenden Fragen können hier nicht beantwortet werden. Es steht offen, welche der hier gegebenen Theorien zur Lösung dieser Problembereiche am nächsten kommt. Die Fragestellungen bilden jedoch eine Art Leitfaden für die Beschreibung und kritische Betrachtung der Modelle. Die Modelle werden nämlich schon in der Absicht ergründet, eine genauere Vorstellung über die Repräsentationseinheiten zu erlangen, um diese in Hinblick auf die anstehende Simulation besser bewerten zu können und ein geeignetes Modell für den Abgleich im Paralexikon der Simulation zu finden.

Zunächst werden einige Stichproben aus den derzeitig vielzitierten computer- und psycholinguistischen Simulationen einen Überblick als Einführung verschaffen, bevor eine vertiefende Beschreibung und Zuordnung zur Abbildung 3.1 erfolgt. Die große Mehrheit der Simulationen unterstellt phonemische Repräsentationen.

Die zeitgenössischen psycholinguistischen Modelle der auditiven Sprachverarbeitung gehen davon aus, dass Phoneme die relevanten sublexikalischen Signale sind (Jacobs, 2003, 132). Für die Modelle, die sich im engeren Sinne auf die Wortsegmentierung im Erstspracherwerb konzentrieren, gilt Gleiches:

Nearly all previous statistical analyses of infant word segmentation have used the phoneme as the unit over which infants are assumed to perceive speech and compute probabilities (Swingley, 2005, 93)

Massaro (1994) kommentiert diesen Umstand jedoch kritisch:

it has been accepted without question that phonemes are perceived. (Massaro, 1994, 222)

Auch Pinker merkt an:

the models have never seriously dealt with the problem of phonological representation. (Pinker, 2001, 161)

In der Tat findet man wenig Abweichendes. McClelland und Elman (1986) unterstellen einfach das Phonem neben phonetischen Merkmalen und Wörtern als grundlegende Einheit in ihrem TRACE-Modell, ohne eine kritische Anmerkung zu geben. Shortlist als dessen verbesserte Modifikation wird an dieser wichtigen Spezifikation nicht kommentiert. Ebenso widmet Brent (1999b) oder Brent (1999a) keine Überlegung auf die *perceptual units* seiner Modellierungen. Batchelder (2002) initialisiert das Lexikon mit phonologischen oder graphemischen Einheiten (Phoneme oder Buchstaben) für ihren BootLex-Algorithmus. Selbst Davis (2003), trotz seiner weitreichenden Einführungen in

⁴Die Diskussion des Einflusses visueller Informationen (vgl. Massaro (1998); Roy und Pentland (2002)) sei hier ausgespart. Sie beinhaltet sicherlich zusätzliche Informationen, aber es kann als gesichert gelten, dass das Sprachsignal auch ohne Rückgriff auf die visuelle Unterstützung hinreicht, bedenkt man die Sprachlernfähigkeit von blinden Menschen.

die frühkindlichen Fähigkeiten, bleibt eine kritische Auseinandersetzung über die Perzeptionseinheiten schuldig. Cairns et al. (1997) stellen genau die interessierenden Fragen in ihrer Einführung:

Does the infant require, at the outset of developing a speech processing capacity, an inbuilt sensitivity to certain structures, such as syllables or words, for instance, or to rhythmic structure itself? (Cairns et al., 1997, 112)

und

What linguistic units are important to the development of segmentation, and how large a window onto the speech is useful? (Cairns et al., 1997, 112)⁵

Leider kommen die Autoren auf den benannten Sachverhalt der linguistischen Einheiten später nicht mehr ausführlich zurück. Das PARSE-Modell (Perruchet und Vinter, 1998) verwendet *Chunks*, in Bezug auf die experimentelle Anordnung der Experimente von Saffran et al. (1996a), Saffran et al. (1996b) und Saffran et al. (1997a). Ihr Modell schafft auch für die Wahl von nach Häufigkeit ausgewählten Lautketten zwischen ein und drei Segmenten (*Chunks*) eine recht überzeugende Segmentierung. Das TRN (*Temporal Recurrent Network*; Dominey (2000)) simuliert die Ergebnisse von Nazzi et al. (1998) mit der Schlussfolgerung, dass schon Neugeborene Vokale und Konsonanten unterscheiden sollten, möchte man die Ergebnisse mit den Experimenten vereinbaren.⁶ Die angeführten Beispiele verdeutlichen, dass das Problem der linguistischen Einheiten entweder in das Bewusstsein einzugliedert, aber nicht gelöst werden konnte, oder dass jegliche Diskussion darüber strikt vermieden wurde. Durch die Ausklammerung dieses wichtigen Themenbereiches schwingt dann aber immer unterschwellig der Eindruck anheim, die Gedanken anderer Wissenschaftler würden aus Sorge außer Acht gelassen, eine mögliche Schwachstelle in den Annahmen des eigenen Modells zugeben zu müssen.

Damit muss die erste der oben genannten Fragen zunächst anhand von Worterkennungsmodellen aufgegriffen werden: Existiert eine vorlexikalische Repräsentation überhaupt, d.h. wird das Sprachsignal direkt auf das mentale Lexikon projiziert oder liegt dazwischen eine weitere Ebene, die zu Abstraktionen in der Lage ist? Wie in Abbildung 3.1 skizzenhaft veranschaulicht, sind beide Spielarten präsent. Dennis Klatt (1979) vertritt die Auffassung einer direkten Abbildung von Spektren zum mentalen Lexikon. Er legt überzeugend dar, dass auf theoretischer Ebene mindestens ein Algorithmus existiert, der nur mit der *Bottom-up*-Strategie den lexikalischen Zugriff direkt erklärt. Sein LAFS (*Lexical-Access-from-Spectra*)-Modell bewältigt alle typischen Problembereiche des Sprachaufnahme Prozesses. LAFS basiert dabei auf dem SCRIBER-System, welches zur Identifizierung 10ms und 20 ms lange Ausschnitte als Diphone definiert. Diese werden

⁵ Angesichts der Bedeutung dieses Aufsatzes in der wissenschaftlichen Diskussion zur Wortsegmentierung ist es umso erstaunlicher, dass diese Fragen nicht in der zu erwartenden Tiefe behandelt wurden. Die Beschreibung der Simulationen ist stark verkürzt dargestellt, sodass es stellenweise der Fantasie des Lesers obliegt, welche linguistischen Einheiten die Autoren als Input verwendeten. Eindeutig ist der Input ihres neuronalen Netzwerks. Hier wurden 9-Bit-Merkmalvektoren zur Beschreibung der phonemischen Einheiten verwendet.

⁶ Dabei ist noch zu bemerken, dass sich der prominente Co-Autor, Jacques Mehler, als ein Verfechter der Silbe als grundlegende Einheit hervorgetan hat.

3 Theorien der Wortsegmentierung

dann in LAFS mittels eines *Forward*-Netzwerkes im mentalen Lexikon phonologischen Repräsentationen zugeordnet.

Das ERIS-Modell von Marcus (1981) baut ebenso wie Klatt (1979) auf der Idee von Wickelgren (1969) auf, das besagt, dass es grundsätzlich keine Repräsentationen für Phoneme geben kann, weil sie Gebilde einer vereinfachenden Theorie sind, aber keine reale Entsprechung haben.⁷ Demzufolge verwendet ERIS 19-Bit-Vektoren, die aus dem Sprachsignal zehn Eigenschaften extrahieren. Darunter sind die Stimmhaftigkeit, die Amplitude und die Formanten sowie ihre Änderungen zwischen 10 ms und die Bandbreite. Auf diese Weise wurden auch Zielwörter in einem Lexikon kodiert. Mittels Wahrscheinlichkeiten der Form $P(S_i|X)$, wobei S_i der Vektor und X das Wort definiert,⁸ konnten schließlich die Wörter des Lexikons mit den Wörtern des Korpus abgeglichen werden. Es stellte sich heraus, dass für alle verwendeten Wörter eine Erkennung eindeutig möglich war.⁹

Stevens (1986) spezifiziert ein Modell (*LAFF: Lexical-Access-From-Features*), welches sich stark an LAFS (Klatt, 1979) anlehnt. Der wesentliche Unterschied liegt in der Verwendung von binären Merkmalen, wie sie von Chomsky und Halle (1968) definiert wurden anstatt der von Klatt verwendeten *Schablonen*. Auf diese Weise vermeidet es den Abgleich mit den komplexen spektralen *Templates* in einem rechenintensiven Netzwerk. Die Abstraktion eines reichhaltigen Spektrums auf ein duales System geht natürlich nur mit dem Nachteil einer eher unzuverlässigen Extraktion der Merkmale aus dem akustischen Signal einher und so bestehen Zweifel, ob ein lexikalischer Zugriff so funktionieren kann (Klatt, 1989, 192 u. 217), zumal eine ingenieurstechnische Implementierung weiterhin aussteht.

Das FUL (*Featurally-Underspecified-Lexicon*)-Modell (Fitzpatrick und Wheeldon, 2000, 140) „... is intended to be both human speech perception ... and automatic speech recognition ...“. Der lexikalische Zugriff erfolgt direkt nach der Extraktion der phonetischen Eigenschaften aus dem Signal. Die phonetischen Eigenschaften wurden deshalb gewählt, weil die Erkennung von segmentalen Einheiten keine befriedigenden Ergebnisse brachte.¹⁰ Natürlich überlappen sich die phonetischen Merkmale und können deshalb nicht vollständig erfasst werden. Dieser unvollständige Teil wird schließlich im Lexikon vervollständigt. Hier liegen Repräsentationen von phonemisch kodierten Wörtern vor, auf welche nun die Eigenschaftsvektoren abgeglichen werden, d.h. mittels logischer Funktionen werden widersprüchliche Szenarien ausgeschlossen.

Die psycholinguistische Forschung nähert sich dem Problem nicht aus der rein akustischen Perspektive, die, wie eben gesehen, für die Ingenieurwissenschaften typisch ist. Hier werden modelltheoretische Vorhersagen nicht mit technischen Implementierungen, sondern mit dem Experiment überprüft und anhand der Ergebnisse können dann ebenso

⁷vgl. Harris und Lindsey (1993) für eine gegenteilige Auffassung.

⁸beziehungsweise für $P(S_i|N)$, wobei N ist das Nichtwort definiert.

⁹Der Autor erwähnt explizit, dass ERIS den Teil der auditiven Analyse des *Logogen*-Modells (Morton, 1969) simuliert, der im *Logogen* hinsichtlich der akustischen Einheiten offen geblieben war.

¹⁰wie bei den Diphonen (Klatt, 1979)

plausible Modelle erstellt beziehungsweise mit Simulationen überprüft werden. Die Idee Stevens (1986) ist aus psycholinguistischer Sichtweise bei Marslen-Wilson und Warren (1994) deutlich erkennbar. Aufgrund unvereinbarer experimenteller Ergebnisse haben die Autoren eine veränderte Version des ursprünglichen Kohortenmodells (Marslen-Wilson und Welsh, 1978)¹¹ vorgestellt.

... featural information extracted from the speech input is projected directly onto the lexical level, where there is no prelexical phonemic level, and where all phonemic or phonetic judgements are fundamentally postlexical and postperceptual in nature ... (Marslen-Wilson und Warren, 1994, 673)

Zusammenfassend lässt sich über die fünf vorgestellten Modelle eines direkten Lexikonzugriffs festhalten, dass eine „Repräsentation“ (Spektra, Vektoren/Merkmale) zwar unterstellt wird, diese aber nicht im Sinne einer vorgelagerten lexikalischen Repräsentation zu verstehen ist. Vielmehr handelt es sich dabei bereits um Repräsentationen, welche dem mentalen Lexikon selbst Struktur geben und die man so auch im Sprachsignal wieder findet.

Worterkennungsmodelle, die eine vorlexikalische Repräsentation unterstellen, stehen in Sachen Plausibilität mit ihren Pendanten auf gleicher Höhe. Diese Modelle konzentrieren sich allerdings auf den Abgleich der vorlexikalischen Repräsentation mit dem Lexikon, nicht aber um die Extraktion dieser Einheiten aus dem Sprachstrom. 1969 hat Morton begonnen, ein einflussreiches Modell (Morton, 1969) vorzustellen, das die Worterkennung erklären helfen sollte. Danach wird jedes Wort durch ein *Logogen* repräsentiert. Auf diese Repräsentationsform nehmen drei Merkmale Einfluss: auditive, visuelle und semantische. Jedes dieser Merkmale setzt sich wiederum aus charakteristischen Eigenschaften zusammen. Immer, wenn eine dieser Eigenschaften eine Entsprechung in einem *Logogen* findet, wird ein Zähler um eins erhöht. Treffen nun mehrere Eigenschaften für genau ein *Logogen* zusammen, erreicht der Zählerstand irgendwann einen Schwellenwert und das *Logogen* „feuert“. An dieser Grenze fällt die Entscheidung über ein bestimmtes Wort. Häufigkeiten werden berücksichtigt, indem der Schwellenwert bei jedem Feuern sinkt. Bei unvollkommenen Stimuli auditiver oder visueller Art stehen nun weniger Eigenschaften der Merkmale zur Verfügung. Sollte das *Context-System* diese Unzulänglichkeiten nicht kompensieren können, kann für diese Stimuli konsequenterweise kein *Logogen* aktiviert werden. Das Wort wird also nicht erkannt. Das Modell baut auf früheren Forschungen von John Morton und einige seiner Mitstreiter auf und wurde immer wieder verbessert, konnte sich aber auch schon damals nicht von seinem unspezifischen Charakter lösen. Darin besteht auch die Hauptkritik, die Pisoni und Luce (1987) passend wie folgt formulierten:

At best, the theory helps to conceptualize how an interactive system may work and how word frequency and contextual effects in word recognition may be accounted for. However, the theory says very little, if anything, about precisely how acoustic-phonetic and higher-level sources of information are integrated, the time-course of word recognition, the nature of the perceptual units, or the role of the lexicon in word recognition. (Pisoni und Luce, 1987, 40)

¹¹Beschreibung siehe weiter unten

3 Theorien der Wortsegmentierung

Man kann das Modell im Nachhinein in Abbildung 3.1 der direkten Aktivierung für den lexikalischen Abgleich zuordnen, weil das *Logogen* simultan mit allen lexikalischen Einträgen verglichen wird. Hinsichtlich des *lexical alignment* trifft hier eine vollständige Einpassung in Abhängigkeit von Wortklassen zu. Auf der lexikalischen Verarbeitungsebene muss es der lexikalischen Konkurrenz zugeschrieben werden.

Wie das *Logogen*-Modell rechnet man das ursprüngliche Kohortenmodell von Marslen-Wilson und Welsh (1978) der direkten Aktivierung beim lexikalischen Abgleich zu. Allerdings erfolgt die lexikalische Einpassung eher positional und ergebnisorientiert, weil die phonologische Repräsentation zugänglich sein muss bevor das Ende des Wortes erreicht wird. Im Kohortenmodell beschreiben Marslen-Wilson und Welsh wie, *Bottom-up*- und *Top-down*-Mechanismen bei der Worterkennung miteinander interagieren. Die Autoren waren zwar nicht direkt an den kleinsten Repräsentationseinheiten interessiert, aber die Erkenntnisse sind auch für die Fragestellung des Phonems oder der Silbe nutzbringend, weil sie auf indirekte Weise das Phonem als linguistische Einheit in ihrem Experiment enthüllen.¹²

The Theory, as currently formulated, would never work if the internal structure of words could not be described formally as a sequence of segment-like units (Pisoni und Luce, 1987, 41)

Zunächst gingen die Autoren von einem unvollkommenen Sprachsignal aus. Das bedeutet, dass nicht alle relevanten Informationen im Eingangssignal vorhanden sind und von höheren kognitiven Prozessen bereitgestellt werden müssen. Zur Überprüfung dieser Hypothese wurden in einem Experiment die Versuchspersonen angehalten, zwei Aufgaben zu erfüllen. Erstens, und diese ist hier für die Fragestellung der kleinsten linguistischen Einheiten wichtig, eine *shadowing task*; zweitens eine *detection task*. Bei der *shadowing task* musste ein Satz wiederholt werden, in dem ein Phonem manipuliert wurde, ohne dass es aber den Versuchspersonen hätte auffallen können. Der Erkennungsprozess lief also unbewusst ab. Bezüglich der Fragestellung, ob Phonem oder Silbe die relevanten linguistischen Einheiten sind, wird damit ein Hinweis für das Phonem gegeben, weil es als abgeschlossene Einheit verarbeitet wurde. Es musste nicht in einer Silbe zur weiteren Verarbeitung eingebettet sein, wie es später etwa von Pallier et al. (1993) und Pallier (1998) behauptet werden würde. In der *detection task* wurden die Versuchspersonen schließlich aufgefordert, die manipulierten Phoneme direkt anzugeben. Aus den Ergebnissen folgern die Autoren dann, dass sowohl *Bottom-up*- und *Top-down*-Mechanismen genutzt werden müssen, diese aber in Abhängigkeit von der Vorhersagbarkeit des Wortes stehen.

Am besten lassen sich diese experimentellen Ergebnisse mit einer sequentiellen Anordnung von prälexikalischen Repräsentationen¹³ beschreiben, welche in strikter links-nach-rechts-Anordnung organisiert sind. Das Sprachsignal wird dabei in zeitlicher Abfolge

¹²vgl. die Studie von Marslen-Wilson (1984, 146) als einen Gegenbeweis zur Silbe und für das Phonem als Repräsentationseinheit.

¹³Marslen-Wilson und Welsh (1978) vermeiden jegliche genauere Spezifizierung der Inputeinheiten. Wie aber bereits Pisoni und Luce (1987, 41) bestätigt haben, ist das Modell ohne einen phonemischen Input nicht haltbar.

mit den Repräsentationen verglichen. Dadurch bilden sich in Abhängigkeit vom Verlauf des Signals so genannte Kohorten, die sich ständig verkleinern, je nachdem wie viele Segmente des Signals noch mit den gespeicherten Repräsentationen übereinstimmen. Sobald eine Kohorte durch ein entsprechendes Signal aktiviert wurde, können auch die *Top-down*-Prozesse eingeschaltet werden. Diese lassen schließlich weitere Informationen über den semantischen Gehalt und syntaktische Strukturen einfließen. Problematisch ist, dass die Häufigkeit bestimmter Wörter vom Modell nicht erklärt werden kann, obwohl derartige Effekte experimentell eindeutig bestimmt sind.¹⁴ Zweitens gestaltet sich die invariable Links-nach-rechts-Anordnung schwierig, da bei Aussprachefehlern der ersten Laute die falsche Kohorte aktiviert würde und deshalb selbst die *Top-down*-Prozesse nicht das richtige Wort finden können. Dieses Verhalten entspricht nicht den experimentellen Beobachtungen.

Forster (1976) stellt ein Suchmodell für den lexikalischen Zugriff vor, das sich äußerlich stark an informatischen Strukturen ausrichtet. Dies spiegelt sich auch in den Begrifflichkeiten wider. So führt er einen *Masterfile* ein, der für jeden seiner Einträge drei verschiedene *Pointer* von *Access files* erhält. Die *Access files* kodieren die orthographische, die phonologische und die semantisch-syntaktische Form eines (Stimulus)Wortes. Dies soll die unterschiedlichen Verarbeitungsprozesse erklären helfen, die bei der akustischen oder visuellen Signalaufnahme anfallen, aber auch die Unabhängigkeit der semantisch-syntaktischen Verarbeitung in das Modell einbeziehen. Die Einträge, also die *Access codes*, werden nach ihrer Häufigkeit geordnet. Die sich daraus ergebenden *Cluster* oder Gruppen nennt Forster *bins*. In einer Weiterentwicklung des Modells (Forster, 1979) wird die grundsätzliche Idee beibehalten, aber die strukturelle Anordnung verändert. Der *Masterfile* wird im Wesentlichen zum Lexikon selbst, welches nun nach wie vor mit *Pointern* aus einem *Lexical Processor* versorgt wird. Dieser enthält aber nur noch die orthographischen und phonologischen *Access codes*. Getrennt davon wird ein syntaktischer Prozessor und ein *Message*-Prozessor für die Semantik eingespielt. Alle drei Prozessoren greifen auf ein *General Problem Solver* (GPS) zurück, der wiederum auf Regeln und Konzeptwissen aufsetzt.

Das *Fuzzy-Logic-Model-of-Phoneme-Identification* (FLMP) von Oden und Massaro (1978) kann auch als ein Modell zur Phonemerkennung beschrieben werden. Das Modell veranschaulicht in der hier angepassten Version¹⁵ die Informationsverarbeitung von konsonantischen Plosiven. Es basiert auf den Annahmen einer getrennten Wahrnehmung der Signaleinheiten und einer graduellen Bewertung der akustischen Eigenschaften im Vergleich zu einer rein diskret angelegten prototypischen Form. Damit fallen zur Phonemidentifikation drei Verarbeitungsphasen an. Zuerst wird bestimmt, welche graduelle Ausprägung ein akustisches Merkmal im wahrgenommenen Signal hat (*feature evaluation*). Danach werden diese Ausprägungen mit vorhandenen Prototypen verglichen und deren Abweichungen bestimmt (*prototype matching*), wobei hier aber auf CV-Repräsentationen

¹⁴Der Effekt der Wortwiederholung (Neisser, 1954) und der Frequenzeffekt (Broadbent, 1967) war auch zur Zeit der Entwicklung des Kohortenmodells gut bekannt. Saint-Aubin und Poirier (2005) führen den Frequenzeffekt auf die Bekanntheit und dessen *co-occurrence* gleichermaßen zurück.

¹⁵FLMP findet in verschiedenen Domänen Anwendung (Massaro, 1998, 61).

3 Theorien der Wortsegmentierung

zurückgegriffen wird, da im Vokal wesentliche Informationen über den Vorgängerkonsonanten erst entdeckt werden können. Letztlich werden daraus die Signaleinheiten, welche die geringsten Abweichungen vom Prototypen im Vergleich zu allen anderen Signaleinheiten haben, ausgewählt (*pattern classification*). Wichtig ist hier anzumerken, dass zur Phonemidentifikation keine diskreten Merkmale herangezogen werden, sondern sich die Entscheidung aus „verschwommenen“ Übergängen beziehungsweise Eigenschaftsdefinitionen ergibt.

Das vielleicht populärste Modell, TRACE (McClelland und Elman, 1986), ist eine Netzwerkimplementation bestehend aus drei Ebenen, die phonetische Merkmale (*features*), Phoneme und Wörter enthalten. Zwischen allen Ebenen und innerhalb einer Ebene (*lateral inhibition*) können Informationen in beide Richtungen ausgetauscht werden.¹⁶ Erfährt nun auf unterster Ebene beispielsweise das Merkmal der Stimmhaftigkeit ([+voicing]) eine entsprechende Erregung, führt dies zur Aktivierung von stimmhaften Phonemen (/b/, /d/, /g/) auf der nächsten Verarbeitungsstufe, welche diese Aktivierung an mögliche Wortkandidaten (*bed, good, dog, bus* oder *back*) zur letzten Ebene weitergibt. In der Wortebene können nun die ausgewählten Kandidaten in einen Wettbewerb treten. Im gegebenen Beispiel würden *bus* und *back* verlieren, weil sie nicht so stark aktiviert wurden, wie zum Beispiel *bed*. Diese Information wird nun wieder auf die Phonemebene zurück geleitet, auf welcher nun die Phoneme /s/ und /k/ gehemmt werden würden, somit werden automatisch andere Wörter mit diesen Phonemen benachteiligt. Je nach Implementation können nun auch phonetische Merkmale gehemmt werden; z.B. einige von denen, welche das /s/ in *bus* kodieren.

Diese Architektur ist in der Literatur sehr stark umstritten, weil es natürlich der Plausibilität in den Kognitionswissenschaften widerspricht, dass entgegen jeglicher experimenteller Beobachtungen bereits aktivierte Wörter in einem *Top-down*-Verfahren andere Kandidaten auf der Phonemebene oder gar tiefer auf der Ebene phonetischer Merkmale einmal mehr oder weniger hemmen. Ein zweiter Kritikpunkt ist die Initiierung eines Netzwerkes für jeden gewählten Zeitabschnitt (*time slice*) im Signal. Dies ist nicht nur sehr rechenintensiv, sondern widerspricht auch einfach der Beobachtung einer kontinuierlichen Verarbeitung des gesprochenen Wortes. Dabei scheint weniger von Belang zu sein, dass der Output der einzelnen Instanzierungen zusammen genommen ein recht brauchbares Ergebnis vorzuweisen hat. Diese zwei Nachteile konnten schließlich in einem Nachfolgemodell, Shortlist, verbessert werden.

In Shortlist (Norris (1994); Norris et al. (1995); Norris et al. (1997)) werden mögliche Wortkandidaten des Inputs in einer kurzen Liste (*short list*) aufgeführt. Diese Wörter stimmen in allen spezifizierten segmentalen Eigenschaften mit dem Input überein. Unter ihnen wird auch ein Wettbewerb durch „Hemmung“ erzeugt. Das bedeutet aber hier, nur wenn eine bestimmte Kombination der Wörter der Shortlist den Inputstrom vollständig segmentiert, gewinnen auch genau diese Wörter den Wettbewerb, ohne eine Zurück-

¹⁶Es ist auch möglich eine Implementierung von *Trace* durchzuführen, bei der Informationen nur zwischen Wörtern und Phonemen in beiden Richtungen erlaubt sind, nicht jedoch zwischen Merkmalen und Phonemen.

führung der Informationen auf die Phonemebene. Auf eine dritte Ebene (phonetische Merkmale) wird vollständig verzichtet. Die Knoten der anderen Wörter werden entsprechend im Netzwerk geringer gewichtet, also in ihrer Aktivierung gehemmt. Beispielsweise würden beim Input von *I take shampoo light* folgende Wörter in einen Wettbewerb treten: *I, take, ache, sham, shampoo, pool,*¹⁷ *light, lie*. Die optimale Kombination würde in diesem Fall nur die Wörter des Inputs treffen können, weil *ache* und *lie* das /t/ in *take* beziehungsweise *light* vernachlässigen und *pool* das /art/ in *light* nicht zuweisen könnte. Eine kleine Erweiterung im Ausgangsmaterial würde diese Eindeutigkeit aufheben. *I take shampoo light embodying the new lifestyle...* Nun gehören nämlich unter anderem auch *item,*¹⁸ *embody, embodying, body, bodying* zur Kandidatenliste und damit würde eine weitere Kombination alle Segmente definieren können. *I take sham pool item bodying the...* In *crucial instance* ist der Fall noch eindeutiger, weil *crew* und die schwache Form von *shall*, sowie *in* und *stance* neben den richtigen Kandidaten ebenso zugewiesen werden könnten. Um diese Unzulänglichkeiten auszuschließen können metrische Informationen des Inputs zuträglich sein (McQueen et al., 1994).¹⁹

SARAH (Mehler et al., 1995) ist eine modelltheoretische Formulierung, welche die Besonderheiten der experimentellen Befunde des lexikalischen Zugriffs erklären könnte. Natürlich kommt auch diese Spezifikation nicht ohne fragwürdige Annahmen aus. Dies ist umso wichtiger, weil keine Ergebnisse einer Simulation vorliegen. Bereits der Name des Modells verrät das Erklärungsziel (*Syllable, Acquisition, Representation, Access Hypothesis*). Das Modell soll erklären wie „the forms of spoken words are acquired“ (Mehler et al., 1995, 255), nicht aber wie der Inhalt des Lexikons aufgenommen wird. Es vereinigt bis zu einem bestimmten Grad beide Varianten der phonemischen und silbischen Verarbeitung in zwei von drei Annahmen des Modells. Die erste Annahme wird mit einem silbischen Filter umschrieben, der dafür sorgt, dass nur eindeutige Silben übernommen werden. Die zweite Annahme ist ein *phonetic analyzer*, der die Silbe anhand der Phoneme beschreibt und drittens definiert der *word-boundary detector*, die Dauer und Betonung des Signals, um die Wortgrenzen zu bestimmen. Dieses Modell

... suggests that a baby represents speech at both at the syllabic and the phonetic levels.
(Mehler et al., 1995, 256)

Jusczyk (2001, 267) sieht in SARAH

... a strong correspondence between the processes used by infants in word segmentation and those underlying lexical access by adults.

Das Modell stellt also einen Zusammenhang zwischen den Wahrnehmungseinheiten bei Säuglingen und Erwachsenen her. Letztere benutzen die Einheiten für den lexikalischen Zugriff und Erstere für die Segmentierung von Wörtern aus dem Sprachstrom. Dabei handelt es sich um die gleiche Einheit.

Das *Neighborhood-Activation-Model* (NAM) von Luce und Pisoni (1998) versucht nun wieder das Problem nach dem Abgleich zwischen Signal und innerer Repräsentation zu

¹⁷Aussprachebedingt könnte man auch noch für *pull* als weiteren Kandidaten argumentieren.

¹⁸Vorausgesetzt, dass „e“ in *item* mehr zum „i“ gesprochen wird.

¹⁹Siehe für die experimentellen Befunde Kapitel 4.1.2

3 Theorien der Wortsegmentierung

beantworten. Die Frage lautet demgemäß: Wie schafft es das Wort, das in einem Umfeld anderer Wörter ausgesprochen wird, die entsprechenden Korrelate im mentalen Lexikon anzuregen? Luce und Pisoni unterstellen, dass die geäußerten Einheiten eine ganze Reihe von *acoustic-phonetic patterns* aktivieren müssen, ähnlich dem Kohortenmodell von Marslen-Wilson und Welsh (1978), nur dass das NAM nicht zwingenderweise Wörter als „Kohorte“ unterstellt und die zeitliche Reihenfolge vernachlässigt. Zwischen den akustisch-phonetischen Mustern wird aber ebenso eine Art Wettbewerb erzeugt. Zur Diskriminierung zwischen den aktivierten Einheiten dienen abermals unterschiedliche Aktivierungsebenen, welche abhängig von der Übereinstimmung des Inputs stärker oder schwächer erregt werden. Hier sind Parallelen zum *Logogen* offensichtlich, weil auch hier die *word decision units*, bevor der Abgleich im mentalen Lexikon erfolgt, Informationen zurückhalten können. Die gleichen Eigenschaften haben auch *Logogens* bevor sie die kritische Schwelle zum „feuern“ erreichen. Im NAM fließen nun weitere Informationen aus einer höheren Verarbeitungsstufe ein. Dies können häufigkeitskodierte Merkmale sein, die aber nicht intrinsisch vorhanden sind oder aber auch alle anderen *Top-down*-Informationen, die den Entscheidungsprozess optimieren würden. Das Modell trifft die Vorhersage, dass die Zeit für den Lexikonabgleich mit der Zahl der aktivierten *word decision units* steigt und auch die Genauigkeit des Abgleichs mindert.

In den Worterkennungsmodellen, die Bezug auf eine vorlexikalische Ebene nehmen, lässt sich erkennen, dass mit Ausnahme von SARAH abgetrennte Segmente, die im groben den Phonemen beziehungsweise Ketten von Phonemen entsprechen, unterstellt werden. Die nun weiter zu führende Diskussion über die Repräsentationseinheiten wird aus diesem Grund das Phonem in den Mittelpunkt rücken und insbesondere seine eventuellen Unzulänglichkeiten gegenüber der Silbe aus modelltheoretischer Perspektive betrachten müssen.

Die recht konzise Abhandlung von Worterkennungsmodellen hat sicherlich einen positiven Effekt auf die Wahl eines Modells für die anstehende Simulation hinterlassen. Aus den Ausführungen kann die Wahl eines Mechanismus für die lexikalische Einpassung gut begründet werden. Es sei aber ausdrücklich betont, dass die Fragen nach dem lexikalischen Zugriff oder Abgleich nicht Thema der Arbeit und der Simulation sind. Eine Einfügung von Worteinheiten muss aber in irgendeiner Weise im Algorithmus vorgenommen werden, um eine Rekursion zu erzeugen. Wie das aber genau geschieht, steht nicht im primären Interesse der Arbeit. Da, wie eingangs angeführt, Modelle, die einen direkten Abgleich zwischen Signal und mentalem Lexikon unterstellen, nicht favorisiert werden können, stehen die Modelle mit vorlexikalischer Repräsentation zur Auswahl. Dabei lassen die angeführten Darstellungen das Kohortenmodell von Marslen-Wilson und Welsh (1978) in seiner ursprünglichen Form für die Ziele dieser Arbeit als attraktiv erscheinen. Das Kohortenmodell ist nicht nur einfach algorithmisch zu beschreiben, also auch zu programmieren, sondern es ist zudem gut durch experimentelle Befunde abgesichert.²⁰ Der

²⁰Natürlich sprechen auch Befunde anderer Wissenschaftler gegen eine derartige Modellierung. Bedenkt man jedoch die Einfachheit einer Wortliste ohne Bedeutungsgehalt, welche sich sechs Monate alte Babys merken können, erscheinen alle Argumente, welche auf die Widersprüchlichkeit von Bedeutungen in einem Satz abzielen, nicht mehr relevant.

Verzicht auf Häufigkeitseffekte innerhalb der Kohorte ließe sich algorithmisch zumindest leicht beheben, sollte sich dieses Kriterium als wichtig herausstellen.

3.2 Wortsegmentierung ohne anfängliches Lexikon

Die genannten Segmentierungsstrategien basieren entweder auf segmentalen (Allophone, Übergangswahrscheinlichkeit) oder auf suprasegmentalen (Prosodien, Phonotaktik) Merkmalen.²¹ In dem, was folgt, wird zuerst ein Abriss über die experimentellen Befunde gegeben und falls vorhanden einige modellhafte Implementierungen vorgestellt. Jeder dieser Ansätze ist zur Wortsegmentierung geeignet und die darin spezifizierten Informationen sind auch ab einem bestimmten Alter dem Kind zugänglich. Allerdings ist für prosodische, allophonische und phonotaktische Strategien ein bestimmtes Quantum an Anfangsmaterial notwendig, um die Muster als Regel abzuleiten. Einleitend sind in Abbildung 3.2 die statistischen und prosodischen Entwicklungsstufen als Übersicht gegeben. Im oberen Teil der Darstellung können die Schlüsselexperimente dem Alter des Kindes zugeordnet werden. Als Beginn für die langsame einsetzende Entfaltung der Segmentierungsfähigkeiten wird der sechste Lebensmonat als sinnvoll angesehen (vgl. Jusczyk, 2002). Aus der räumlichen Anordnung der Quellen kann auch die zeitliche Entwicklung der Experimente ersehen werden.

3.2.1 Prosodie: Rhythmus und Betonung

Die suprasegmentalen Merkmale lassen sich im Sprachsignal auf drei verschiedenen Dimensionen physisch erfassen: Die Frequenz, die Dauer und die Lautstärke. Erstere gibt an wie oft sich die Perioden der überlagernden Sinuswellen während einer Sekunde wiederholen und wird auch oft mit Tonhöhe umschrieben, denn je häufiger sich eine Welle in einem bestimmten Zeitabschnitt wiederholt, desto schneller werden auch die Luftdruckänderungen auftreten, die dann einen höheren Ton erzeugen müssen. Die Dauer oder Länge gibt an wie viel Zeit für die Produktion eines Schallereignisses benötigt wird. Das Schallereignis kann ein einzelner Laut, eine Silbe, ein Wort oder eine anders spezifizierte Einheit sein. Letztlich ist die Lautstärke oder Intensität der Amplitudenausschlag, also die Veränderung auf der y-Achse im Oszillogramm. Kombiniert man nun diese Dimensionen entsteht ein Rhythmus, der durch die Betonung oder im linguistischen Sinne durch den Akzent auf Silben- oder Wortbasis und die Intonation auf phrasaler oder Satzbasis, bestimmt wird (vgl. Pickett (1999); Schukat-Talamazzini (1995); Reetz (2003)).

Der Rhythmus unterscheidet sich bei natürlichen Sprachen grundsätzlich. Die Unterschiede zwischen den Sprachen sind dabei graduell. Es gibt also Sprachen, die prosodisch ähnlich und andere, die sehr unterschiedlich sind (Norwegisch vs. Niederländisch

²¹Myers et al. (1996) und Mattys und Clark (2002) zeigen, dass auch die Wahrnehmung von Pausen als Anhaltspunkt genutzt werden könnte. Da aber Pausen nur ergänzend herangezogen werden, dementsprechend auch nur marginalen Einfluss besitzen, spricht man nicht von einer eigenständigen „Strategie“, sondern führt sie als zusätzlichen Einfluss an.

3 Theorien der Wortsegmentierung

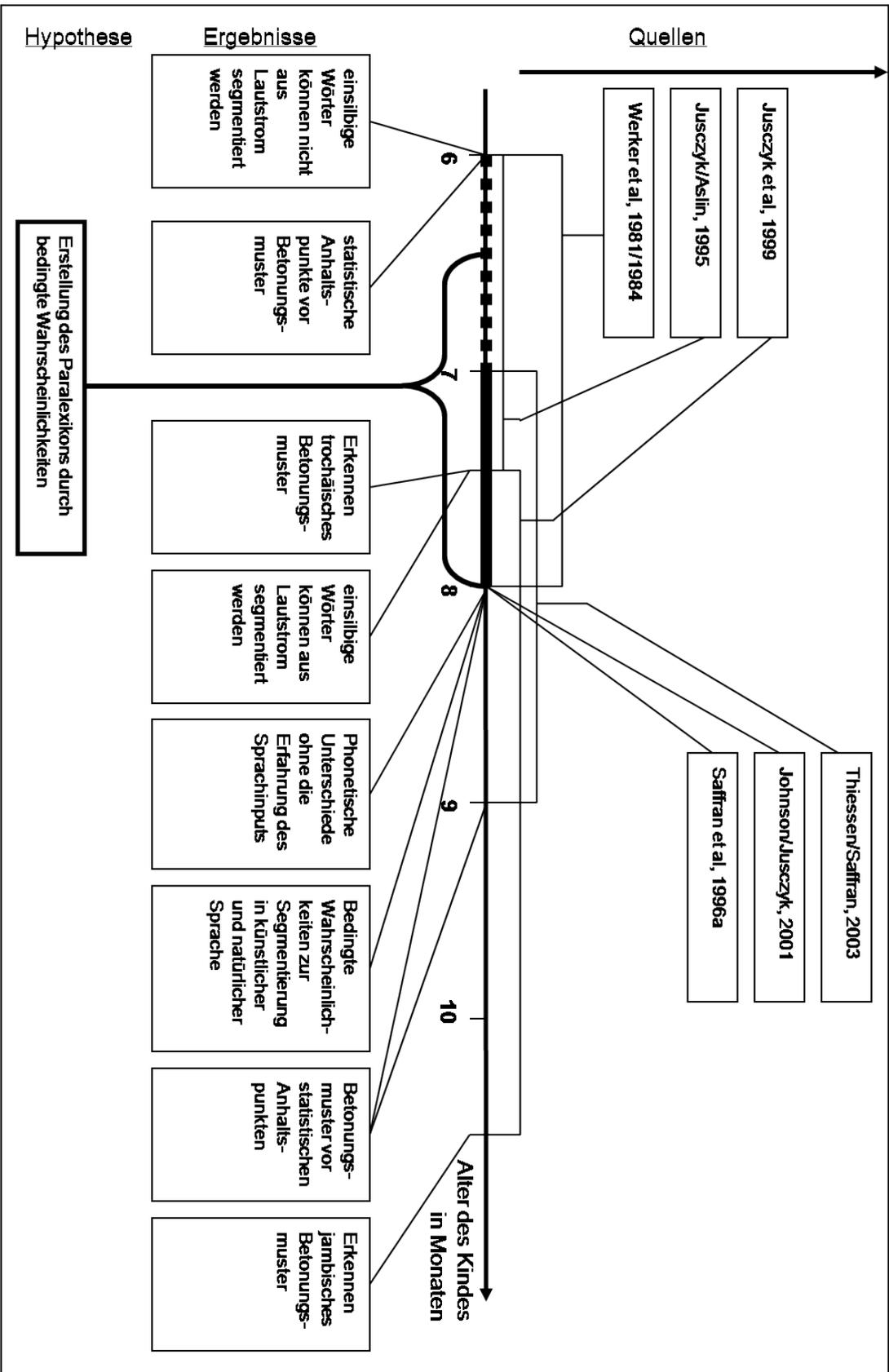


Abbildung 3.2: Entwicklung prosodischer und statistischer Sensibilität, Quelle: Eigene Darstellung

vs. English). Der genaue Rhythmus einer Sprache muss vom Kleinkind erlernt werden. Ab einem bestimmten Alter können Kleinkinder die Betonungsmuster auch zur Segmentierung nutzen. Dazu ist es notwendig, das spezifische Muster einer Sprache mit dem Zusammenfallen des Wortanfangs, -endes oder beidem zu verbinden.

Bei Experimenten von Kleinkindern konzentriert man sich vor allem auf Akzente, weil für das Kleinkind ein sich entwickelndes Lexikon unterstellt werden muss, das aus wortähnlichen Einheiten besteht. Erst im weiteren Verlauf der frühkindlichen Entwicklung erlauben die Kombinationen der Wörter größere Einheiten zusammenzustellen und damit auch Intonationen zu berücksichtigen.²² Dies soll nicht bedeuten, dass keine Anstrengungen für die Erkennung von Prosodien auf Satzbasis unternommen würden. Nazzi et al. (2000b) oder Soderstrom et al. (2005) stellen eine diesbezügliche Studie für sechs Monate alte Babys vor (siehe Kap. 4.3). Weitaus bekannter sind aber die Untersuchungen des Wortakzents, z.B. im Vergleich von rhythmischen Klassen bei fünf Monate alten Babys (Nazzi et al., 2000a).

Eine erste systematische und repräsentative Erfassung des rhythmischen Musters des Englischen haben Cutler und Carter (1987) durchgeführt. Das Ergebnis lässt sich recht einfach, wenn auch noch etwas ungenau, zusammenfassen.

Two computerized dictionaries were found to list approximately three times as many words beginning with strong syllables . . . as beginning with weak syllables. (Cutler und Carter, 1987, 133)

In der Tat gehen von dieser Untersuchung eine Reihe von Modellen in der metrischen Phonologie aus.²³ Cutler und Norris (1988) führen experimentell vor, dass starke Silben, wenn von schwachen Silben gefolgt, in Nichtwörtern als Anhaltspunkt für eine mögliche Segmentierung bei Erwachsenen genutzt werden. Bei einer konkurrierenden Umgebung (schwache Silben gefolgt von starken Silben) ist dies hingegen nicht der Fall. Beispielsweise wird *mint* in *mintesh* schneller erkannt als in *mintayne*. Während ersteres dem mehrheitlichen trochäischen Muster des Englischen entspricht (stark – schwach), folgt das zweite Nichtwort einer Ausnahme (stark – stark). Damit setzen die Autoren einen der Anfangspunkte zur Entwicklung der metrischen Segmentierungsstrategie.²⁴

Um dieses Verhalten auch bei Kleinkindern nachzuweisen, demonstrieren Jusczyk et al.

²²Cutler (1976) stellt ein Experiment diesbezüglich mit Erwachsenen (*undergraduates*) vor, in der sie den Zusammenhang zwischen Intonation und Akzent (*stress*) eines Satzes untersucht. Ein Zielphonem wurde schneller erkannt, wenn die Intonation als auch der Akzent dem gewöhnlichen Aussprachemuster entsprechen. Wird die Betonung des Wortes geändert, erhöht sich auch die Reaktionszeit.

²³In Cutler (1995) finden sich detailliertere und besser aufbereitete Daten zur *Metrical Segmentation Strategy*, die aus probabilistischer Perspektive diskutiert werden.

²⁴Norris et al. (1997) und Norris et al. (2001) fügt später der spezifischen Natur der metrischen Segmentierungsstrategie mit dem *Possible Word Constraint* (PWC) eine weitere Theorie hinzu und versucht sie in den Stand einer universalen Strategie zu heben. In Norris et al. (1997) entdeckten die Wissenschaftler, dass ein einzelner Konsonant schwieriger zu entdecken war als eine Silbe (*apple* in *fapple* brauchte mehr Reaktionszeit als in *vuffapple*). In Norris et al. (2001) sollte dann gezeigt werden, dass dieser Effekt universalen Charakter besitzt. Dabei wurden Zusammensetzungen verwendet, die für das Englische nicht möglich sind und es zeigte sich die gleiche Verzögerung. In der Diskussion schließen sich dann weitere Beispiele aus der Literatur an (z.B. Sesotho und Niederländisch).

3 Theorien der Wortsegmentierung

(1993a), dass auch Versuchspersonen im Alter von neun Monaten das dominante Akzentmuster des Englischen der Form stark – schwach²⁵ dem umgekehrten Fall (schwach – stark) vorziehen. Diese Präferenz lässt sich bei sechs Monate alten Babys noch nicht nachweisen. Das Herausfiltern der tiefen Frequenzen unter 400 Hertz stellt dabei sicher, dass sich die Probanden tatsächlich nach dem Akzentmuster richten müssen, da nun lediglich die Prosodien erkennbar bleiben, nicht aber die rein phonetischen Informationen, deren Grundfrequenz nicht mehr vorhanden, deshalb unkenntlich geworden ist.

Wie bereits im Kapitel 3.1 angeklungen ist, versuchen Mehler et al. (1995) mittels des SARAH-Modells den lexikalischen Zugriff zu erklären. Eine wichtige Komponente dabei sind die prosodischen Informationen, die als Anhaltspunkte während der *acquisition* herangezogen werden sollen:

... given a handful of words sufficiently representative of the language, the algorithm decides whether the language is stress initial, terminal stress, etc. (Mehler et al., 1995, 254)

Abermals wird angeführt, dass diese Informationen aus isoliert vorkommenden Wörtern abgeleitet werden kann. Damit wird die metrische Segmentierungsstrategie von Cutler und Norris (1988) übernommen. Die Grundfrage bleibt jedoch bestehen: Woher kann das Kind wissen, ob es sich um isolierte Wörter handelt?²⁶ Die im Aufsatz so sehr betonte Frage nach dem Wie stößt spätestens an dieser Stelle an ihre Grenzen. In der gleichen Ausgabe bemerkt Cutler (1995, 118), dass die Betonung der Silbe im Prinzip in genügendem Maße in den Vokalen enthalten ist. Hier wäre ein möglicher Lösungsansatz gegeben, da die Wahrnehmung von Vokalen eine Sonderrolle einnimmt. Andererseits müssen auch hier wieder Parallelen zu anderen Sprachen gezogen werden. Cutler (1995, 119) gesteht selbst ein, dass das Englische als *free-stress language*, das Französische als *nonstress* und das Polnische als *fixed-stress language* anzusehen ist. Eine einfache Segmentierungsstrategie auf universaler Ebene kann man mit der Qualität der Vokale allein also nicht definieren.

Juszyk und Aslin (1995) zeigen, dass siebeneinhalb Monate alte Babys einsilbige Wörter in einem Lautstrom wiedererkennen können (Experiment 1). Nur wenige Wochen früher (bei sechs Monaten), sind sie dazu nicht in der Lage (Experiment 2). Ändert man auch nur ein phonetisches Merkmal der Wörter bei der Konditionierung, können auch die siebeneinhalb Monate alten Babys die Wörter nicht mehr im Lautstrom erkennen (Experiment 3). Dieser Befund legt nahe, dass die Probanden schon sehr detaillierte Informationen verarbeiten müssen. Schließlich lässt sich das erste Experiment wiederholen, wenn man die Wörter in der Trainingsphase nicht isoliert, sondern in anderen Sätzen vorspielt (Experiment 4). So kann man doch mit den Worten der Autoren schlussfolgern:

Information about such regularities in the input, be they prosodic or phonetic properties, is apt to prove useful in identifying sequences of sounds that are likely to be an important component in forming and building up a lexikon. (Juszyk und Aslin, 1995, 17)

²⁵Das bedeutet, dass die betonte (starke) Silbe von der unbetonten (schwachen) Silbe gefolgt wird, wie in *pliant, falter, donor, comet, neighbor, butter, final*, etc. Das umgekehrte Muster (unbetonte Silbe wird von betonter Silbe gefolgt) existiert zwar auch ist aber sehr viel seltener. Beispiele hierfür sind: *comply, befall, define, restore, resent, assign*, etc.

²⁶Für eine detaillierte Diskussion dazu siehe Kapitel 3.3.

3.2 Wortsegmentierung ohne anfängliches Lexikon

Newsome und Jusczyk (1995) replizieren zunächst noch einmal die Ergebnisse von Jusczyk und Aslin (1995) mit mehrsilbigen Wörtern und positiver Evidenz. Aufbauend dazu können nun die Prosodien näher untersucht werden. In zwei Folgeexperimenten soll gezeigt werden, ob der Sprachlerner die starke Silbe als entscheidenden Anhaltspunkt zum Erkennen von Einheiten nutzt, wie es z.B. von Brown (1968) oder Gleitman und Wanner (1982) vorgeschlagen wird. Die Idee ist hier, den Sprachlerner mit der starken Silbe eines Wortanlauts zu konditionieren (*king* oder *ham*) und in der Testphase das gesamte Wort (*kingdom* oder *hamlet*) vorzuspielen. Lassen sich signifikante Wiedererkennungseffekte messen, so wäre doch ein Hinweis gegeben, dass das Kind diese Teilmenge im zweisilbigen Wort erkennt. Die Experimente konnten dies aber nicht belegen.²⁷ Aus diesem Ergebnis muss man schließen, dass immer das gesamte Wort abgeglichen wird, nicht aber nur die silbischen Bestandteile. Das Muster, stark–schwach, in zweisilbigen Wörtern wird wohl nur unteilbar zugeordnet. Dies entspricht natürlich auch den Vorhersagen der metrischen Segmentierungsstrategie (Cutler und Norris, 1988). Trifft diese aber zu, so dürften die Sprachlerner das gegenteilige Muster, schwach – stark, nicht erkennen. Sie müssten annehmen, dass zwischen beiden Silben Wortende beziehungsweise -anfang liegen. Diesen Befund lieferte in der Tat das vierte Experiment. Denn wiederholt man die Experimente zwei und drei mit dem neuen Stimulusmaterial der schwach – starken Muster, gibt es einen signifikanten Effekt. Die Kinder hören *guitar*, eingebettet in einer Textpassage, wesentlich länger zu, wenn sie mit *tar* konditioniert wurden. Im Zusammenspiel mit den Vorgängerexperimenten kann dieses Ergebnis nur bedeuten, dass sie *guitar* nicht als ein Wort anerkennen, sondern möglicherweise nur *tar*.

Morgan und Saffran (1995) untersuchen wie sequentielle und suprasegmentale Informationen beim Herauslösen von wortähnlichen Einheiten zusammenwirken. Sie finden heraus, dass

9-month-olds appear to be capable of integrating sequential and suprasegmental information in forming wordlike (multisyllabic) phonological percepts, 6-month-olds are not. (Morgan und Saffran, 1995, 911)

Der Nebeneffekt dieser Studie, ein Nachweis über die Fähigkeit der phonologischen Analyse, könnte ergänzend in Kapitel 4.1.2 angeführt werden. Hier soll dieser Befund nicht weiter interessieren, sondern eher die Fähigkeit der suprasegmentalen Wahrnehmung im Zusammenspiel mit den sequentiellen Informationen. Die ersten Experimente zeigen, dass neun Monate alte Babys zwei Silben immer dann als zusammengehörig betrachten, wenn zwei weitere Regelmäßigkeiten als Bedingungen erfüllt sind. Erstens muss die Rhythmik und zweitens die Reihenfolge der Silben dem erlernten Muster entsprechen. Wird nur eine der Bedingungen verletzt, kann kein Wiedererkennungseffekt nachgewiesen werden. Für sechs Monate alte Probanden sind diese Bedingungen nicht zwingend. Sie können die Reihenfolge der Einheiten vernachlässigen und meinen auch bei richtigem rhythmischen Muster, aber falscher Reihenfolge, den gleichen Stimulus vor sich zu haben. Das bedeutet, dass sich die Wahrnehmungsleistung innerhalb von wenigen Wo-

²⁷Weder für die Konditionierung mit der Wortliste und einem Test mit den Textpassagen (Experiment 2), noch, um andere Effekte auszuschließen, für den umgekehrten Fall (Experiment 3): Training mit ganzen Passagen und Test mit der Wortliste.

3 Theorien der Wortsegmentierung

chen sehr entscheidend verfeinert hat. Während mit sechs Monaten, eine Verknüpfung von rhythmischen und segmentalen Mustern nicht möglich ist, beherrschen neun Monate alte Babys dieses Zusammenspiel recht überzeugend.

Echols et al. (1997) führte drei Untersuchungen durch, um die Wahrnehmungsleistung der sprachlichen Rhythmik bei Erwachsenen und Babys herauszuarbeiten. Erwachsene englische Sprecher bevorzugten erwartungsgemäß das trochäische Betonungsmuster. Dies tun auch neun Monate alte Babys, aber noch nicht sieben Monate alte. Das dritte Experiment zeigt dann, dass die neun Monate alten Versuchspersonen diese Fähigkeit auch zur Segmentierung von Wörtern nutzen können. Dies geschieht aber nur mit dem trochäischen und nicht mit dem jambischen Betonungsmuster.

Da im Englischen die jambischen Betonungsmuster, wenn auch sehr viel seltener als die trochäischen, präsent sind und diese auch erlernt werden müssen, um einen Lautstrom segmentieren zu können, wäre es nützlich zu wissen, ab welchem Zeitpunkt beide Fähigkeiten denn zur Verfügung stehen. Die Beantwortung dieser Frage ist Jusczyk et al. (1999a) gelungen. Nach einer Reihe von 15 Experimenten können sie das Alter auf zehneinhalb Monate eingrenzen. Zuerst zeigen die Wissenschaftler, dass siebeneinhalb Monate alte Babys in der Lage sind, Wörter mit trochäischen Muster in einem Lautstrom zu erkennen. Diese Fähigkeit ist aber noch sehr begrenzt und verfügt noch nicht über das notwendige Abstraktionsvermögen, das zur umfangreichen Regelableitung notwendig wäre. Die Erkennungsleistung tritt nämlich nur dann auf, wenn sie vorher die betreffenden Wörter in Listen oder die Textpassagen präsentiert bekommen haben. Diese Ergebnisse sind schon von Jusczyk und Aslin (1995) oder Newsome und Jusczyk (1995) bekannt. Verändert man nun nur leicht das Trainings- und Testmaterial ist kein Hinweis für eine mögliche Segmentierung gegeben. So kann man lediglich die starken Silben im Test verwenden und sie werden mit Teststimuli der trochäischen Form nicht wiedererkannt. Das gleiche gilt natürlich für den umgekehrten Fall und andere Kombinationsmöglichkeiten von Test- und Trainingsstimuli (isoliert oder in Textpassagen).

Abermals hat man hier den Beweis erbracht, dass Kinder diesen Alters auf das gesamte Wort und nicht nur auf den Silbenanlaut der starken Silbe reagieren. Im weiteren Verlauf der Studie wurden die Stimuli des jambischen Betonungsmusters verwendet. Für siebeneinhalb Monate alte Babys ist dieses Betonungsmuster noch nicht anwendbar. Das Gegenteil ist der Fall. Wie das 11. Experiment zeigt, glauben sie, dass sie das bekannte trochäische Muster anwenden müssen, nämlich dann, wenn ein Wort der Form stark – schwach von einer weiteren schwachen Silbe gefolgt wird. Aus den Wörtern *guitar is* (schwach – stark – schwach) wird beispielsweise *taris* (stark – schwach) als Einheit extrahiert. Genau diesen Segmentierungsfehler begehen zehneinhalb Monate alte Sprachlerner nicht mehr. Im Gegensatz zu ihren jüngeren Mitstreitern machen sie auch nicht mehr den Fehler, die starke Silbe aus einer jambischen Konstruktion herauszulösen, selbst wenn sie mit genau dieser starken Silbe konditioniert wurden.

Mattys et al. (1999) untersuchen weitere Überlagerungseffekte mit anderen Segmentierungsstrategien. Sie möchten wissen, wie phonotaktische mit prosodischen Einflüssen zusammenarbeiten. Das Ergebnis dieser Untersuchung zeigt, dass die Prosodie gegenüber

3.2 Wortsegmentierung ohne anfängliches Lexikon

der Phonotaktik Vorrang genießt. Da man die Wahrscheinlichkeit von phonotaktischen Mustern gut kontrollieren und dementsprechende Lautverbindungen am Wortanfang, -ende oder -mitte mit prosodischen Mustern kombinieren kann, ist man in der Lage, Aussagen über deren gegenseitige Beeinflussung zu treffen. Dabei lassen sich zwei weitere Ergebnisse festhalten. Natürlich besitzen neun Monate alte Babys eine ausgeprägte Sensitivität hinsichtlich phonotaktischer Muster und deren Zusammenfallen mit Wortgrenzen (siehe auch Kapitel 3.2.3). Interessanterweise beeinflusst die Prosodie jedoch die Wortgrenze. Sie wird mit einem bestimmten prosodischen Muster verbunden. Eine Änderung des Betonungsmusters hat unweigerlich auch eine geänderte Präferenz für bestimmte phonotaktische Kombinationen (*cluster*) zur Folge.

Ramus (2002) stellt eine Studie vor, die eine neue Experimentiermethodik (Ramus und Mehler, 1999) verwendet, welche sich zum gezielten Nachweis einzelner Komponenten der Rhythmik besser eignet, weil Überlagerungseffekte mit Intonationen durch Resynthese kontrolliert werden können. Ramus (2002) äußert sich positiv zur Lösung des *Bootstrapping*-Problems mittels einer phonologisch-prosodischen Ansatzes. Er möchte genauer wissen welche Qualität die Abbildung zwischen rhythmischer Klasse und Segmentierungsstrategie hat (vgl. Ramus, 2002, 88). Sicherlich nicht unbeeinflusst von den Arbeiten Werkers und Tees, verwendet Ramus auch unterschiedliche Sprachen (Japanisch und Niederländisch), um die Betonungsmuster kontrollieren zu können. Den gewünschten Segmentierungseffekt gibt es aber für den Rhythmus²⁸ nicht isoliert, sondern nur in Verbindung mit weiterführenden phonotaktischen Informationen. Wie er selbst ausführt (Ramus, 2002, 110), kann dies während des gegebenen Alters aber nicht stimmen und so wird als Grund die noch neue Methodik angeführt, die für das frühe Alter wohl eher ungeeignet ist.

In Johnson et al. (2003) werden fünf Experimente vorgestellt, welche die PWC-Strategie²⁹ auch für einjährige Kleinkinder belegen sollen. Dazu wird gemessen, ob zwölf Monate alte Versuchspersonen einen freistehenden Konsonanten als segmentierte Einheit akzeptieren würden oder nicht. Es zeigte sich auch hier, dass die Kleinkinder die trainierte Silbe (z.B. *rush*) in den einsilbigen Testwörtern (z.B. *prush*) schlechter wahrnehmen als in den zweisilbigen Einheiten (z.B. *niprush*). Allerdings ist dieser Effekt positionsabhängig. Tritt die Zielsilbe am Wortanfang auf, fällt es den Versuchspersonen in beiden Umgebungen signifikant schwerer die Zielsilbe zu erkennen. Werden nun keine Kunstwörter mehr verwendet (z.B. *win* in *wind* versus *window*) und diese werden in einen Satzkontext eingebunden, lässt sich dieser markante Unterschied zwischen Wortanfang und Wortende nicht mehr messen. Wie das letzte Experiment dann beweisen konnte, gilt dies auch, wenn die umgebende Satzkonstruktion wieder weggelassen wird. Entscheidend ist also die Bekanntheit des Wortes.

²⁸Unter dem Begriff Rhythmus legt der Autor offenbar eine engere Definition zu Grunde als die zu Beginn gegebene. Er sieht die Intonation (Satzmelodie) nicht als Bestandteil des Sprachrhythmus.

²⁹*Possible Word Constraint* (siehe Fußnote 24 auf Seite 57 und Norris et al. (1997) und Norris et al. (2001)).

3.2.2 Allophonie

Jusczyk et al. (1999b) gibt fünf Bedingungen an, die gelten müssen, wenn Allophone zur Segmentierung von wortähnlichen Einheiten genutzt werden sollen. Weil diese zum Verständnis der Gesamtproblematik sehr dienlich sind, sollen sie hier in voller Länge angeführt werden.

First, there must be some indication that allophones are orderly manifestations of phonemic contrasts in the language, rather than simply some random acoustic variation produced by speakers. Second, the distribution of these allophones in fluent speech must correlate with word boundaries. Third, infants must be able to discriminate one allophone from another. Fourth, infants must be sensitive to the systematic distribution of these distinct allophones within native language words. Fifth, they must use these allophonic cues as markers of word boundaries during on-line speech processing. (Jusczyk et al., 1999b, 1466)

In Ladefoged (1975, 88ff) findet sich eine Zusammenfassung, die belegt, dass die erste Bedingung für das Englische gilt. Ebenso kann Church (1987) aufzeigen, dass die zweite Bedingung erfüllt ist. Er beschreibt einen einfachen Parser, der den Sprachinput mit allophonischen und phonotaktischen Input kodiert und daraus eine silbenartige Struktur segmentiert. Diese wiederum verwendet er, um die lexikalische Erschließung (*lexical retrieval*) modellhaft zu beschreiben. Der Erklärungsgehalt dieser Arbeit zielt nicht direkt auf die Segmentierungsprozesse bei Säuglingen ab, da er ein Lexikon annimmt – passt also schon eher zu den Beschreibungen in Kapitel 3.1 – aber die Beobachtungen der Allophonie im Englischen können durchaus ein Vorbild für nachfolgende Experimente gewesen sein. So legt er zum Beispiel dar, dass der aspirierte Verschlusslaut [t^h] im Englischen nur am Wortanfang beziehungsweise Silbenanfang vorkommt (wie in *time*). Das hat zur Folge, dass bei diesen Wörtern der Wortanfang vorhersagbar wäre, wenn dieses Muster einmal als Regel gelernt wurde. /h/ tritt ebenso nur am Silbenanfang auf (z.B. *hit*) und die glotalisierten Verschlusslaute /ʔ/ am Silbenende (z.B. *hit* und *it*).

Hinsichtlich der dritten, anfangs genannten, Bedingung beschreiben Hohne und Jusczyk (1994) ein Experiment, in dem sie zwei Monate alte Babys mit *niterate* und *nikerate* eingebettet in dem Satz *Teddy said . . . to me* habituierten. In der *Postshift*-Phase spielten sie dann *night rate* beziehungsweise *nike rate* ein. In einem zweiten Experiment wurde mittels *Cross Splicing* die interessierenden Übergänge (*t-r*) zwischen *night rate* und *nirate* ausgetauscht. Die Ergebnisse zeigen, dass zwei Monate alte Babys in der Lage sind, diese rein allophonischen Unterschiede zu erkennen. Der *t-r*-Übergang zwischen beiden Wortpaaren besteht darin, dass das [t] bei *night* weder aspiriert noch retroflexiv ist. Das folgende [r] in *rates* ist stimmhaft. All das trifft bei *nirate* nicht zu. Das [r] ist stimmlos und das [t] davor behaucht und retroflexiv gebildet. Damit konnte nachgewiesen werden, dass Kleinkinder nicht nur die rein phonetischen Segmente unterscheiden können, sondern auch sehr feine Variationen innerhalb der segmentalen Kategorien.

Die Ergebnisse sind allerdings wenig überraschend, berücksichtigt man die Arbeiten von Werker und Kollegen (Werker und Tees, 1984). Ab dem zehnten Lebensmonat beginnen die Sprachlerner feste kategorische Entscheidungen für Sprachlaute vorzunehmen. Die allophonischen Unterschiede von [t] und [r] mögen für den erwachsenen Sprecher

3.2 Wortsegmentierung ohne anfängliches Lexikon

des Englischen gering sein und man ist deshalb schnell geneigt, diese Fähigkeit zu bewundern. Für das Kleinkind dürften diese Unterschiede jedoch kaum bemerkenswerter sein als den Stimmtoneinsatz von 20 Milisekunden richtig wahrzunehmen.

Die Fähigkeit der reinen Erkennung von unterschiedlichen Lauteigenschaften sagt allerdings noch nichts darüber aus, ob das Kleinkind den Unterschied im Laut auch mit einer weiteren Eigenschaft – z.B. Ende und Anfang einer Wortes – verbinden kann (fünfte Bedingung). Das letzte Experiment dafür steht nach wie vor aus. Allerdings haben Jusczyk et al. (1999b) eine Reihe von Experimenten durchgeführt, welche die Erkennung der allophonischen Verteilung innerhalb von Silben beziehungsweise Wörtern belegen soll (vierte Bedingung). Eine positive Evidenz kann zwar nicht beweisen, dass die allophonische Verteilung für die Wortsegmentierung genutzt wird; aber sie gibt einen deutlichen Hinweis für die Möglichkeit einer solchen Fähigkeit. Die Autoren bemerken dazu selbst

A demonstration of this sensitivity would be consistent with, although it would not prove, the possibility that they [neun Monate alte Babys] also . . . use allophonic cues to segment words from fluent speech (Jusczyk et al., 1999b, 1467)

Ein negativer Befund bedeutet andererseits wohl aber, dass Allophone bei neun Monate alten Babys nicht für die Segmentierung zur Verfügung stehen. In drei Experimenten wurden neun Monate alte Babys einer *Head Turn Preference Procedure* unterzogen. In der Sensibilisierungsphase hörten sie entweder *night rates* oder *nitrates*; in einer zweiten Gruppe hörten andere Probanden *hamlet* oder *doctor*. Im Test traten diese Begriffe in längeren Texten auf. Es zeigte sich dann, dass zwischen den allophonischen Varianten kein Unterschied zu messen war, schon aber bei *hamlet* und *doctor*. Im zweiten Experiment wurde das Stimulusmaterial zu einsilbigen Wörtern vereinfacht (*night* und *dock*). Das Ergebnis änderte sich jedoch nicht. *Night* wurde in *nitrates* wie in *night rates* gleichermaßen bemerkt. Der Unterschied zwischen *dock* und *doctor* sollte sich abermals als leicht signifikant herausstellen. Im dritten Experiment musste schließlich noch die Möglichkeit überprüft werden, ob die Babys bei *nitrates* von zwei Wörtern ausgehen. Also wurde *rates* durch andere Wörter ersetzt und zusammen mit *night* eingespielt. Nun spiegelten die Hörzeiten, das gleiche Verhalten wider wie bei den *dock* und *doctor* Passagen, was eindeutig belegt, dass die Säuglinge *nitrates* als zwei Wörter ansehen, wenn sie zuvor mit *night* sensibilisiert wurden. Mit diesem letzten Experiment konnte die Sensibilität der Säuglinge hinsichtlich von verteilungsmäßigen Strukturen in der Sprache abermals belegt werden, jedoch nicht für allophonische Regelmäßigkeiten. Das vierte und letzte Experiment replizierte die Anordnung des ersten Experiments, allerdings mit zehneinhalb Monate alten Babys als Versuchspersonen. Hier fand man einen sehr deutlichen Unterschied zwischen den Zeiten von *nitrates* und *night rates*.

Zusammenfassend muss man also sagen, dass zehneinhalb Monate alte Babys die allophonische Verteilung innerhalb von Wörtern erkennen und diese möglicherweise auch zur Wortsegmentierung nutzen können. Für neun Monate alte Babys ist dies nicht der Fall. Sie kennen weder die allophonische Verteilung, noch können sie dieses Wissen dann nutzen, um Wortgrenzen zu markieren.

3.2.3 Phonotaktik

Die Phonotaktik einer Sprache bestimmt die zulässigen Kombinationen von Lauten beziehungsweise Phonemen und deren Position im Wort. Diese Erkennungsleistung bildet sich ebenfalls schon sehr frühzeitig heraus. Um erstens den Zeitpunkt und zweitens die genauen Erkennungsfähigkeiten zu bestimmen, führten Friederici und Wessels (1993) eine Reihe von Experimenten durch. Im ersten Experiment konnte das kritische Alter ermittelt werden. Wortlisten mit richtigen und falschen phonotaktischen Mustern am Wortanfang und -ende konnten von neun Monate alten Babys, aber nicht von viereinhalb oder sechs Monate alten Babys erkannt werden. In den folgenden drei Experimenten manipulierten die Wissenschaftler das Inputmaterial. Die Wortliste wurde in einen Sprachstrom eingefügt. Hier konnte keine Erkennungsleistung nachgewiesen werden. Auch die Vereinfachung der direkten Lautumgebung auf zwei gleiche einsilbige Kunstwörter (Experiment 3) oder die Beschränkung des Stimulus auf das Wortende in kindgerichteter Sprache (Experiment 4) boten keinen Anhaltspunkt. Erst die Verkürzung der Zeit zwischen den einzelnen Stimuli (Triplets z.B. in der Form /*miɡ* NT*it miɡ*/) zeigte einen statistisch relevanten Anstieg in der Hörzeit.³⁰ Im siebten Experiment konnte letztlich der Einfluss prosodischer Information durch Filterung tiefer Frequenzen³¹ ausgeschlossen werden. Die Kleinkinder ließen keinen Unterschied zwischen den unterschiedlichen Stimuli erkennen.

Eine interessante Möglichkeit diese Ergebnisse weiter zu elaborieren, ist, prosodisch ähnliche und unähnliche Sprachen zu verwenden, welche sich jedoch phonotaktisch nicht wesentlich voneinander unterscheiden. Jusczyk et al. (1993b) benutzten Wortlisten des Englischen, Niederländischen und Norwegischen. Zunächst konnte man einfach feststellen, dass neun Monate alte amerikanische Babys der Liste mit englischen Wörtern länger zuhörten als der Liste mit niederländischen Einträgen beziehungsweise, wie bei der *Head Turn Preference Procedure* üblicherweise gemessen, die Zeit des Zusehens signifikant länger war.³² Ein Wiedererkennungseffekt liegt also vor und muss entweder auf die phonotaktischen, phonetischen oder prosodischen Eigenschaften des Stimulus zurückzuführen sein. Letzteres kann man ausschließen, weil bei Verwendung von tiefpassgefiltertem Material keine Präferenz der Säuglinge auszumachen ist (Experiment 2).³³ Um die phonetischen Eigenschaften zu kontrollieren bietet es sich an, das Stimulusmaterial zu ändern, d.h. nur die Lautschnittmenge des Niederländischen und Englischen als Grundlage für die Wortlisten zu nehmen, die phonotaktischen Unterschiede jedoch beizubehalten. Diese Wortlisten werden dann sowohl amerikanischen als auch niederländischen Probanden

³⁰Das sechste Experiment replizierte das erste Experiment mit den verkürzten Zeiten allerdings nur für neun Monate alte Subjekte.

³¹Hohe Frequenzen sind dann im Lautstrom nicht mehr vorhanden. Dadurch bleiben nur die Prosodien erkennbar (Jusczyk et al., 1993a), die phonetischen Informationen gehen jedoch verloren (Jusczyk et al., 1993b).

³²Dieses Verhalten zeigen auch nur neun Monate aber nicht sechs Monate alte Babys (Experiment 3).

³³Wie bereits beim Vorgängerexperiment beschrieben, führt ein Abschneiden der niedrigen Frequenzbereiche (unter 400 Hertz) zu Unkenntlichkeit der phonetischen Information, während die prosodische (Intonation und Rhythmus) erhalten bleibt.

3.2 Wortsegmentierung ohne anfängliches Lexikon

vorgesetzt. In diesem vierten Experiment zeigte sich, dass die Babys ihrer jeweiligen Muttersprache länger zuhörten. Auch hier wird der prosodische Einfluss mit *Low-pass*-Filterung und das Alter kontrolliert.³⁴ Interessanterweise lassen sich diese Ergebnisse mit einer Liste des Norwegischen nicht replizieren. Das Norwegische unterscheidet sich prosodisch derart vom Englischen, dass schon sechs Monate alte Sprachlerner die zwei Listen unterscheiden können, also die Prosodien als Anhaltspunkt nehmen. Dieses letzte Ergebnis lässt sich gut mit anderen Forschungen (z.B. Best et al., 1988) in Einklang bringen. Die anderen Befunde geben ihrerseits einen sehr überzeugenden Beweis für die Fähigkeit von neun Monate alten Babys, phonotaktische Informationen zu erkennen.

Juszyk et al. (1994) bestätigten noch einmal die von Friederici und Wessels (1993) und Juszyk et al. (1993b) ermittelte Altersgrenze. Da sich phonotaktische Muster aus der Häufigkeit zusammenhängender Phonemkombinationen ableiten lassen, präsentierten sie neun und sechs Monate alten Babys Listen mit Phonemketten, die mit hoher Wahrscheinlichkeit innerhalb eines Wortes auftreten würden und solche Phonemketten, die eher unwahrscheinlich in einem englischen Wort anzutreffen wären. Neun Monate alte (aber nicht sechs Monate alte) Babys hörten deutlich länger der Liste mit wahrscheinlichen Phonemketten zu. In einem letzten Versuch konnten sie den Einfluss der Vokale ausschließen, indem sie beide Listen mit den gleichen Vokalen versahen und nur die konsonantische Umgebung manipulierten.³⁵

Die Einbeziehung phonotaktischer Information zur Wortsegmentierung ist in zahlreichen Simulationen dargelegt worden. Eine der frühen modelltheoretischen Umsetzungen ist der zu Anfang des Kapitels 3.2.2 benannte Parser von Church (1987), der auch phonotaktische Informationen des Englischen zur Silbifizierung heranzieht. Church (1987, 58) argumentiert dort, dass Konsonatencluster mit mehr als fünf Komponenten nicht auftreten. Die Einbeziehung der phonotaktischen Informationen macht es dann möglich, die gewünschten Silben zu bestimmen.

Brent und Cartwright (1996) präsentieren eine Reihe von Simulationen, in denen sie verschiedene Einschränkungen des Inputmaterials vornehmen und den Output der Segmentierung messen. Dabei sehen sie die Segmentierung eines Sprachstroms als eine Optimierungsaufgabe an, die durch eine Funktion beschrieben wird.³⁶ Zunächst zeigten die Forscher, dass ein Text, welcher mittels dieser Funktion segmentiert wurde, höhere Präzisions- und Vollständigkeitswerte aufweist³⁷ als ein Text, bei dem die vorgegebene Anzahl der Lücken zufällig eingefügt wurde.³⁸ Dann testeten sie die Verbesserung bei Zunahme der Vokal- (*vowel constraint*: Jedes Wort besitzt mindestens einen Vokal) und

³⁴Allerdings ist der Effekt bei englischen Babys sehr viel ausgeprägter, was sich durch anglophone Einflüsse in den niederländischen Haushalten erklären lässt.

³⁵Kuhl (1991); Kuhl et al. (1992) und Polka und Werker (1994) haben herausgefunden, dass Vokale sehr viel früher erkannt werden können als Konsonanten.

³⁶Diese Funktion erhält man durch Anwendung der *Minimum Representation Length* (MRL)-Methode (Rissanen, 1989); siehe auch 3.2.4

³⁷Brent und Cartwright (1996) verwenden die Begriffe *Accuracy* und *Completeness*; welche aber nach den Definitionen (S.109-110) äquivalent zu den gebräuchlicheren Begriffen *Precision* and *Recall* sind.

³⁸Es handelt sich hierbei nicht um eine reine Zufallsauswahl der Lücken, da die korrekte Anzahl der Lücken bereits vorgegeben wurde. Der reine Zufall müsste hier ein schlechteres Ergebnis liefern.

3 Theorien der Wortsegmentierung

der Wortgrenzenbedingung (*boundary cluster constraint*: Jedes Wort beginnt und endet mit zulässigen Phonemkombinationen). Wiederum zeigte sich eine starke Verbesserung in beiden Messdimensionen. Die dritte Simulation testete schließlich die Phonotaktik explizit. Die Phonotaktik ist im Prinzip bereits Teil der Optimierungsfunktion und lässt sich im Grunde schon in der Wortgrenzenbeschränkung entdecken, wenn auch in beiden Fällen implizit. Mithin ist es wichtig, diesen Einfluss separat zu messen.³⁹ Es stellte sich heraus, dass beide Texte, der Kontrolltext sowie der manipulierte Input, ungefähr die gleichen Zuwächse verzeichneten. Daraus lässt sich schließen, dass die phonotaktischen Informationen vom Algorithmus unabhängig sind; im Kontrolltext also die gleichen prozentualen Verbesserungen hervorrufen.⁴⁰

Mattys und Jusczyk (2001) zeigen, dass neun Monate alte Versuchspersonen in der Lage sind, die Phonotaktik zur Segmentierung von Worteinheiten aus dem Lautstrom zu nutzen. Die Probanden hörten einer Liste von CVC-Wörtern länger zu, wenn diese vorher in einem Satzkontext mit günstigen phonotaktischen Umgebungen vorkamen als wenn die Phonotaktik einem eher unwahrscheinlichen Muster folgte. Zudem kann das Erkennen von Wortanfang oder -ende ebenso durch phonotaktische Umgebungen konditioniert werden.

Saffran und Thiessen (2003) legen dar, dass neun Monate alte Babys auf verschiedene Silbennuster konditioniert werden können.⁴¹ Allerdings kann das Experiment nicht klären, ob der Erkennungseffekt auch auf die Dauer der Silben zurückzuführen sei. Schließlich unterscheiden sich beispielsweise CV und CCCVCCC Cluster auch hinsichtlich ihrer zeitlichen Länge. Die Probanden sind ebenso in der Lage, bei zweisilbigen Wörtern einfache Muster der Stimmhaftigkeit zu erkennen.⁴² Werden diese Muster jedoch auf weitere phonologische Zusammenhänge ausgedehnt, kann weder ein Erkennungs- noch ein Neuigkeitseffekt gemessen werden. Allerdings sind die präsentierten Muster wohl auch nicht in natürlichen Sprachen anzutreffen, was darauf hinweist, dass die Versuchspersonen diese

³⁹Um hier eine Vorstellung von der Entwicklungsrichtung zu bekommen, sollten zwei weitere Experimente von Luce und Large (2001) bedacht werden. Eigentümlicherweise nehmen die Autoren keinen Bezug zu Brent und Cartwright (1996), obwohl dies bei der Nähe des Themas doch einleuchten würde. Im ersten Experiment können Luce und Large die Vorhersagen verschiedener Simulationen (Shortlist, TRACE, ARTPHONE; PARSYN) für erwachsene Sprecher bestätigen, nämlich dass eine hohe phonotaktische Wahrscheinlichkeit die Wortverarbeitung gesprochener Sprache erleichtert, ein erhöhter Wettbewerb unter lexikalischen Einträgen jedoch verlangsamend wirkt. Letzteres wurde durch sehr ähnlich klingende Wörter erreicht (*similarity neighbourhood density*-Bedingung). Wird das Experiment mit Nichtwörtern durchgeführt ergibt sich überraschenderweise ein anderes Ergebnis. Hier lassen sich sowohl synergetische als auch neutrale Effekte nachweisen. Für Kinder zwischen zwei bis drei Jahren siehe Coady und Aslin (2004).

⁴⁰Der Vollständigkeit wegen sei hier kurz noch die vierte Simulation angeführt. Sie testete, ob die Liste mit Phonemkombinationen, welche aus den Wortenden und Anfängen der ausgewählten Korpora das gleiche Ergebnis liefern würde, wie eine komplette Liste von erlaubten Phonemkombinationen am Wortanfang und -ende. Das Experiment konnte keinen Hinweis für einen signifikanten Unterschied belegen.

⁴¹Dabei können die verschiedenen Konsonanten und Vokalkombinationen als phonotaktische Muster einer Sprache angesehen werden.

⁴²Anfangskonsonant stimmlos, Auslaut stimmhaft und umgekehrt.

3.2 Wortsegmentierung ohne anfängliches Lexikon

Fähigkeit für komplexere Muster nicht benötigen. Die Experimente geben damit weiter Aufschluss über die Fähigkeit, einfache phonotaktische Zusammenhänge in natürlichen Sprachen zu erkennen.

Als Abschluss kann ein Zitat von Friederici und Wessels (1993, 294) sehr treffend eine Zusammenfassung geben.

Whether the observed sensitivity to phonotactic word-boundary constraints in 9-month-old infants can be taken to reflect fully developed phonotactic knowledge about possible words in their native language, or whether it must be characterized as a sensitivity to frequency of occurrence of certain syllables in certain positions cannot be decided on the basis of the data at hand.

Wie auch McQueen (1998) ausführt, gilt dies ebenso für erwachsene Sprecher. Die phonotaktischen Informationen bieten zweifellos eine wesentliche Unterstützung für die Segmentierung. Allerdings ist die Rolle der Phonotaktika eher beschränkt:

... different cues will provide different degrees of assistance in segmentation. Phonotactics may have a relatively limited role to play in segmentation. The sequence constraints between segments are not powerful enough to mark all word boundaries. (McQueen, 1998: 41)

Er führt weiter aus, dass 37% der Wortgrenzen durch sogenannte Randbedingungen der Abfolge der Segmente (*sequence constraints*) entdeckt werden können. Definiert man die einzelnen Segmente jedoch ungenau, sinkt die Erfolgsquote dabei auf ein Prozent. Eine vollständige Segmentierung ist mit den Phonotaktika zwar nicht möglich, ein wesentlicher Beitrag zur Unterstützung der Segmentierung ist aber auf jeden Fall gegeben.

3.2.4 Übergangswahrscheinlichkeiten und regelmäßige Verteilungen

Nach Brent und Cartwright (1996) lassen sich grundsätzlich zwei Ansätze zur Segmentierung eines Lautstroms unterscheiden. Die *Subsequence frequency strategies* basieren entweder auf Wörtern (*word-based*)⁴³ oder auf Wortgrenzen (*boundary-based*). In einem späteren Aufsatz unterteilt Brent (1999b) die Ansätze zur Wortgrenzensegmentierung noch einmal weiter in *utterance-boundary* und *predictability* Strategien. Diese Unterteilung dankt ihre Existenz einem neuen Forschungsansatz zur Bestimmung von Wortgrenzen, der sich erst 1996 langsam etablierte (Aslin et al. (1996); Christiansen und Seidenberg (1998)). Diese *Utterance-boundary*-Strategien nutzen die Eigenschaften von ganz bestimmten Phonemkonstruktionen, die nur am Wortende auftreten. Kennt man diese Phonemkonstruktionen, kann man das Ende des betreffenden Wortes und damit auch den Anfang des folgenden Wortes bestimmen.

Die ersten *Predictability*-Strategien fußen auf den Arbeiten von Harris (1954) und Harris (1955). So wie auch die späteren Arbeiten von Saffran et al. (1996b) wurden bei Harris die distributionalen Eigenschaften der Phoneme in einem Lautstrom genutzt, um Vorhersagen über die nachfolgenden Laute zu geben. Elman (1990) und später auch Cairns

⁴³ *BootLex* (Batchelder, 2002); *INCDROP* (Brent (1999b); Brent (1997)); Olivier (1968) zitiert in Brent und Cartwright (1996); Brent (1999a); Batchelder (2002); de Marcken (1996) und Goldsmith (2001).

3 Theorien der Wortsegmentierung

et al. (1997) haben anhand eines neuronalen Netzwerkes die Vorhersage des Folgephonomens erfolgreich simuliert. Je größer die Abweichung vom vorherberechneten Phonem war, je höher wurde hier die Wahrscheinlichkeit einer Wortgrenze definiert. Saffran et al. (1996b) entwickelten Harris' Arbeit insofern weiter als sie nun die Übergangswahrscheinlichkeiten längerer Phonemketten (Silben) zueinander untersuchten. Motiviert durch Experimente an acht Monate alten Babies (Saffran et al., 1996a, siehe unten), konnte der Einfluss von statistischen Verteilungen bei der Wortsegmentierung nachgewiesen werden. Dieser experimentellen Beobachtung liegen die Überlegungen einer Vorhersagestrategie zu Grunde. Sie besagt hier, dass die Phonemkombinationen im Wort eine andere Wahrscheinlichkeit haben als Phonemketten, die zwischen den Wörtern auftreten. So folgt mit einer bestimmten Wahrscheinlichkeit auf eine Silbe eine nachfolgende Silbe. Im Wort und beim Wortübergang sind die Lautumgebungen jedoch verschieden und somit ändern sich auch die Wahrscheinlichkeiten der Silbenfolgen. Berechnet man die Übergangswahrscheinlichkeiten der einzelnen Silben in einem gegebenen Lautstrom, treten bestimmte Lautkombinationen zwischen Wörtern sehr selten oder nie auf. Dieser Umstand kann zur Segmentierung von Wörtern genutzt werden. Leider bleiben die Überlegungen von (Saffran et al., 1996a) auf wenige Beispiele beschränkt.

Das zentrale entwicklungspsychologische Experiment für die tatsächliche Nutzung von Übergangswahrscheinlichkeiten von Lauten bei Babys gelang Saffran et al. (1996a) mittels einer modifizierten Form der *Head-Turn-Preference-Procedure* (Jusczyk und Aslin, 1995). Acht Monate alte Säuglinge wurden zwei Minuten lang mit einem Sprachstrom konditioniert, der CV-Silben in einer bestimmten Anordnung enthielt. Der Synthesizer enthielt keine weiteren akustischen oder prosodischen Anhaltspunkte, sodass die Versuchspersonen lediglich durch die bedingten Wahrscheinlichkeiten zwischen den Silben, ein Muster im Sprachstrom erkennen konnten. Silben, die zu einer Einheit („Wort“) gehörten, folgten einander mit einer hundertprozentigen Vorhersagbarkeit, während die Silben von verschiedenen Einheiten („Wörtern“), eine Vorhersagbarkeit von nur einem Drittel aufwiesen. In der Testphase hörten die Säuglinge schließlich Anordnungen der gleichen Silben, die aber nicht die gewohnten Übergangswahrscheinlichkeiten aufwiesen. Da die Versuchspersonen deutlich länger dem modifizierten Lautstrom zuhörten, war letztlich der Beweis erbracht, dass die Babys die Änderung der Übergangswahrscheinlichkeiten eindeutig erkannt haben müssen.

Diese Ergebnisse sind weitestgehend akzeptiert. Kritik wurde lediglich an den typischen methodischen Randbedingungen dieser Art von Experimenten geäußert, wie die Festsetzung der Übergangswahrscheinlichkeit auf 0.3 und 1.0, die von der Realität eines realen Sprachstroms zu stark abstrahiert (Monahan (2006); Gambell und Yang (2005, 24)),⁴⁴ oder die künstliche Darbietung des Sprachstroms (Johnson und Jusczyk, 2001). Beachtenswerter ist dagegen abermals die unterschiedliche Interpretation des Experiments und dessen Argumentation hinsichtlich der Nativismus-Empirismus-Debatte (Pe-

⁴⁴Aslin et al. (1998) zeigten aber, dass auch bei Verwendung unterschiedlicher Übergangswahrscheinlichkeiten die Ergebnisse wie bei Saffran et al. (1996a) korrekt sind. Zudem konnten die reinen *co-occurrences* der Silben als intervenierende Variable ausgeschlossen werden.

3.2 Wortsegmentierung ohne anfängliches Lexikon

setsky et al. (1997); Pinker (1997); Jenkins und Maxam (1997); Clark et al. (1997); Elman (1997); Saffran et al. (1997a); Bates und Elman (1996).⁴⁵

Aslin et al. (1998, 322) gingen sodann dazu über „...to identify the particular type of statistics infants utilize to segment words from fluent speech.“ Es wäre nämlich auch möglich, dass die Babys die absolute Häufigkeit der Silben heranziehen und auf dieser Grundlage Segmentierungen vornehmen. Um genau das auszuschließen, manipulierten Aslin et al. (1998) einen weiteren Sprachstrom dergestalt, dass sie im Trainingsmaterial zwei Wörter doppelt so oft einfügten wie die anderen zwei Wörter (*part-words*). Dadurch änderte sich natürlich die absolute Häufigkeit dieser Silben innerhalb der Wörter, nicht jedoch deren Übergangswahrscheinlichkeit, welche innerhalb der Wörter weiterhin 1.0 betrug. In der Testphase wurden dann nur die *part-words* verwendet; die absolute Häufigkeit war nun gleich und die Übergangswahrscheinlichkeiten betrugen 0.5 und 1.0. Da die Versuchspersonen den neuen *part-words* bedeutend länger zuhörten, konnte daraus ein Erkennungseffekt der Übergangswahrscheinlichkeiten geschlossen werden.

Eine alternative Strategie konnte ein weiteres Experiment beweisen. Babys des gleichen Alters sind bereits in der Lage, einfache Muster im Sprachstrom zu erkennen und diese zu abstrahieren. Sie nutzen spezifische Informationen für eine allgemeine Schlussfolgerung – also eine Regelableitung. Marcus et al. (1999) zeigten mit der gleichen experimentellen Methode wie Saffran et al. (1996a), dass Säuglinge Muster der Form ABA und ABB aus einem Sprachstrom extrahieren können, ohne auf eine statistische Verteilung zurückzugreifen. Um eventuelle statistische Berechnungen gänzlich auszuschließen, verwendeten Marcus et al. (1999) unterschiedliche Stimuli in der Habituations- und Testphase. Mit zwei Folgeexperimenten konnte schließlich noch die Möglichkeit eliminiert werden, dass eventuell die Stimmhaftigkeit der präsentierten Silben für die Erkennung des Musters verantwortlich sei. Da man derartige Interventionen ausschließen kann, müssen die Kleinkinder abstrakte Regeln extrahiert haben. So muss der Sprachlerner in der Lage sein, zumindest zwei Strategien der Mustereextraktion zu verwenden: eine statistische und eine algebraische. Analog zum Experiment von Saffran et al. (1996a) meldeten sich die üblichen Verdächtigen schnell zu Wort und ordneten die Ergebnisse geschickt ihren Philosophien zu (Pinker (1999); McClelland und Plaut (1999); Seidenberg und Elman (1999); Christiansen und Curtin (1999); Negishi (1999); Eimas (1999); Altmann und Dienes (1999); Marcus (1999a); Marcus (1999b); Marcus (1999c); Marcus (1999d)).

⁴⁵Saffran et al. (1997a) versuchen einen möglichen Konsens in der Diskussion herbeizuführen. Sie sprechen den neuronalen Netzwerken einen großen Wissensgewinn zu und akzeptieren gleichzeitig in begrenzter Weise den Gedanken der *Innateness*:

Linguistic structure cannot be learned through indirect analyses of input sentences, no matter how complex or numerous these analyses may be. Such analyses must in some fashion be focused or oriented by innate predispositions of the learner; otherwise, there is no way to explain why human languages or why language recurrently develop certain types of structures ... Innate biases in statistical learning may be different in important ways from innate knowledge of linguistic principles. But both of these implementations involve types of innateness. (Saffran et al., 1997a, 1276)

3 Theorien der Wortsegmentierung

In diesem Zusammenhang sollte sicherlich auch Wolff (1975b) und Wolff (1977) Erwähnung finden. Mit Computersimulationen für eine künstliche (Wolff, 1975b) und eine natürliche Sprache (Wolff, 1977), gelang ihm eine erste und recht überzeugende Präsentation von Daten über die Herauslösung von bedeutungstragenden Einheiten sowie ein Entwicklungsverlauf der Wortschatzgröße und Wortlänge. Perruchet und Vinter (1998) entwarfen das PARSE-Modell, welches unter leichter Abänderung der Annahmen der Experimente von Saffran et al.⁴⁶, die wichtigste Eigenschaft der Übergangswahrscheinlichkeitenberechnung in den Mittelpunkt stellt. Die Änderung bezieht sich auf die Repräsentationseinheiten (*percepts*), welche, wenn

... repeatedly perceived, its components are associated and form a new representational unit as an automatic by-product of the joint attentional processing of the components.
... These initial representational units in turn become able to guide perception and to enter as components of other percepts, and this process continues recursively. (Perruchet und Vinter, 1998, 258)

Die Information zur Segmentierung der Wortgrenzen ist aber identisch mit der auch die Übergangswahrscheinlichkeiten berechnet werden. Die Zuweisung von verschiedenen Gewichten zu den Einheiten erfolgt aufgrund ihrer Häufigkeit. Da sich die Einheiten nur in Bezug zu den anderen in ihrer Gewichtung erhöhen oder dekrementieren, entsteht genau die Relation der Häufigkeit von einer Einheit zur Gesamtverteilung der Einheiten. PARSE nutzt damit die gleichen Informationen, welche auch zur herkömmlichen Berechnung der Übergangswahrscheinlichkeiten herangezogen werden müssen. Somit handelt es sich um einen unterschiedlichen Rechenweg, dem aber das gleiche Phänomen zu Grunde liegt. Die Arbeit illustriert jedoch recht anschaulich, welche Prozesse bei der statistischen Wortsegmentierung ebenso ablaufen könnten.

Sicher nicht ohne dieses Wissen, versuchen sich Gambell und Yang⁴⁷ an einer Simulation mit dem Bestreben, die Universalgrammatik eher anzupreisen als generelle kognitive Fähigkeiten dafür verantwortlich zu machen. Sie benutzen die herkömmliche Formel zur Berechnung der Übergangswahrscheinlichkeiten, $P(AB)/P(A)$, und einen Korpus der CHILDES Datenbank (MacWhinney, 1995) mit 226.178 Wörtern. Per Zufall wählten sie Stichproben aus der Datenbank.⁴⁸ Alle Wörter wurden in Silben auf Basis des *Onset*-Prinzips zerlegt.⁴⁹ Da die Ergebnisse der Simulation (Vollständigkeit 0.23 und Präzision 0.42) aus Sicht der Autoren nicht ausreichend erschienen, um eine solide Ausgangsbasis für den Sprachlerner unterstellen zu dürfen, leiteten sie daraus die Existenz einer Universalgrammatik ab, denn wenn die Übergangswahrscheinlichkeiten als einziger Algorithmus ohne Vorkenntnis sprachspezifischen Wissens zu viele Fehler aufweisen,

⁴⁶Saffran et al. (1996a); Saffran et al. (1996b); Saffran et al. (1997a).

⁴⁷Gambell und Yang (2003); Gambell und Yang (2004); Gambell und Yang (2005).

⁴⁸Hier scheint eine erste Kritik angebracht, weil eine Zufallsauswahl alle Korpora mit kindgerichteter Sprache berücksichtigt, also auch jene, welche bereits sprachliche Dialoge und Konstruktionen beinhalten, die untypisch für die anfängliche Phase des Spracherwerbs sind.

⁴⁹Hier muss eine zweite Kritik folgen, weil erstens unklar bleibt durch welchen Mechanismus das offensichtlich erlernte Prinzip angeeignet werden konnte und zweitens, weil das Prinzip auch für durchschnittlich 10% der englischen Wörter falsche Ergebnisse liefern muss (vgl. Diskussion in Kapitel 3.3)

müssen zusätzliche Informationen universalsprachlicher Natur gegeben sein. Schließlich sind sich alle einig, dass das Segmentierungsproblem am Anfang des Spracherwerbs gelöst werden müsse. Der eigentliche Trick ihres Aufsatzes besteht aber darin, dass sie auch unabhängig von den Ergebnissen ihrer Simulation, immer schon die Universalgrammatik als letztendliche Wahrheit hervorheben, weil die Voraussetzungen der Simulation diese eben auch erfordern würden. Genauer fragen sie:

First, How [sig] do infants know the primacy of syllables? . . . Second, where do the syllables come from? . . . Third, syllabification of speech is far from trivial, involving both innate knowledge of phonological structures as well as discovering language-specific phonotactic constraints. Finally, in a realistic learning environment, statistical information is bound to be far less perfect than the constructed in artificial language learning experiments . . . (Gambell und Yang, 2005: 24).

Die ersten drei Aussagen setzen eine Silbifizierung voraus. Mögliche Alternativen hierzu wären der Ansatz von Perruchet und Vinter (1998) oder eine zufällige Auswahl von Segmenten oder segmentähnlichen Gebilden, wie im vierten Kapitel noch ausführlich auszuführen sein wird. Die letzte Aussage des gegebenen Zitats kann getrost durch die Ergebnisse der hier vorgelegten Simulation verneint werden.

3.3 Vereinbarung konträrer Standpunkte in der Wortsegmentierung

Nachdem sich die Forschung zunächst auf die Begründung von möglichen Segmentierungsstrategien beschränkte, konzentriert sie sich in jüngerer Zeit auf die Abfolge und das Ineinanderwirken der einzelnen Mechanismen. Dabei konnten der Anfang der Verwendung und der Einfluss der Phonotaktik und der Allophonie relativ unstrittig festgesetzt werden. Die Mechanismen der Rhythmik und der statistischen Verteilung lösten jedoch eine kleine Kontroverse aus, die sich nach dem jetzigen Stand der Experimente für die Übergangswahrscheinlichkeiten entscheiden dürfte.⁵⁰ Saffran et al. (1996b) zeigten, dass erwachsene Sprachlerner statistische Verteilungen dem Sprachrhythmus vorziehen. In einem weiteren Experiment (Saffran et al., 1996a) gelang ihnen der Nachweis, dass auch acht Monate alte Babys bereits die statistischen Muster eindeutig berechnen können. Johnson und Jusczyk (2001) legten dann aber anschaulich dar, dass die prosodischen Eigenschaften eines Lautstroms Vorrang vor statistischen Verteilungen bei acht Monate alten Kindern genießen. Im Vergleich zum Experiment von Saffran et al. (1996b) würden sich hier die Präferenzen im Zeitverlauf vom Kleinkind- bis ins Erwachsenenalter umkehren. Schließlich erwiderten Thiessen und Saffran (2003), dass sechs bis acht Monate alte Babys die statistischen den prosodischen Informationen vorziehen.

Eine Reihe von Wissenschaftlern bestätigen, dass Prosodien, insbesondere die speziellen Betonungsmuster des Englischen, von neun – aber nicht sieben – Monate alten Kindern zur Wortsegmentierung genutzt werden können (Echols et al., 1997). Morgan

⁵⁰Diese Entwicklung ist keineswegs unstrittig. Oft wird recht unkritisch die metrische Segmentierungsstrategie als der anfängliche Mechanismus unterstellt (vgl. z.B. Siegmüller, 2007, 121).

3 Theorien der Wortsegmentierung

(1996b) und auch Jusczyk et al. (1993a) hat mit sechs und neun Monate alten Babys das gleiche Ergebnis erhalten. Eine sprachunabhängige Studie gibt für neun Monate alte Babys die Erkennung bestimmter rhythmischer Muster im Niederländischen und Englischen an.⁵¹ Das weist daraufhin, dass Babys den Unterschied zwischen jambischen und trochäischen Mustern mit dem neunten Monat gemeistert haben. Gambell und Yang (2005) deuten aber auf ein Zirkelschlussproblem hin, wenn man die prosodischen Eigenschaften als Ausgangspunkt für die anfängliche Lautsegmentierung nimmt:

To use the Metrical Segmentation Strategy, the learner must be able to identify the language-specific stress pattern, for the metrical systems in the world's languages differ considerably. This can only be achieved after the learner has accumulated a sufficient and representative sample of words to begin with – but where do these words come from? There appears to be a chicken and egg problem at hand. (Gambell und Yang, 2005, 4)

Johnson und Jusczyk (2001) schlagen dazu vor, das anfängliche Paralexikon⁵² könnte aus dem Aufnehmen der isoliert gesprochenen Wörter einer Sprache entstehen. Allerdings kann dies bis heute nicht experimentell nachgewiesen werden. Somit folgen die Autoren ansatzweise den Gedanken von Peters (1983) und Pinker (1984). Isolierte Wörter machen in der englischen Mutter-Kind-Sprache nicht mehr als 10% von der Gesamtzahl gesprochener Wörter aus (Brent und Siskind (2001); Aslin et al. (1996); Weijer (1999)); eine Größenordnung die ausreichend erscheint. Allerdings ist diese Hypothese⁵³ nicht gut begründet. Das Problem bleibt unverändert bestehen, denn woher soll der Sprachlerner nur wissen, dass es sich bei einem isolierten Wort nicht um zwei oder drei Wörter handelt, wie es beispielsweise bei *I go* und *ago* oder *I do it* und *idiot* oder *isolate* und *I'm so late* der Fall ist. Die Beispiele zeigen bereits, dass die Wortlänge keinen zuverlässigen Anhaltspunkt geben kann. Gambell und Yang (2005) bringen deshalb die angeborenen Sprachstrukturen als einzig verbleibende Lösung ins Spiel. Die Universalgrammatik sei, so die Autoren, die einzig mögliche Alternative zu dem Dilemma des Zirkelschlusses, weil die Übergangswahrscheinlichkeiten als einziges Modell, bei dem das Zirkelschlussproblem nicht auftritt, zu viele Fehlsegmentierungen verursachen. Letztlich sprechen sie sogar den statistischen Modellen ihre Subsistenz ab:

The fact that statistical learning does not extract words reliably, combined with the fact that simple linguistic constraints and algebraic learning do extract words reliably, raise the possibility that statistical learning may not be used in word segmentation at all. (Gambell und Yang, 2005, 25)

Wie noch zu zeigen sein wird, kann mit einer einfachen sprachunabhängigen und statistisch fundierten Heuristik ein anfängliches Lautlexikon erstellt werden, das nicht den universalgrammatischen Ansatz heranziehen muss. Im Grunde führt der in Kapitel 5 entwickelte Ansatz zwei bestehende Modelle – *distributional regularities* (Brent und Cartwright, 1996) und *transitional probabilities* (Saffran et al. (1996a); Saffran et al. (1996b)) – zusammen. Dabei tritt weder das Zirkelschlussproblem auf, noch widerspricht es den experimentellen Ergebnissen der Psycholinguistik oder Entwicklungspsychologie.

⁵¹vgl. dazu auch Houston et al. (2000)

⁵²Lexikon ist hier im Sinne einer Liste mit der rein lautlichen Repräsentation zu verstehen, d.h. Bedeutungen wurden der lautlichen Struktur noch nicht zugeordnet.

⁵³Isolierte Wörter werden als Ausgangspunkt einer prosodischen Regeldeduktion angenommen.

Das Zirkelschlussproblem ist bei dieser Herangehensweise deshalb nicht von Belang, weil kein anfängliches Lexikon unterstellt wird. Die einzigen Annahmen sind, dass das Kleinkind Häufigkeiten von bestimmten Lautsequenzen in Beziehung zu den Häufigkeiten von anderen Lautsequenzen bestimmen kann. Dies wurde in Experimenten von Saffran et al. (1996a) und Johnson und Jusczyk (2001) gezeigt. Die Ergebnisse sind in der Literatur nicht umstritten und sind auch aus neurobiologischer Perspektive stimmig (Kandel (2006); Spitzer (2000); und Spitzer (2002); Ewert (1998, 94 und 209ff)). Die Fähigkeit einer statistischen Verarbeitung tritt auch bei anderen kognitiven Prozessen auf (Lalonde und Werker (1995); Hasher und Chromiak (1977); Saffran (2002); Saffran (2003d); Aslin et al. (1999); Clark (2004)) und ist auch nicht auf den Menschen beschränkt. (Hauser (2001, 428); Gelman und Cordes (2001)).⁵⁴ Der Ansatz fußt damit ebenso auf dem neurobiologischen Prinzip des Lernens (Breitenstein und Knecht (2003, 139f); vgl. Spitzer (2002)), welches in Kurzform besagt, dass die Häufigkeit eines Reizes zu einer Verstärkung der entsprechenden Synapsen und zusätzlichen dendritischen Dornen führt, damit die Signalübertragung beschleunigt.

Wie ist dieses Modell nun aber mit den Experimenten in Einklang zu bringen?⁵⁵ Erstens muss man die zeitlichen Abfolgen berücksichtigen. Mit Thiessen und Saffran (2003) sollte man annehmen dürfen, dass die Lautstromsegmentierung erst dann einsetzt, wenn das Kind Statistiken bilden kann. Das ist zwischen dem sechsten und achten Monat möglich. Johnson und Jusczyk (2001) legen für die Verwendung der Prosodien eine Zeitspanne zwischen den siebten und neunten Monat an. Damit wäre eine zeitliche Verschiebung von einigen Wochen zur Bildung der Statistiken gegeben. Diese würden den Grundstein für das rhythmische Muster legen und, wenn die prosodische Regel internalisiert wurde, schließlich zwischen dem siebenten und neunten Monat auch die vorherrschende Segmentierungsstrategie sein. Eine Bevorzugung von prosodischen Eigenschaften soll also nicht in Abrede gestellt werden. Diese tritt aber nicht zeitgleich mit der vorrangigen Bildung von statistischen Verteilungen auf.

Erst nach dem neunten Monat beginnt das englischsprachige Kind die rhythmischen Feinheiten seiner Muttersprache zu erlernen. Wie bereits im Kapitel 3.2.1 ausführlich dargelegt wurde, wird das trochäische Muster schon mit neun Monaten recht überzeugend beherrscht. Die jambische Silbenbetonung kann jedoch erst mit zehneinhalb Monaten richtig erkannt werden (Jusczyk et al., 1999a). Eine mögliche Erklärung hierzu wäre, dass zwischen dem neunten und dem zehnten Monat die prosodische Regel weiter verfeinert und aufgrund des erstellten Paralexikons richtig deduziert werden konnte.

Gerade weil das Kind noch nicht in der Lage ist, die *schwach – stark / stark – schwach*-Unterscheidung zu treffen, ist es unwahrscheinlich, dass sich die Sprachlerner auch nach dem achten Monat lediglich auf die Betonungsmuster verlassen, d.h., das erkannte Muster tritt nun nicht häufig genug auf oder gerät in Konflikt.⁵⁶ In diesem Fall würde das Kind

⁵⁴vgl. Massaro (1994) und Tomasello (1995, insbesondere 137f) als zusammenfassende Darstellungen von anderen Quellen.

⁵⁵insbesondere hinsichtlich Johnson und Jusczyk (2001).

⁵⁶Beispielsweise für folgende Kombinationen von starken und schwachen Silben an Wortübergängen:
s-w / s-w / w-s / s-w-s.

3 Theorien der Wortsegmentierung

wieder auf die statistischen Verteilungen zurückgreifen müssen. Hierbei erkennt es bald, dass sich das Muster wiederholt. Unterstellt man lediglich die Rhythmik, so wie sie im Englischen zutreffend ist, als universale Segmentierungsstrategie, könnten beispielsweise französische Kinder wohl nicht ihre Muttersprache lernen. Sie würden die typischen Anhaltspunkte dieser universalen Metrik nicht auf die Wortgrenzen von französischen Wörtern beziehen können. Die betonte starke Silbe steht in der Regel im Französischen nicht am Anfang des Wortes. So wird es aber in der aus dem Englischen abgeleiteten universalen Segmentierungsstrategie gefordert.⁵⁷

3.4 Zusammenfassung

Zuerst wurde ein resümierender Blick auf die Wortsegmentierung im Allgemeinen geworfen, um grundsätzliche Mechanismen der lexikalischen Erschließung kennenzulernen und eine Vorstellung von den vorhandenen Modellen zu erhalten. Zum einen war dies notwendig, um Simulationsentwürfe auszuschließen, welche mit den Theorien der Ersterwerbsforschung unvereinbar sind, und ein geeignetes Modell für die lexikalische Erschließung im Paralexikon der Simulation auszuwählen. Zum anderen kann man aus den Beschreibungen abermals eine Entwicklungsrichtung der frühkindlichen Fähigkeiten ersehen. Die Mehrzahl der Modelle, die eine vorlexikalische Repräsentation unterstellen, zeigen die Vorteile einer segmentalen Verarbeitung.

Das derzeit sehr populäre Kohortenmodell wurde in seiner ursprünglichen Formulierung (Marslen-Wilson und Welsh, 1978) auch für die Simulation des Paralexikonabgleichs für den hiesigen Segmentierungsalgorithmus gewählt. Es bietet den Vorteil einer einfachen Implementation und verzichtet, zumindest in den frühen Versionen, auf Häufigkeitseffekte innerhalb der Kandidatenliste.⁵⁸ Die Fortführung des Kohortenmodells (Marslen-Wilson und Warren, 1994) ist allerdings ungeeignet, weil unklar ist, wie genau die Extraktion der phonetischen Eigenschaften und deren postlexikalische Bewertung vonstattengeht. Die derzeitige algorithmische Beschreibung ist für eine Simulation noch zu unspezifisch und erfordert zu viele subjektive Annahmen. Sicherlich wäre auch das NAM-Modell eine Alternative. Die schon recht komplexen Vorgänge im Netzwerk und die Verwendung von *Top-down*-Mechanismen sprechen jedoch dagegen. Dies trifft auch für das Trace- und Shortlist-Modell zu.

Im zweiten Teil des Kapitels wurden die Segmentierungstheorien ohne anfängliches Lexikon, welche für Erstsprachler relevant sein dürften, repliziert und deren zugrunde liegende Experimente ausführlich beleuchtet. Lässt man die Gesamtheit der experimentellen Befunde Revue passieren, fällt eine gewisse Chronologie bei der Entwicklung der einzelnen Fähigkeiten auf. Wie in Kapitel 3.3 ausgeführt, könnte man grob die Gene-

⁵⁷Ebenso kann man das Koreanische als Beispiel dagegen anführen, da im Koreanischen die Prosodie keine lexikalische Einheit oder ähnliches spezifizieren kann (Kim, 2003).

⁵⁸Die Berücksichtigung der Häufigkeiten und deren positive Auswirkungen sind im BootLex-Algorithmus (Batchelder, 2002) ausgeführt. Eine entsprechende Erweiterung wäre also denkbar, aber für das Erkenntnisziel dieser Arbeit nicht notwendig.

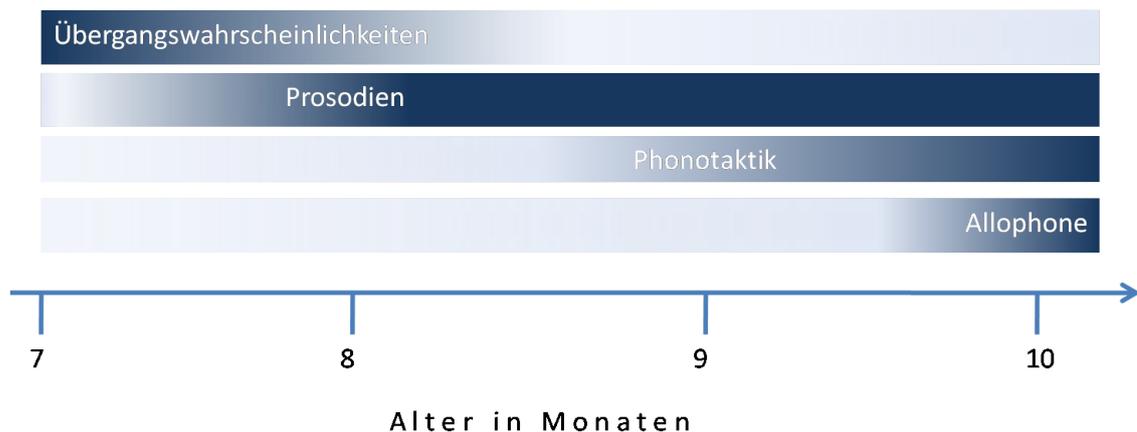


Abbildung 3.3: Vereinfachte Darstellung einer möglichen Abfolge und zeitgleichen Auftretens von Segmentierungsstrategien, Quelle: Eigene Darstellung

se der Segmentierungsfähigkeiten zusammenfassend skizzieren (Abbildung 3.3).⁵⁹ Dazu lässt sich festhalten, dass der Sprachlerner verschiedene Anhaltspunkte für die Segmentierung benutzt, aber wie bereits McQueen (1998) deutlich ausführt, kann keine dieser Anhaltspunkte allein zu einem befriedigenden Ergebnis führen.⁶⁰

Mattys et al. (1999) löst die Frage nach dem Verhältnis zwischen Prosodien und Phonotaktik. Wenn sich beide Strategien widersprechen, verlassen sich Kleinkinder eher auf prosodische Informationen. Dabei dauert die Herausbildung von phonotaktischem Wissen der Sprache auch einige Wochen länger als die prosodische Sensibilität (Friederici und Wessels, 1993), die, wenn einmal mit den Wortgrenzen verknüpft, immer als vorrangiges Instrument behandelt wird. Morgan (1994, 401) stellt den komplementären Charakter von Rhythmus und distributionalen Eigenschaften für acht Monate alte Babys, beruhend auf Experimenten mit veränderter Methodik, heraus. Dieses Ergebnis sagt natürlich noch nichts über die genaue Abfolge aus oder darüber, welcher Mechanismus vorherrschend benutzt wird. Johnson und Jusczyk (2001) und Thiessen und Saffran (2003) klären dieses Ineinandergreifen von prosodischen und statistischen Segmentierungsmechanismen. Letztlich sind die Mechanismen der allophonischen Variation mit zehn Monaten verfügbar (Jusczyk et al., 1999c). Auf diese Weise verdichten sich die einzelnen Theorien zur Wortsegmentierung zu einem einheitlichen Theoriegebäude. Die anstehende Simulation kann nun klären, ob die vorgeschlagene Abfolge der übergeordneten Theorie hinsichtlich ihrer Informationsdichte des ersten Segmentierungsmechanismus durchführbar ist. Davor müssen allerdings die mit Experimenten abgesicherten Randbedingungen der Simulation erläutert werden.

⁵⁹Für eine feinere Aufschlüsselung der Entwicklung von Prosodien und Übergangswahrscheinlichkeiten des interessierenden Zeitraums siehe Abbildung 3.2.

⁶⁰Dies ist nicht zuletzt auch deshalb so, weil es Umgebungen ohne die notwendigen akustischen Anhaltspunkte gibt.

4 Die frühkindlichen Sprachperzeptionsfähigkeiten als Randbedingungen der Simulation

In der dieser Arbeit zugrunde liegenden Simulation wurde von fünf kognitiven Fähigkeiten eines sechs bis acht Monate alten Kindes ausgegangen: die Wahrnehmung von sprachlichen Einheiten, das Erkennen von Übergangswahrscheinlichkeiten und hochfrequenten Einheiten im Lautstrom, das Erinnerungsvermögen von Lautsequenzen und *Top-down*-Mechanismen von mentalen Repräsentationen. Bei der Entwicklung einer Simulation müssen die Annahmen und Voraussetzungen des Systems gut begründet sein, soll sie denn dem Anspruch einer erklärenden Theorie gerecht werden. Möchte man erklären, wie die Lautsegmentierung im frühen Kindesalter funktioniert ohne dabei auf ein Sprachmodul im Sinne der Nativisten zurückzugreifen, dann dürfen in den Annahmen keine „sprachorganischen“ Eigenarten untergebracht sein. Die modelltheoretischen Annahmen der Simulation werden im folgenden Kapitel hinsichtlich ihrer empirischen Realität aus der Perspektive der Psycholinguistik und Entwicklungspsychologie beschrieben. Es soll also gezeigt werden, dass die gemachten Annahmen tatsächlich experimentell nachweisbar sind und deshalb als plausible Voraussetzungen in der Simulation verwendet werden dürfen. Zudem soll so gezeigt werden, dass in den Randbedingungen keine sprachspezifischen Informationen enthalten sind.

Während die Frage nach der Wahrnehmung von sprachlichen Einheiten eine Kontroverse in der Psycholinguistik auslöste, sind die restlichen Fähigkeiten in Experimenten nachgewiesen und gaben weniger Anlass zu einer ausgeweiteten Diskussion. Deshalb wird auf die Wahrnehmung der sprachlichen Einheiten detaillierter eingegangen, um Entwicklung und kontroverse Standpunkte besser beurteilen zu können. Ein zweiter Grund für eine umfassendere Beschreibung der Wahrnehmungsfähigkeiten sprachlicher Einheiten ist der, dass die Gefahr bestehen würde, von einer widersprüchlichen Annahme im Modell auszugehen wie z.B. eine perfekte Silbifizierung des Sprachinput (Gambell und Yang (2005) oder Swingley (2005)¹). Unterstellt man nämlich hier schon höher geordnete Einheiten, welche auf eine sprachliche Hierarchie zurückzuführen sind, steht die Frage im Raum, wie diese Einheiten überhaupt mit allgemeinen kognitiven Fähigkeiten wahrgenommen werden können oder ob nicht dann die hierarchische Struktur der Spracheinheiten, wie in einem *x-bar*-Schema, genetisch kodiert sein müsste.²

¹Hinsichtlich seines Erkenntniszieles mag diese Vorgehensweise vertretbar. Swingley modifizierte in der zweiten Simulation den Silbifizierungsalgorithmus der Korpora möglicherweise deshalb, weil er nicht erklären konnte, wie ein fehlerfreier silbifizierter Korpus in der Realität des Kindes als Sprachinput zustande kommen kann. Er widmet dieser Problematik immerhin eine Seite, auf der er aber nicht auf die wirklichen Widersprüche eines silbifizierten Inputs hinsichtlich der Sprachenvielfalt eingeht.

²Vgl. Vroomen et al. (1998) für ein SRN zur Silbifizierung eines Sprachstroms auf Grundlage der Sonorität der Phoneme, welches „... regarding the syllabic structure were far from perfect“ (217).

4.1 Wahrnehmung von sprachlichen Einheiten

Eine intensive Auseinandersetzung mit dem Thema der kleinsten mentalen Repräsentationseinheiten bei der Sprachperzeption³ ist für die Gesamtkonzeption der Arbeit unerlässlich, weil zuerst geklärt sein muss, von was denn eine „Statistik“ errechnet werden soll. An dieser Stelle muss geklärt werden, was die Grundlage der Kalkulationen der kleinen Sprachlerner ist. Sollen die einzelnen Phoneme als Repräsentationen der Sprache, die Silben oder Einheiten irgendwo dazwischen verwendet werden?⁴

Die erste Frage, die bei der vorliegenden Modellentwicklung für bedingte Wahrscheinlichkeiten auftritt, läuft also auf die grundlegende Einheit einer statistischen Berechnung hinaus. Von welchen Einheiten soll eine Statistik erstellt werden? Erwartungsgemäß findet die empiristisch-rationalistische Debatte auch hier ihren Niederschlag. So stehen sich Phonetiker und formale Phonologen gegenüber, konzentrieren sich allerdings auf unterschiedliche Entwicklungsstufen des Kindes. Erstere bevorzugen die frühe Entwicklung und überlassen den Rationalisten das Feld ab einem Alter von zwei Jahren (vgl. Macken, 1995). Letztlich gibt es genug Hinweise, dass Sprache sowohl erlernte als auch angebohrte Eigenschaften hat (vgl. Macken, 1995, 672). Hierin spiegelt sich abermals die im Kapitel 2 angerissene Diskussion wider. Im Hinblick auf die psychologische Plausibilität, die im hiesigen Modell angestrebt wird, muss diese Frage in die Sprachentwicklungsforschung übergeleitet werden. Es hängt also davon ab, welche Einheit das Kleinkind als Repräsentationsschablone verwendet. In der Sprachentwicklungsforschung ist dies trotz intensiver Bemühungen für die frühe Sprachperzeption bislang ungeklärt (vgl. Goldinger und Azuma, 2003, 307). Trotz einer Reihe von unterschiedlichen Repräsentationsmo-

³Um begriffliche Interferenzen zwischen ähnlichen Fragestellungen zu vermeiden, sei hier explizit darauf hingewiesen, dass die Verwendung des Ausdrucks der mentalen Repräsentationseinheiten nicht mit der Idee der *Mental-Imagery*-Debatte in Zusammenhang steht, zumindest nicht direkt. Repräsentationseinheiten sind hier im Sinne von kleinsten Wahrnehmungseinheiten (*minimal perceptual units*) zu verstehen, die zur Lautsegmentierung verwendet werden. Sie sind also nicht als abstrakte Konzeptvorstellungen für semantische Bilder und deren mentale Verknüpfungen auszulegen.

⁴Auch die Vorstellung phrasaler linguistischer Einheiten wurde in die Debatte eingebracht (Peters (1983) oder Hirsh-Pasek et al. (1987)). Peters stellt die These auf, dass Kleinkinder *speech formulas* (2) verwenden. Das sind satzähnliche Fragmente, welche als abgeschlossene Einheiten aufgenommen werden. Da wir die kleinsten Spracheinheiten aus Sicht der Erwachsenenlogik betrachten, verfehlen wir die Ausgangslage der Kleinkinder zu berücksichtigen. Als Erklärung schlägt Peters die Gestalttheorie in Kombination mit anderen analytischen Ansätzen vor. Ihre Beobachtungen stützen sich aber hauptsächlich auf Kinder, die bereits Äußerungen produzieren können. Ab diesem Zeitpunkt ist die kognitive Entwicklung des Kindes jedoch unzweifelhaft weiter fortgeschritten und die Anfänge der semantischen Zuordnung zur lautlichen Gestalt ist in vollem Gang. Offensichtlich ist ihre Auffassung der *minimal units of speech* doch etwas weit reichender. Wie sonst würde man erklären können, dass Kleinkinder doch die lautliche Erscheinung kleinerer Einheiten sehr wohl unterscheiden können. Genau die richtigen Diskriminierungen dieser Einheiten sind aber wichtig. Folgt man nämlich der Auffassung, dass das mentale Lexikon in wörterähnlichen Einträgen organisiert ist, dann müssen die sublexikalischen Einheiten kleiner als das Wort selbst sein (vgl. Frauenfelder und Floccia, 1999, 4). Wäre das Wort allerdings selbst eine Segmentierungseinheit, würde das Segmentierungsproblem gar nicht bestehen können.

dellen kann man aus der Literatur der letzten 40 Jahre entnehmen,⁵ dass der Disput sich nunmehr auf die Silbe und das Phonem als grundlegende Wahrnehmungseinheiten beschränkt.

Für die Erörterung der Frage, welche Repräsentationseinheiten in der Statistik von den Kleinkindern verwendet werden, ist eine allein auf das Kind gerichtete Untersuchung zu eng gefasst. Zum einen sind die methodischen Grundlagen der Experimente eingeschränkter als bei Erwachsenen und zum anderen können bei experimentellen Fehlern, die bei bestimmten Anordnungen hin und wieder vorkommen, bei Erwachsenen leichter überprüft beziehungsweise ausgeschlossen werden. Derartige Experimente braucht man dann nicht für Kleinkinder zu modifizieren. Die Betrachtung des Gesamtzusammenhanges ermöglicht es dann, die Validität der Ergebnisse zu verfeinern (vgl. Dietrich, 2007, 92).

Einen weiteren Vorteil einer breiter angelegten Diskussion gewinnt man zudem durch die Entwicklungsrichtung des Phänomens (vgl. Nusbaum und Goodman, 1994, 308ff). Unter der Annahme einer fehlerfreien empirischen Untersuchung erfährt man, ob die mentalen Repräsentationen bei Erwachsenen und Kindern, gleich oder verschieden sind. Bei Gleichheit kann man die Ergebnisse der Säuglingsstudien in lineare Entwicklungsmodelle, die bis ins Erwachsenenalter reichen, integrieren. Lineare Entwicklungen, erlauben die Fähigkeiten von Erwachsenen auf die der Kinder rückzuschließen. Sind die mentalen Repräsentationen jedoch verschieden und weisen keine offensichtlichen Zusammenhänge zu denen von Erwachsenen auf, muss überlegt werden, wie das Kind seine mentalen Repräsentationen im Zeitverlauf möglicherweise ändern wird. Kennt man die Phänomene, welche die Sprachentwicklung beim Kind bestimmen, können widersprüchliche Entwicklungsszenarien im Vorfeld ausgeschlossen werden.⁶ Aus diesem Grund achtet man in der psycholinguistischen Forschung streng auf die Vereinbarkeit zwischen den Ergebnissen der Spracherwerbsforschung und den experimentellen Voraussetzungen, die bei Erwachsenen unterstellt werden.

According to one major psycholinguistic constraint on theory construction, the processing mechanisms postulated in adults should be tied to acquisition mechanisms in infants. (Frauenfelder und Floccia, 1999, 28)

Mithin erscheint es also sinnvoll, einen Abriss über die Forschungsergebnisse von Erwachsenen und Kleinkindern zu geben, um anhand dessen einen Einblick in die Problem-

⁵Seit dem Aufsatz von Eimas et al. (1971), vgl. auch Pisoni (1981) als eine Zusammenfassung der bis dahin 30 jährigen Forschung zur Sprachwahrnehmung.

⁶Das Erlernen des englischen Präteritums beispielsweise verläuft gemessen an der Bildung korrekter Vergangenheitsformen im Zeitverlauf nicht linear, da es nach dem Erlernen der Bildungsregel erst noch die Unterscheidung zwischen unregelmäßigen und regelmäßigen Verben erkennen muss. Auf diese Weise ergeben sich unterschiedliche Richtungsvektoren für die Entwicklung der Vergangenheitsform. Mit einem Jahr könnte das Kleinkind mit *go* alle Zeitformen des Vorgangs „gehen“ umschreiben. Wenige Monate darauf weiss es die Vergangenheit dieses Vorgangs mit *went* zu kennzeichnen. Bald lernt es aber die Bildungsregel des englischen Präteritums und wird mit vielleicht drei Jahren plötzlich *goed* hervorbringen. Meist noch innerhalb eines Jahres bemerkt das Kind aber die Ausnahmen zu unterscheiden und sagt von da an wieder *went*. Der Lernprozess bei Erwachsenen, welche Englisch als Zweitsprache erwerben, verläuft grundsätzlich anders.

4 Randbedingungen der Simulation

stellung zu erhalten und einen plausiblen Schluss für eine modelltheoretische Annahme zu ermöglichen.⁷

Die Vorgehensweise in diesem Unterkapitel sei wie folgt gegliedert: Zuerst wird eine Darstellung über die relevante Forschung der Repräsentationseinheiten im Allgemeinen gegeben. Darauf aufbauend kann der Fokus zur Forschung der frühkindlichen Wahrnehmungsfähigkeiten für kleinste linguistische Einheiten überschwenken. Letztlich wird aus den so gewonnenen Erkenntnissen, die beste Wahl einer Repräsentationseinheit für die anstehende Simulation getroffen.

4.1.1 Wahrnehmung linguistischer Einheiten im Allgemeinen

Eine tiefer gehende Diskussion wird hier mit den Erkenntnissen des letzten Drittels des vergangenen Jahrhunderts beginnen.⁸ Die seitdem verwendeten Methoden messen, von we-

⁷Jakobson (1972) beziehungsweise (1968) und ihm folgend eine Reihe weiterer Vertreter der starken Gleichheitsthese (*Strong Identity Thesis*) vertreten die Ansicht, „... that both children and adults create such sound patterns independently because they possess the same physical phonetic apparatus“ (Macken, 1995, 675). Dies klärt zwar nicht eine eventuelle Gleichheit der mentalen Repräsentationen in frühkindlichen und erwachsenen Gehirnen und damit die Übertragbarkeit der Ergebnisse aus Untersuchungen mit Erwachsenen, die aufgrund der breiteren Palette an experimentellen Anordnungen größere Validität der Ergebnisse aufweisen, aber es gibt zumindest Anlass, die Möglichkeit, dass die mentalen Repräsentationen bei Kindern und Erwachsenen gleich sind, nicht auszuschließen. Dehaene-Lambertz und Gliga (2004) argumentieren in diesem Sinne „... that infants have access at the beginning of life to phonemic representations, which are modified without training or implicit instruction, but by the statistical distribution of speech input in order to converge to the native phonemic categories.“ (1375). Ein ähnliches Ergebnis ist schon in Dehaene-Lambertz und Baillet (1998) für drei bis vier Monate alte englische und französische Babys bestätigt wurden.

⁸Sicherlich lässt sich die Problematik schon davor mehr oder weniger genau thematisieren oder in abgewandelter, philosophisch abstrakter Form wiederfinden. Die hier vorgeschlagene zeitliche Festsetzung ist dennoch nicht willkürlich getroffen. Erstens hat sich die Psycholinguistik seit dieser Zeit vom Einfluss der generativen Transformationsgrammatik emanzipiert und tritt in eine neutrale, von Knobloch (2003, 16) implizierte, fünfte Entwicklungsphase ein. Ein zweiter Grund besteht in der wissenschaftshistorischen Bedeutung der Psycholinguistik für diese Zeit. Die experimentelle Psycholinguistik betont im Nachhinein ihre bereits länger als hundertjährige Tradition der empirischen Arbeit seit Broca, Wernicke und Wundt. Obwohl es natürlich bei Broca und Wernicke auch um sprachpsychologische Erkenntnisgewinne ging, ordnet man diese Ergebnisse gewöhnlich doch noch der angewandten Psychologie zu, da sich die Psycholinguistik, wenn überhaupt als solche identifiziert, als reines Teilgebiet der Psychologie, die ebenfalls noch in den Kinderschuhen steckte, oder gar der Medizin beziehungsweise Physiologie verstand. Drittens erlaubten technische Neuerungen weitergehende Erkenntnisse. Die Reaktionszeitmessung ist eine der ältesten Verfahren (Scharlau et al., 2003, 190), die in psychologischen Experimenten Anwendung fand. Spezielle Verfahren für die Sprachperzeption (wie das Tonband, der Spektrograph oder der Synthesizer), entstanden aber erst im oder nach dem zweiten Weltkrieg und fanden in der universitären Forschung erst in den 50er und 60er Jahren allgemeine Verbreitung (vgl. Fischer-Jørgensen, 1961, 116ff); sicher auch, um die behavioristischen Gedanken experimentell vorzuführen. Letztendlich dankt die Psycholinguistik einen Großteil ihrer Eigenständigkeit der kognitiven Revolution, die sich spätestens in den 70er Jahren vollzogen hatte, und sich unter dem Druck zweier gegenteiliger Theorien, als eine brauchbare Zweickwissenschaft emanzipierte. Sicherlich nicht ganz zufällig finden sich seitdem auch in der Linguistik experimentelle Vorgehensweisen (vgl. Sichelschmidt und Carbone, 2003, 115). Genau in dieser Zeit

nigen Ausnahmen in jüngster Vergangenheit abgesehen, die interessierenden Sachverhalte auf indirektem Weg.⁹ Die Folge davon ist der relativ große Interpretationsspielraum, den die Ergebnisse gestatten. Es kann demnach keine, im naturwissenschaftlichen Sinne, eindeutige Lösung geben.¹⁰ Mit einer angemessenen Abwägung aus der Qualität der vorgebrachten Argumente und unter Berücksichtigung des wissenschaftlichen Umfeldes wird sich aber dennoch eine Tendenz zur Wahrheit finden lassen. Es geht also in erster Linie um eine tendenzielle Sympathienahme für eine Theorie, die nicht willkürlich, sondern durch die sorgfältige Beachtung der als relevant erachteten Einflussfaktoren geschieht.

Im Zuge der universalgrammatischen Einflüsse, die sich spätestens seit Chomsky und Halle (1968)¹¹ auch auf die phonetischen und phonologischen Modelle ausbreiten sollten, untersuchten Foss und Lynch (1969), wie sich die Reaktionszeit der Phonemerkennung in Abhängigkeit von der Komplexität der umgebenden Oberflächenstruktur verhält. Sie fanden heraus, dass die Reaktionszeiten für ein bestimmtes Phonem in rechtsverzweigten Sätzen kleiner ist als in den komplexeren selbsteingebetteten Konstruktionen. Zur

konnte man erstmals die experimentelle Methodik zur Datenerhebung bei Säuglingen so modifizieren (von Eimas et al. (1971) vorgestellt), dass die Ergebnisse wissenschaftlich nutzbar waren, also auch replizierbar und valide. Im letzten Drittel des vergangenen Jahrhunderts treffen also zwei wichtige Faktoren aufeinander. Einmal die positive Neuausrichtung der linguistischen Wissenschaft und zweitens eine sprichwörtliche Explosion an psycholinguistischen Fragestellungen, die gerade im Bereich der akustischen Sprachperzeption nicht mehr nur auf Reaktionszeitmessungen beschränkt blieben, sondern ein breiteres Spektrum an wissenschaftlichen Instrumentarien zur Verfügung hatten, wie beispielsweise die Fähigkeit der genauen experimentellen Überprüfung der Phänomene des frühkindlichen Spracherwerbs. Das gestiegene Interesse an der Sprachpsychologie lässt sich schon allein an dem gestiegenen internationalen Interesse erkennen, welches insbesondere in Europa erste psycholinguistische Forschungen vorantrieb (vgl. Knobloch, 2003, 27). Vor diesem Hintergrund ist der gewählte Zeitpunkt als Themeneinstieg zu verstehen (siehe auch Harnish, 2002, 1).

⁹Oft wird speziell im Zusammenhang mit den phonetischen Repräsentationen auf die Arbeit von Jakobson (1972) verwiesen. Sicherlich auch vor dem Hintergrund seiner Vielsprachlichkeit hat er das Phonem als grundlegende Spracheinheit nicht in Frage stellen können, bedenkt man die unterschiedlichen Silbensysteme des Tschechischen und des Französischen. Nur kurz schneidet er die Frage nach der Strukturzusammensetzung höherer Spracheinheiten an (84ff) und führt sie im Wesentlichen auf das Phonem zurück. Schon in dem wenige Jahre zuvor erschienenen Büchlein (Jakobson et al., 1967) widmet sich Jakobson der Phonemanalyse. Wenn auch nicht in Bezug zum Erstspracherwerb, so werden doch die Präferenzen der Autoren deutlich, wie die Sprache analysiert werden müsste (S. 1ff). Stellt man hier noch einen Zusammenhang mit der Jakobsonschen *strong identity thesis*, die zu diesem Zeitpunkt schon gut formuliert war, her, kann kaum mehr Zweifel an der Haltung Jakobsons zu den grundlegenden Repräsentationseinheiten erhoben werden.

¹⁰Natürlich könnten in wenigen Jahren die technischen Entwicklungen soweit gediehen sein, dass auch direkte Nachweise mittels bildgebender und elektrophysiologischer Verfahren gelingen und Einblicke in die sprachverarbeitenden Regionen des Gehirns geben. Erst damit könnte der direkte Beweis geführt werden und die Debatte der kleinsten Repräsentationseinheiten beendet werden. Da aber unklar bleibt, zu welchem Zeitpunkt diese Verfahren zu dieser Reife gelangen, steht es sicherlich im Interesse des schnelleren Erkenntnisgewinns eines jeden Forschers auch mittels indirekter Methoden die Geheimnisse unserer Welt zu lüften.

¹¹Siehe aber auch schon Halle (1956, 510) oder Licklider (1952, 590). In beiden Quellen wird eine Sequenz von Phonemen als Repräsentationsform eines Wortes vorausgesetzt. Vgl. zudem Fitzpatrick und Wheeldon (2000, 137) sowie Lahiri und Marslen-Wilson (1991, 248) und als Gegenargument die Studie von Morais et al. (1979).

4 Randbedingungen der Simulation

Erklärung nahmen die Wissenschaftler an, dass alle linguistischen Ebenen bis hin zur Tiefenstruktur analysiert werden und folglich bei komplexeren Sätzen mehr Zeit anfällt (*time-sharing hypothesis*).

Wenig später ermittelten dann Savin und Bever (1970) in einer experimentellen Anordnung die Reaktionszeit zur Erkennung von Silben und Phonemen bei Erwachsenen. Sie verwenden dabei ausschließlich einsilbige Kunstwörter der Form CVC¹² (*baeb* oder *saeb*). Die Reaktionszeiten für die Erkennung für Silben lagen weit unter denen der Phoneme, was als Motiv für die Silbe als Perzeptionseinheit gewertet wurde. Um einen offensichtlichen Widerspruch mit der Morphologie, die Bildung des Plurals im Englischen mit /s/ oder /z/ beispielsweise, zu vermeiden, erklärten sie „that phonemes are perceived only by an analysis of already perceived syllables“ (300). Weiterhin erachten Savin und Bever die Phoneme „neither perceptual nor articulatory“, sondern als „psychological entities of a nonsensory, nonmotor kind“ (301), daher gelten sie als abstrakte Einheiten. Dies implizierte allerdings, dass die Silben als real angesehen werden mussten.

Kritik an der methodischen Vorgehensweise von Savin und Bever (1970) übten Foss und Swinney (1973) und kurz darauf McNeill und Lindig (1973). Foss und Swinney (1973) zeigten, dass zweisilbige Zielwörter schneller entdeckt wurden als deren Anfangsilbe und diese wiederum schneller als die Anfangsphoneme dieser Silbe. Aufgrund der umgekehrten Proportionalität zwischen Reaktionszeit und Silben- beziehungsweise Phonemanzahl schließen die Autoren, dass die alt bewährte Reaktionszeitmessung nur begrenzte Aussagen über dieses Problemfeld erlaubt. So waren die Reaktionszeiten zur Erkennung der Silben nur halb so kurz wie deren akustische Präsentation, was zumindest auf überlagerte Effekte hinweist. Zur Lösung des Problems schlugen sie eine Unterscheidung zwischen *identification* und *perception* (254) für die Repräsentationseinheiten vor, wobei die Identifikation des Phonems direkt erfolgen müsse, aber größere Einheiten immer weiter in ihre Bestandteile heruntergebrochen werden. Damit ändert Donald Foss seine Auffassung, die er noch einige Jahre zuvor vertreten hatte (Foss und Lynch, 1969). Natürlich gestehen die Autoren den abstrakten Charakter der Phoneme ein, bestehen aber gleichzeitig auf der abstrakten Natur der Silben.¹³ Zudem lassen die Ergebnisse den Schluss zu, dass zuerst ein lexikalischer Abgleich erfolgt und erst dann die phonologische Verarbeitung eintritt. Sie weisen zusätzlich darauf hin, dass „the representation of the input in terms of phonetic syllables will be very uneconomical“ (247). Dieses Argument, auch als *Occam's Razor* bekannt, sagt zwar erst einmal nichts über die psychologische Richtigkeit aus, aber angesichts besserer Alternativen, nämlich der Repräsentation von etwa 40 bis 50 abstrakten Phonemen, fällt es schwer, die rechenaufwendigen Repräsentationen in Tausenden von Silben vorzustellen.¹⁴

¹²C = *consonant* = Konsonant und V = *vowel* = Vokal.

¹³Diesen Punkt hat wenig später auch Studdert-Kennedy (1974, 2366) noch einmal betont und dabei die Abstraktheit der Silben unterstützt.

¹⁴Wickelgren (1969) vertritt die Auffassung, dass Phoneme als abstrakte Einheiten überhaupt nicht existieren, sondern Produkte unserer geistigen Vorstellung sind. In der neuronalen Realität existiert nur eine Vielzahl an Allophonen, die er auf 106 interne Repräsentationen schätzt – eine Zahl, die er im Vergleich zur Gesamtzahl der Repräsentationen (1010) als unvernachlässigbar impliziert.

4.1 Wahrnehmung von sprachlichen Einheiten

Noch deutlicher legen die Ergebnisse von McNeill und Lindig (1973) das Defizit der Reaktionszeitmessung offen. Eine erste Verzerrung würde auf jeden Fall auftreten, weil die Listen von Savin und Bever (1970)) lediglich aus Silben bestünden. Diese Korrelation belegten sie, indem sie Listen mit Einheiten unterschiedlicher Länge präsentierten. Ein ganzer Satz wurde schneller erkannt als eine darin enthaltene Phrase, die wiederum schneller als ein Wort und in gleicher Weise Silben und Phoneme. Damit ist klar, dass die Reaktionszeit nichts über die Repräsentation der linguistischen Einheiten aussagt.

Healy und Cutting (1976) schlagen versöhnliche Töne an. Sie räumen den Silben wie auch den Phonemen in gleicher Weise ihre Eigenschaft als grundlegende Repräsentationseinheit ein. Wie auch schon McNeill und Lindig (1973) beobachteten Healy und Cutting, dass sich die Reaktionszeiten in Abhängigkeit von den präsentierten Listen änderten. Folglich führten sie Experimente mit heterogenen Listen durch, die Phoneme wie auch Silben beinhalteten. Phoneme und Silben waren so geordnet, dass einmal Phoneme und zum anderen Silben einfacher zu identifizieren waren. Wie man anhand der Listen voraussagen konnte, änderten sich auch tatsächlich die Reaktionszeiten. In einer Liste wurden die Phoneme schneller erkannt und in der anderen Liste die Silben.

Morton und Long (1976) interpretierten dann noch einmal die Ergebnisse von (Foss und Lynch, 1969) entsprechend ihrem derzeitigen Kenntnisstand um. Sie führten die Unterschiede in den Reaktionszeiten im Wesentlichen auf die Vorhersagbarkeit des Wortes zurück, indem das Phonem enthalten ist. Ist die Übergangswahrscheinlichkeit vom Vorgängerwort groß, so verringert sich die Erkennungszeit bei konstanter Silbenzahl und gleicher Vorkommenshäufigkeit der betreffenden Wörter. So werden in ihrem Satz, „At the sink she washed a . . . , thinking of the time when she was younger“ *plate* und *pan* eingesetzt und die Zeiten gemessen, wie lange gebraucht wird, das Phonem /p/ zu erkennen. Hängen die Reaktionszeiten vom Wort ab, dann muss das Wort die Einheit sein, wie auch McNeill und Lindig (1973) oder Foss und Swinney (1973, 48) zustimmen würden. „The present results rule out this possibility for a model in which word recognition and phoneme detection are independent“.

Diese Interpretationen stehen noch ganz im Geiste von Chomsky und Halle (1968, 7ff und 24ff), die annahmen, dass der Hörer phonologische Repräsentationen bildet und nicht auf die phonetischen Spezifika reagiert. Foss und Blank (1980) zeigten nun allerdings, dass der Hörer durchaus eine duale Strategie verfolgen kann. Sie unterschieden zunächst zwischen den beiden Repräsentationsformen, einem phonetischen und einem phonologischen Code. Ersterer ist prälexikalisch und wird vor dem lexikalischen Abgleich erkannt. Letzterer ist postlexikalisch, d.h. hier erfolgt zuerst ein Zugriff auf das mentale Lexikon. Die Autoren zeigten nun experimentell, dass die Reaktionszeiten des Zielphonems deutlich kleiner sind als die Dauer des Wortes, in welchem das Phonem am Anfang enthalten ist. Dies gilt auch für Kunstwörter, die ja nicht im Lexikon enthalten sein können. Die Reaktionszeit änderte sich nicht in Abhängigkeit von der Vorkommenshäufigkeit der Wörter. Allerdings übersteigen die Reaktionszeiten der Phoneme die ihrer Wörter, wenn das Vorgängerwort ein Kunstwort ist oder die Übergangswahrscheinlichkeiten im Sinne von Morton und Long (1976) geändert werden. Im ersten Fall können die

4 Randbedingungen der Simulation

Probanden die Wortgrenze nicht eindeutig festlegen und greifen deshalb auf das Lexikon zurück. Der zweite Fall ist in Übereinstimmung mit Morton (1969) auf die Vorhersagbarkeit zurückzuführen. Alle Ergebnisse dieser Experimentenreihe lassen sich schließlich mit der *Dual-Code*-Hypothese erklären: Ist die phonetische Repräsentation nicht eindeutig, erfolgt eine Überprüfung mit den lexikalischen Einträgen, die in Übereinstimmung mit den phonologischen Regeln eindeutig bestimmt werden. Ansonsten erfolgt eine direkte phonetische Identifikation. Nur knapp drei Jahre danach nimmt Foss (Foss und Gernsbacher, 1983) von seinem ursprünglichen Vorschlag dualer Codices abermals Abstand. Aufgrund der unerwarteten Ergebnisse der Experimente schreiben die Autoren, fast schon überraschend ehrlich, der reinen vorlexikalischen Repräsentation die höhere Wahrscheinlichkeit zu. Allerdings lassen auch ihre Ergebnisse noch keinen eindeutigen Schluss für eine bestimmte Repräsentationsform zu. Sie geben für den Fall einer CV-Silbe als Repräsentationseinheit jedoch zu bedenken, dass sich diese sehr schnell in ihre phonemischen Bestandteile zerlegen lassen müsse (vgl. Foss und Gernsbacher, 1983, 631).

Einen entscheidenden Beitrag zur Diskussion der grundlegenden linguistischen Einheiten haben Mehler et al. (1981), Cutler et al. (1983) und Cutler et al. (1986) geleistet. Seitdem kann eine sprachspezifische Unterscheidung zwischen dem Englischen und dem Französischen hinsichtlich ihrer Repräsentationseinheiten getroffen werden.¹⁵ Zunächst entdeckten Mehler et al. (1981), dass französische Sprecher die Silbe als Segmentierungseinheit verwenden. Wenn eine CV-Silbe in einem CV-CVC-Wort erkannt werden sollte (wie *ba* in *balance*), reagierten die Probanden signifikant schneller als wenn ein CVC-CVC-Wort (wie z.B. *balcon*) vorgegeben wurde. Das umgekehrte Resultat erhielten die Wissenschaftler bei Verwendung einer CVC-Silbe (wie z.B. *bal*) als Vergleichseinheit. Die Silbe wurde deshalb als kleinste Repräsentationseinheit verkündet.¹⁶ In einer weiteren Reihe von Experimenten sollte dies auch für Sprecher des Englischen gezeigt werden (Cutler et al., 1983). Es stellte sich jedoch heraus, dass englische Sprecher, die Silbe nicht als Segmentierungsraster nutzten. Bei letzteren konnte kein statistisch signifikanter Zeitunterschied gemessen werden. Die Autoren hielten damit fest:

We conclude, therefore, that the syllabification strategy is characteristic of listeners rather than of stimulus language. (Cutler et al., 1983, 169)

1986 folgte dazu ein weiteres Experiment (Cutler et al., 1986), welches die bereits vorhandenen Ergebnisse zweifellos bestätigte. Französische Sprecher teilten französische sowie auch englische Wörter nach dem französischen Silbenmuster auf. Bei den englischen Probanden konnte sich das weder für französische noch für englische Wörter feststellen lassen.

Treiman (1983) stellt die Frage nach der internen Struktur der Silbe. Eine Silbe darf nicht als reine Abfolge von Phonemen gewertet werden, weil sie eine interne Struktur aufweist, die eine Analyse höherer Verarbeitungsmechanismen erfordert. Sie verwendet „Wortspiele,“ bei denen die Probanden durch Regelinduktion die Lösung des Spiels herbeiführen konnten. Diese Methode ist sicher nicht unstrittig, da anhand des präsentierten

¹⁵Erst später dehnte man diese Unterteilung auch auf das Japanische aus (Cutler, 1994).

¹⁶vgl. Segui (1984), Segui et al. (1995), Segui et al. (2001).

Materials bereits eine bestimmte Regelauswahl vorgegeben wird. Trotzdem werden ihre Ergebnisse oft als Untermauerung für die Silbe als Perzeptionseinheit herangezogen. Die Wortspiele gaben für die Autorin Anlass, eine interne Struktur der Silbe anzunehmen, die auch als mentale Repräsentation existent sei. Die Struktur der Silbe ließe sich zumindest in Silbenkopf und -reim begründen. Offen bleibt allerdings, ob sich der Reim nochmals hierarchisch in Kern und Koda teilen lässt.

[t]he syllable structure . . . suggests that a hierarchy of units is involved, including syllables, phonemes, and intrasyllabic units such as onset and rimes. That is, there exist units intermediate between the syllable and the phoneme-units. (Treiman, 1983, 73)

Letzteres, so Treiman, müsse in einer Theorie der Sprachperzeption berücksichtigt werden.

Pisoni und Luce (1987) geben in ihrer *state of affairs* zu erkennen, dass die Frage nach der linguistischen Einheit falsch beantwortet wird. Es gibt nicht nur eine Einheit, sondern eine von der Aufgabe beziehungsweise von der Situation abhängige Einheit. Des Weiteren behalten sie die vier Phasen *auditory*, *phonetic*, *phonological* und *lexical, syntactic, semantic* von Studdert-Kennedy (1974) bei, fügen aber noch eine Vorstufe, die *peripheral auditory analysis*, hinzu. Während der Echtzeitanalyse des Signals (*peripheral auditory analysis*) sind Einheiten wie die Phoneme eher unwahrscheinlich, da die fehlende direkte Beziehung (*one-to-one-relation*) zwischen dem Sprachsignal und einer Einheit nicht zu bestimmen ist. Hier sind kontextsensitive Diphone (Spektra) vielversprechender. Für die nachgelagerte akustisch-phonetische Analyse ist eine phonetische Matrix (Chomsky und Halle, 1968) vorteilhaft. Letztlich sind aber segmentale Repräsentationen unerlässlich für die Sprachperzeption. Im Grunde werden dazu Muster aus dem ursprünglichen Signal (*peripheral auditory analysis*) generiert und auf höheren Ebenen zu diskreten Einheiten weiterverarbeitet.¹⁷

Ein späterer Aufsatz (Norris und Cutler, 1988) legte eine weitere Unzulänglichkeit in der bis dato verwendeten Methodik offen. Insbesondere geht es um die Problematik, was genau mit der Reaktionszeit gemessen wird. Einmal mehr entfachen (Norris und Cutler, 1988) die Diskussion, welche auch schon bei Savin und Bever (1970), Foss und Swinney (1973) oder McNeill und Lindig (1973) unterschwellig anklang, von Foss und Swinney (1973) letztendlich nicht überzeugend widerlegt werden konnte. Ändert man das Stimulusmaterial auf die eine oder andere Weise, so treten derart gegensätzliche Reaktionszeiten auf, dass man zu dem Schluss kommen muss: „. . . that faster detection times for syllables than for phonemes are completely artifactual“ (Norris und Cutler, 1988, 542). Die Autoren führen als Lösung eine Reihe sogenannter *foils* ein. Beispielsweise dürfen keine Einheiten des dargebotenen Materials mit der Zieleinheit gänzlich übereinstimmen. Nur eine Einheit weist eine hohe Ähnlichkeit auf. Dieses Verfahren würde sicherstellen, dass sowohl Silben als auch Phoneme gleichermaßen von den Versuchspersonen analysiert würden und keine Entscheidung aufgrund einer Teilanalyse getroffen

¹⁷Harris und Lindsey (1993) sind dabei der gegenteiligen Ansicht. Sie legen dar, dass was wir zu hören glauben, bereits eine phonologische Repräsentation ist. Der menschliche auditive Perzeptionsapparat sei nicht in der Lage eine phonetische Repräsentation in ihrer Vollständigkeit aufzunehmen, sondern nur gefilterte Informationen, welche dann als „elementare Muster“ weiterverarbeitet werden.

4 Randbedingungen der Simulation

werden könne. Mit dieser neuen Methodik werden nun Phoneme schneller erkannt als Silben.

Ein weiterer Hinweis für eine phonemische Kodierung wird durch den Befund von Miller und Dexter (1988) gegeben. Sie zeigten, dass bei variierendem Sprechtempo in einsilbigen Wort- und Nichtwortpaaren, die sich nur im Stimmtoneinsatz des ersten Konsonanten unterscheiden (*beef/peef* oder *beace/peace*), ausreicht, das erste phonetische Segment zu entschlüsseln. Dieses Ergebnis spricht wieder gegen die Silbe als *processing unit*.

Pitt et al. (1990) unterstellen, wenn auch nicht in direkter Erwähnung, eine lineare Abfolge von Phonemketten – auch wenn dadurch eine Struktur entsteht, wie sie von Treiman (1983) abgelehnt wurde, weil sie wahrscheinlich von der Möglichkeit einer phonetisch-phonologischen Wahrnehmungseinheit im Voraus überzeugt sind. In drei Experimenten verfolgen sie das Ziel, die Aufmerksamkeit ihrer Probanden auf bestimmte Phoneme innerhalb einer Silbenstruktur während der Sprachverarbeitung zu lenken. Ihre Argumentation besagt, wenn sich die Sprecher auf ein einziges Phonem, das ja kleiner als die Silbe ist, konzentrieren können, dann muss dies auch die kleinste linguistische Einheit sein, weil dies ohne eine entsprechende Repräsentation nicht möglich wäre. Ihre Ergebnisse lassen sich in vier Punkten festhalten. Erstens sind die Probanden in der Lage, ihre Aufmerksamkeit äußerst präzise, zweitens sehr flexibel und drittens sehr schnell auf die vorbestimmten Einheiten zu lenken. Letztlich beeinflussen kleinere Störungen im Sprachmaterial das Ergebnis nicht. Bei größeren Verzerrungen des Inputs ist dies aber der Fall.

Pallier et al. (1993) bemängelten, dass die betreffenden Phoneme oft als Anfangskonsonanten der zweiten Silbe gesetzt wurden und dass der dabei gemessene Effekt der internen Silbenstruktur zufallen müsse. Die Silbenstruktur wurde also als eine Art Strukturierungsgitter über die Wörter gelegt, aufgrund dessen dann die Identifizierung des Phonems erfolgen konnte. Für ihren Beweis verwendeten die Autoren zwei Probandengruppen, die ein Zielphonem einmal in der Koda der Silbe und zum Anderen im Silbenkopf lokalisieren sollten. Es traten tatsächlich zeitliche Verzögerungen bei spanischen und französischen Probanden auf. Später replizierte Pallier (1998) das Experiment mit französischen Wörtern, bei denen die gleichen Ergebnisse hervorkamen wie bei Pitt et al. (1990). Erwachsene französische Sprecher können also auch sehr genau Phoneme lokalisieren, wie englische Sprecher, obwohl sie eine unterschiedliche Sensibilität zur Silbe haben (wie von Cutler et al. (1986) vorgelegt).

Allerdings konnte auch dieses Experiment nicht klären, warum spanische Sprecher keine Silbifizierung zeigen, wenn das *phoneme detection paradigm* benutzt wird (Sebastian-Gallés et al., 1992) trotz einer eindeutigen Silbenstruktur wie im Französischen. Die vorgeschlagene Erklärung von Pallier (1998), dass dies auf die Zahl der Vokale in beiden Sprachen zurückzuführen sei, mutet wenig überzeugend an.

Einen ganz anderen Ansatz nehmen Morais und Kolinsky (1994) ein. Sie beschreiben eine bewusste und unbewusste Ebene der phonemischen Wahrnehmung. Die bewusste Ebene tritt erst mit dem Erlernen des Alphabets auf. Die unbewusste Ebene steht da-

gegen schon sehr früh zur Verfügung. Sie ist für die Frage der Wahrnehmungseinheit interessant. Hier wird wie folgt argumentiert.

If speech attributes can be wrongly combined, they must have been separately registered as independent units at some earlier stage of processing. (Morais und Kolinsky, 1994, 290)

Goldinger (1998)¹⁸ vertritt die Meinung, dass keine abstrakten Einheiten, ganz gleich welcher Form, existent seien. Er gründet diese Aussage auf eine Reihe von Experimenten (*shadowing tasks*), welche belegen, dass umfangreiche Details des Sprachsignals, mögen sie auch überflüssig sein, im Langzeitgedächtnis gespeichert werden. Mit dem Abrufen einzelner Wörter werden auch umfangreiche episodische Informationen der Stimme mitgeliefert, die beim Hören des Lemmas präsent waren. Wenn ein bestimmtes Wort aber mit seinen höchst unterschiedlichen Ausprägungen detailliert gespeichert wird und als solches auch abrufbar ist, dann kann dies keine abstrakt diskrete Einheit sein und deren Bestandteile ebenso wenig. Es verwundert wenig, dass der Autor später eine neurale Netzwerktheorie¹⁹ heranzieht und damit tatsächlich die bereits vorgestellten, widersprüchlichen Ergebnisse gut erklären kann (Goldinger und Azuma, 2003). Allerdings muss man anmerken, dass die Architektur des Netzwerkes nicht vollständig der Aussage entspricht, dass es keine Phoneme gibt. Immerhin werden die Knotenpunkte mit *best matches* aktiviert und stehen somit auch für einzelne Phoneme bereit. Der Unterschied ist aber der, dass die Neuronen (Knoten) ebenso auch Wörter oder Silben repräsentieren können. Das gleiche Neuron beziehungsweise die gleiche Neuronengruppe kann also unterschiedliche Repräsentationen gleichzeitig speichern. In ihrem Aufsatz gehen Goldinger und Azuma spezifischer auf die Resultate von Savin und Bever (1970), Foss und Swinney (1973), McNeill und Lindig (1973), Healy und Cutting (1976) und Norris und Cutler (1988) ein und bringen sie mit ART in Einklang.²⁰

Pisoni und Levi (2007) fassen die Forschung der Repräsentationseinheiten in einem Abriss zusammen. Sie beschreiben eine anfängliche und traditionelle Sichtweise, in der die Wörter durch eine lineare Abfolge von abstrakten Phonemen charakterisiert werden und die sich analytisch verarbeiten lassen. Die Autoren kommen dann auf neuere Ansätze zu sprechen, welche versuchen, den Problemen der traditionellen Perspektive Rechnung zu tragen. Insbesondere stehen dabei die starken akustisch-phonetischen Abweichungen von Sprachsignal und interner Verarbeitung im Vordergrund. Dazu zählen insbesondere die Nichtlinearität des Sprachsignals und die damit verbundenen Probleme bei der Sprachsegmentierung. Die neuen Paradigmen proklamieren für sich, vollständige Informationen zu berücksichtigen und bestreiten eine analytische Kognition. Die episodischen Informationen, welche durch die Sprechsituation gegeben sind, werden mit dem Problem der Variabilität verknüpft. Obwohl die Abstraktheit der Repräsentations-

¹⁸Erste Ergebnisse dieser Forschung lagen schon Anfang der 90er Jahre vor (Goldinger et al., 1991) und auch später mit Goldinger (1997) und Goldinger (1996).

¹⁹Die Wissenschaftler benutzen eine Variante der *Adaptive Resonance Theory* (Grossberg, 1980). Eine verständlichere Darstellung findet sich allerdings in Carpenter und Grossberg (2003).

²⁰Wie sich erst im weiteren Verlauf der Arbeit zeigte, liefert die Simulation sehr ähnliche, wenn nicht sogar völlig übereinstimmende Ergebnisse. Somit kann eine Sympathie mit ART nicht bestritten werden, obwohl es keine Übereinstimmung hinsichtlich der Abstraktheit der Spracheinheiten gibt (siehe auch Kap. 7.2).

4 Randbedingungen der Simulation

einheiten dadurch weitestgehend vermieden wird, betonen die Verfasser nachdrücklich, dass ein hybrider Ansatz die beste Wahl sei; das Ende der abstrakten Einheiten, wie es aus der künstlichen Intelligenzforschung von Brooks (1991) zu hören war, solle also nicht erfolgen.

Zusammenfassend lässt sich für die Ergebnisse der Sprachaufnahme von linguistischen Einheiten bei Erwachsenen sagen, dass sie unstrittig in der Lage sind, phonemische und phonetische Einheiten zu erkennen. Nach Durchsicht der relevanten Literatur mit unterschiedlichen Befunden für und gegen Silben und Phoneme spricht einiges dafür, dass die Frage falsch gestellt ist (Pisoni und Levi (2007); Goldinger und Azuma (2003)). Ebenso scheint es sich tatsächlich um ein sprachspezifisches Problem zu handeln. Als gesichert gilt, dass die Reaktionszeit (*Response Time*, RT) in *monitoring tasks* kein zuverlässiges Maß für die Messung von Silben oder Phonemen sein kann. Dass Silben schneller entdeckt werden als Phoneme oder umgekehrt, ist von Norris und Cutler (1988) widerlegt. Wie schon Healy und Cutting (1976) feststellten, kommt es darauf an wie man das Material präsentiert. Eine gewisse Unklarheit besteht darin, in welcher strukturellen Form die Einheiten zugänglich sind. Deshalb kann man noch nicht eindeutig schlussfolgern, ob Silben oder einzelne phonetische Merkmalsvektoren Verwendung finden. In der Forschung zeigte sich sehr bald eine Tendenz, *Top-down*-Ansätze verstärkt zur Erklärung hinzuzuziehen, weil man einerseits bei der Extraktion aus der Informationsmasse keine ausreichende Antwort finden konnte und andererseits überhaupt zweifelhaft war, ob alle Informationen im Eingangssignal enthalten waren.

Dieser Umstand hat die Erforschung dieses Phänomens auf Kleinkinder und die frühkindliche Entwicklung gelenkt. Hier waren Zugriffe auf ein mentales Lexikon grundsätzlich auszuschließen, weil es keine Hinweise auf ein genetisch oder in sonst einer anderen Weise kodiertes mentales Lexikon gibt. Es ist also plausibel bei einem Neugeborenen ein *Bottom-up*-Vorgehen anzunehmen und damit die überlagernden Effekte des *Top-down*-Ansatzes wie sie bei Erwachsenen Probanden vorkommen, zunächst auszublenden. Das wirft sogleich die Frage auf, wie nun das mentale Lexikon aus dem Sprachinput entwickelt wurde und man ist gezwungen, die Erkennung der Sprachlaute in direkter Abhängigkeit zum Input einzugestehen. Vorgegebene Phonemkategorien²¹ wie sie in der generativen Phonologie beschrieben werden, sind angesichts der Vielzahl an gegenteiligen Experimenten nicht mehr haltbar und werden wohl auch deshalb heute selbst von ihren Verfassern kaum noch in Erwähnung gezogen. Aufgrund der Vielfältigkeit der Sprachmerkmale sind vorgegebene Kategorien schon aus Gründen der Ineffizienz eines Gehirns wenig sinnvoll. Natürlich bilden sich Phonemkategorien, aber nicht, weil sie genetisch als solche vorgegeben wurden, sondern durch das häufige Auftreten eines Umweltreizes, einen Laut einer bestimmten und komplexen Frequenzzusammensetzung.²²

²¹Vorgegebene Phonemkategorien stehen zwar auch im Bezug zum Input, der hier als Auslöser betrachtet wird, aber sie verzichten auf die schwierige Aufgabe der genauen Extraktion der Merkmale des phonetischen Materials, eben weil man glaubt, dass diese Informationen nicht aus dem Input zu erfassen sind.

²²Ein möglicher Mechanismus wäre hier analog zur Populationskodierung der Neuronen zu sehen. Das Gehirn bildet einen statistischen Mittelwert von Schallereignissen in einem für ein bestimmtes Neu-

4.1.2 Wahrnehmung von linguistischen Einheiten bei Säuglingen

Eimas et al. (1971) kann insofern als Anfangspunkt gesehen werden, weil hier erstmals ein Messverfahren Anwendung fand, das auch zur Sprachperzeption für Neugeborene geeignet war.²³ Das Experiment zeigte, dass nur wenige Monate alte Säuglinge (ein bis vier Monate) stimmhafte Laute sehr genau unterscheiden können und damit das Kontinuum des Stimmtoneinsatzes diskret zuordnen, also Kategorien bilden. Im damaligen Experiment wurden [pa] und [ba] als Stimuli verwendet. In den darauf folgenden Jahren wurden diese Fähigkeiten auch für unterschiedliche Artikulationsorte (z.B. Hillenbrand (1983) u. Hillenbrand (1985)) und -arten von Konsonanten (z.B. Jusczyk und Thompson (1978); vgl. Jusczyk (1982, 376)), Stimmtoneinsatz (vgl. Eimas, 1975) sowie für Vokale (Kuhl, 1980), die nicht diskret wahrgenommen werden, bei sechs Monate alten Babys oder jüngeren nachgewiesen.²⁴ Die Interpretation dieser Ergebnisse wurde von Anfang an im Licht der universalgrammatischen Idee geführt.²⁵ Eimas et al. (1971, 306) schrieben:

The implication of these findings is that the means by which the categorical perception of speech, that is, perception in a linguistic mode, is accomplished may well be part of the biological makeup of the organism.

In der Literatur werden bevorzugt auf vier bis fünf weitere Aufsätze (Bertoncini und Mehler (1981); Jusczyk und Derrah (1987); Bertoncini (1988); Bijeljac-Babic et al. (1993)) verwiesen, welche seit der Studie von Eimas et al. (1971) die Silbe als universale Repräsentationseinheit etablieren sollen (Jusczyk (1997, 115); Gerken und Aslin (2005, 6f); Houston et al. (2004, 422ff)). Bertoncini und Mehler (1981) habituierten zwei Monate alte französische²⁶ Babys in vier Gruppen mit eindeutig silbischen Stimuli [tap], konsonantischen Segmentfolgen [tsp] sowie eine Einbettung des konsonantischen Clusters in Vokale [utspu].²⁷ Wie vorhergesagt, wurden in der Testphase die silbischen Strukturen [pat] schneller erkannt als die konsonantischen [pst]. Dies verleitet die Autoren dazu, die Silbe als natürliche Einheit der Lautsegmentierung anzusehen. Dazu bemerken sie abschließend:

We cannot make any definitive conclusion as to the comparative status of the syllable and the phoneme from the research with infants. (Bertoncini und Mehler, 1981, 259)²⁸

In dieser Zeit forschten Werker und Kollegen an einem entwicklungswissenschaftlichen

ronenensemble, deren Zusammengehörigkeit das Gehirn häufigkeitsbedingt erlernt (vgl. Roth und Prinz (1996, 226); Roth (1996, 159) oder Singer (1995, 91ff)).

²³High Amplitude Sucking Procedure (HAS).

²⁴vgl. Houston et al. (2004, 418) als einschlägige Zusammenfassung.

²⁵vgl. auch Eimas (1975, 193).

²⁶Diese wichtige Information wurde in der Methodenbeschreibung des Experimentprotokolls wohl bis Werker und Tees (1984) ausgelassen. Da aber alle Subjekte aus der *Pariser Port Royal Maternity Clinic* stammen, kann man davon ausgehen, dass es sich um Kinder französisch sprechender Mütter handelt.

²⁷Die Kontrollgruppe als vierte Gruppe erhielt das konsonantische Cluster.

²⁸Die VCCCV-Struktur sollte Hinweise auf die Herauslösung der Silben aus dem Lautstrom geben. Dies stellte sich jedoch als problematisch dar, weil Dishabituation gut mit der CVC-Struktur übereinstimmte, die Erkennungszeit jedoch gut drei- bis viermal so hoch war. Zudem kann die VCCCV-Struktur in verschiedene Silben aufgeteilt werden.

4 Randbedingungen der Simulation

Phänomen (Werker et al. (1981); Werker und Tees (1983) u. Werker und Tees (1984)). Wie konnte man schlüssig erklären, dass sechs Monate alte Babys in der Lage waren, Laute zu diskriminieren, die in ihrer Sprache zu einer Kategorie gehörten? Wenn universale Lautkategorien vorhanden sind, müssten sich diese nicht nur in verschiedenen Sprachen finden lassen, sondern in Abhängigkeit von der Sprache auch bei Kleinkindern durch den Stimulus der Zielsprache entsprechend ausgelöst werden. Werker et al. (1981) zeigten, dass erwachsene Sprecher des Englischen die retroflexiven Verschlusslaute ([ʈ],[ʈ]) des Hindi und die aspirierten Verschlusslaute ([tʰ],[dʰ]) erwartungsgemäß nicht von den dentalen Verschlusslauten ihrer Sprache unterscheiden konnten. Erwachsene Sprecher des Hindi hatten damit keine Erkennungsprobleme, da dieser Unterschied im Hindi bedeutungsunterscheidend ist. Kleinkinder, ganz gleich ob Englisch oder Hindi sprechend, verhalten sich jedoch wie erwachsene Sprecher des Hindi. Damit konnte ein „decrease in speech perceptual abilities with age“ (349) vermutet werden.

Diese entwicklungsabhängige und unbewusste Habituation auf bestimmte Lautkategorien, sollte mit Werker und Tees (1983) zeitlich genauer auf das Alter eingegrenzt werden. Vier, acht und zwölf Jahre alte Kinder wurden mit dem gleichen Stimulusmaterial getestet wie in Werker et al. (1981). Überraschenderweise zeigte sich aber keine lineare Verschlechterung der kategorialen Lautwahrnehmung. Acht und zwölf Jahre alte Kinder waren in der Erkennung des Stimmtoneinsatzes der aspirierten Verschlusslaute deutlich besser als die Vierjährigen. Die genaue zeitliche Eingrenzung gelang schließlich ein Jahr später (Werker und Tees, 1983). Das erste Experiment replizierte die Ergebnisse der Vorgängerstudie (Werker et al., 1981) mit dem Salischen, einer nordamerikanischen Indianersprache.²⁹ Das zweite Experiment grenzte dann die Zeit bei Kleinkindern sehr genau ein. Acht bis zehn Monate alte Kinder besitzen noch die Fähigkeit, innerhalb der phonemischen Kategorie ihrer Sprache einzelne Laute zu diskriminieren. Bereits einen Monat später nimmt diese Fähigkeit sehr stark ab und ist in einem weiteren Monat nicht mehr existent. Diese Untersuchungen wurden auf eine Reihe weiterer Konsonanten ausgeweitet mit gleichem Ergebnis (Werker und Lalonde, 1988). Interessant ist aber in diesem Zusammenhang, dass Zulu-Klicks (⊙) bei zwölf bis vierzehn Monate alten Babys genauso gut erkannt werden wie bei erwachsenen Sprechern des Englischen. Es scheint also auch darauf anzukommen, wie viele und welche phonetischen Eigenschaften im phonologischen System vorhanden sind (Best et al., 1988).³⁰

Für die Frage der kategorialen Wahrnehmung gelang Maye et al. (2002) ein entscheidendes Experiment. Die Wissenschaftler zeigten, dass in Abhängigkeit von der Häufigkeit des dargebotenen Stimulus, [d]-[t] als bimodale oder einen Laut dazwischen als unimodale Verteilung, sechs und acht Monate alte Babys immer genau das habituierte Muster bevorzugten. Die Gruppe der unimodalen Verteilung zeigte keine Präferenzen für die bimodale Verteilung, obwohl diese ja vorher sicherlich durch die Sprachumwelt des Kindes bereits gelernt sein musste. Mit diesem Experiment konnte sehr klar bewiesen werden, dass die kategoriale Zuordnung von Lauten auf eine statistische Verteilung zurückgeht.

²⁹Hier sollten die glotalisierten Velare /kʰ/ und Uvulare /qʰ/ unterschieden werden.

³⁰Etwas später wurde diese Eigenschaft auch für jüngere Babys (sechs bis zwölf Monate) nachgewiesen (Best (1991) in Houston (2005, 419)).

4.1 Wahrnehmung von sprachlichen Einheiten

Jusczyk und Derrah (1987) merkten an, dass die herkömmliche Art der Messung bei Kleinkindern nur nominale Aussagen erlaubte. So könne man nur herausfinden, ob ein Kleinkind ein bestimmtes Lautmuster wahrnimmt oder nicht. Wichtig wäre aber auch zu wissen, wie viel mehr oder um wie viel besser das eine Lautmuster im Gegensatz zu einem anderen wahrgenommen würde. Ob die Methodik tatsächlich ordinale Unterschiede misst, ist nach wie vor strittig. Für die Fragestellung der linguistischen Einheiten ist sie aber auf jeden Fall wertvoll.

Zwei Monate alte Babys wurden in vier Gruppen³¹ geteilt und mit Silbensequenzen der Form [bi], [bo], [bər], [ba] konditioniert. In der Testphase erhielten die ersten drei Gruppen jeweils eine andere Silbe [bu], [du], [da] für [ba] im vorhandenen Material. Die vierte Gruppe erhielt [di], [do], [dər] und [da] als neues Stimulusset. Nach der logischen Vorhersage sollte die letztgenannte [du]-Gruppe höhere Saugraten haben als die [bu]-Gruppe, weil hier nicht nur der Vokal [u], sondern auch ein neuer Konsonant [d] dazu gekommen ist. Die [da]-Gruppe hingegen hat nur den Anfangskonsonanten geändert, sollte sich also ähnlich der [bu]-Gruppe verhalten. Letztlich müsste die vierte Gruppe die höchsten Saugraten aufweisen. Die Ergebnisse zeigten, dass der Unterschied zwischen den ersten beiden Gruppen signifikant war. Die [da]-Gruppe verhielt sich aber wie die [du]-Gruppe, sodass man hier den Wechsel des Konsonanten als das entscheidende Kriterium betrachten muss. Die letzte Gruppe zeigte die Ergebnisse entsprechend der Vorhersage. Die Ergebnisse kann man natürlich unterschiedlich deuten. Die Verfasser kommentieren das Ergebnis wie folgt:

These results demonstrate not only do 2-month old infants have the ability to discriminate speech contrasts, but they also possess some capacity for representing speech sounds. . . . they encoded not only the fact that the syllables differed from one another but also something about the way in which they differed from each other. (Jusczyk und Derrah, 1987, 652)

Hinsichtlich dieses Experiments drängt sich die Frage auf, welche Implikationen die Ergebnisse auf die Repräsentationsform der linguistischen Einheit haben könnte.

Because these findings can be conveniently described in familiar linguistic terms that refer to properties of speech such as phonetic segments or phonetic features, it is tempting to assume that infants' responses to speech are based on such units. (Jusczyk und Derrah, 1987, 653)

Dieser Gedanke liegt sehr nahe, aber Informationen über die Struktur und Organisation dieser Einheiten lassen sich daraus nicht eindeutig schließen.

Die gleiche Methodik wurde in Bertocini (1988) verwendet und mit Neugeborenen und wiederum zwei Monate alten Babys durchgeführt.³² Wiederum sollten die phonetischen Unterschiede innerhalb der Silbe mit der *High-Sucking-Procedure* untersucht werden. Für die älteren Säuglinge konnte natürlich das gleiche Ergebnis nachgewiesen werden wie auch schon in Jusczyk und Derrah (1987). Es ließ sich allerdings auch zeigen, dass diese Fähigkeit bei den Neugeborenen noch nicht vorhanden war, wohl aber eine

³¹und Kontrollgruppen

³²Leider wurde in der Experimentbeschreibung keine Angabe zur Sprache der Mutter gemacht.

4 Randbedingungen der Simulation

Erkennung der Silben. Somit können die Autoren eine Entwicklungsrichtung skizzieren, welche von der Silbe als übergeordnete Einheit hin zu phonetischen Segmenten verläuft.³³

Such an account fits well with the view that young infants representations of syllables are holistic or undifferentiated with respect to phonetic segments. (Bertoncini, 1988, 30)

There are strong indications in the present study that the infants representations are developing even during the first 2 months of life. Thus, the newborns give no evidence of distinguishing a new token that differs only in its initial consonant from the other syllables in the preshift set. (Bertoncini, 1988, 31)

Dieser Befund provozierte nun schon fast ein Folgeexperiment, welches aber erst einige Jahre später gelang. Bijeljic-Babic et al. (1993) testeten vier Tage alte französische Babys auf ihre Unterscheidungsfähigkeiten. Können die wenige Tage alten Versuchspersonen mehrsilbige Äußerungen aufgrund der Anzahl der Silben oder Phoneme unterscheiden? Das erste Experiment untersuchte, ob die Säuglinge CV-Doppel- von CV-Dreierkombinationen und umgekehrt unterscheiden können. Das positive Resultat führte sodann zur zweiten Untersuchung. Hier konnte die Zeit als mögliche Hilfe ausgeschlossen werden. Die Logik des dritten Experiments war jedoch die, dass die Säuglinge, wenn sie denn phonemische Repräsentationen hätten, unterschiedliche Phoneme innerhalb gleicher Silbenstrukturen auch bemerken müssten, ähnlich, wie sie im ersten Experiment unterschiedliche Silbenkombinationen unterscheiden. Hier konnte natürlich kein positiver Nachweis erbracht werden und so sind diese Experimente als der Beweis für eine angeborene Silbenstruktur in den kleinen Kreis jener Wissenschaftler eingegangen. Dabei sind die Annahmen des dritten Experiments kritikwürdig. Wie Jusczyk (1997, 112) anmerkt, besteht eine große Versuchung, die Ergebnisse aus den Erkennungsauf-

³³Walley et al. (1986) hat diese Entwicklung für Kinder des Vorschulalters (fünf Jahre) und der 2. Klassenstufe (sieben Jahre) beobachtet. Dabei war es aber unwichtig, ob es sich um ein Phonem oder eine Silbe handelte.

The developmental trend is not simply from correct classifications given a single corresponding phoneme. Rather, our results suggest that position, as well as the number, of shared constituents (either phonemes or syllables) matter. There appears to be a developmental trend whereby (1) attention to the ends of sound and (2) attention to final correspondences is stronger than attention to whole syllables. In brief, in the present task at least, it did not matter greatly whether corresponding phonemes formed the syllable unit or not, but it did matter where the correspondences were located. (Walley et al., 1986, 227)

Sie führen dann weiter aus:

Syllables may have no special status at the level at which the perceptual similarity of two speech sounds is calculated. (Walley et al., 1986, 227)

und beenden die Diskussion:

But the syllable is not always the organizing unit for children and it is not always more controlling of performance than is the phoneme. We speculate that syllables may be special at the level of consciousness, but that at the level of perceptual similarity, the level at which a sound contacts a representation in memory, the syllable may play no special role. (Walley et al., 1986, 228)

gaben über minimale Unterschiede von phonetischen Segmenten bei Säuglingen, falsch zu interpretieren. Erstens ist überhaupt nicht klar, dass derartige Strukturen überhaupt angeboren sein müssen. Sie können sich ebenso häufigkeitsbedingt aus den Umweltreizen erlernen lassen. Zweitens ist das dritte Experiment ungleich schwieriger. Das Erkennen der feinen phonetischen Unterschiede im Stimulusmaterial des dritten Experiments kann nicht mit der Unterscheidungsleistung zwischen zweier-CV-Kombinationen von dreier-CV-Kombinationen gleichgesetzt werden. Drittens wurden ausschließlich französische Säuglinge getestet, die natürlich den Rhythmus dieser Sprache schon vorgeburtlich erlernt haben.³⁴

Gegen die letztgenannte Kritik versuchte Bertoncini (1995), einen Einwand einzubringen. Französische Neugeborene (bis fünf Tage) wurden in zwei Experimenten getestet. Die Vorgehensweise vollzog sich analog zum Experiment von Bijeljac-Babic et al. (1993). Das Stimulusmaterial des ersten Experiments bestand aus zwei Listen mit zweisilbigen japanischen Wörtern der Form VCV und CVCV, bei denen das Silbenende immer genau auch mit der Mora zusammenfiel. Jede Silbe entsprach also genau auch einer Morastruktur. Die Ergebnisse stehen in vollen Einklang mit Bijeljac-Babic et al. (1993). Französische Neugeborene silbifizieren also auch japanische Wörter. Die Listen des zweiten Experiments enthielten japanische Wörter mit zwei und drei Moras, eingebettet in zweisilbigen Wörtern, deren Strukturen jedoch nicht kollaborieren. Für diese Listen wurden keine erhöhten Saugraten gemessen. Die Babys erkannten die Mora also nicht als Segmentierungseinheit.

Ohne entsprechende Daten für japanische Probanden lassen die Ergebnisse keinen eindeutigen Schluss zu. Man kann, wie Bertoncini (1995), argumentieren, dass die Subjekte, nicht die Mora als Segmentierungseinheit nutzen, sondern das Silbenmuster, in welchem die Mora-Kombinationen eingebettet sind. Andererseits könnte hier eine ähnliche Überraschung wie bei Cutler et al. (1986) für die englischen und französischen Erwachsenen folgen, auch wenn es sich im Bertoncini-Experiment um Neugeborene handelt. Es ist doch ebenso möglich, zu sagen, dass die französischen Babys die Silbenstruktur im zweiten Experiment nicht erfasst haben, weil konkurrierende Strukturen, die Moras, präsent waren. Immerhin verwenden erwachsene Japaner auch die Mora als Segmentierungseinheit (Cutler, 1994).

Das dritte Experiment von Bertoncini (1988) unterstellt, dass Neugeborene bereits eine gewisse Vorstellung von einzelnen Phonemen angelegt haben. Eine phonemische Repräsentation erfordert immer auch eine Kategorisierung³⁵ eines bestimmten, aber auch

³⁴Wie bereits in Kapitel 4.1.1 ausgeführt wurde, lassen sich bei erwachsenen Sprechern des Englischen und Französischen die rhythmischen Eigenschaften der Sprache eindeutig auf die besondere Silbenstruktur des Französischen zurückführen. Französische Sprecher segmentieren das Englische in silbenähnliche Strukturen, während dies bei englischen Sprechern weder in ihrer Muttersprache noch im Französischen zu messen ist (Cutler et al., 1986).

³⁵Für die Definition der Simulationsannahmen dieser Arbeit ist es unwichtig, wie genau eine phonemische Kategorie entsteht. Das bedeutet, dass es möglicherweise eine Art Prototyp eines Lauts oder einen *magnet effect* (Kuhl, 1991) gibt oder dass die Grenzen zwischen den Kategorien, die Eigenschaften eines Lauts (Liberman et al., 1967) festlegen. Für diese Arbeit ist demgegenüber nur wichtig, dass

4 Randbedingungen der Simulation

variablen Frequenzspektrums (Pierrehumbert, 2003). Caplan et al. (1999) glauben, dass die Fähigkeit zur Erkennung von Konsonanten und Vokalen bereits mit der Geburt vorhanden ist (vgl. Caplan et al., 1999, 1494). Damit können die Autoren natürlich nicht spezifische Laute in einem Jakobsonschen (Jakobson, 1972) oder Libermanschen (Liberman und Mattingly, 1989) Sinne gemeint haben, sondern wohl eher die Voreinstellungen zur Aufnahme und Anpassung der Frequenzspektren. Eine anderweitige Auffassung ist nach heutigem Wissensstand nicht mehr ernsthaft zu vertreten (vgl. Werker (1995); Hickok und Poeppel (2004, 72f)). Die auditive Entwicklung, welche die einzelsprachlichen Besonderheiten der Phonemkategorien definiert, muss sich also erst noch vollziehen. Die kortikale Entwicklung zur aditiven Informationsverarbeitung ist mit dem vierten Monat in vollem Gang (Moore, 2002). Die Verbindung von der Gehörnde und anderen neurologischen Strukturen ändert sich entscheidend erst von da an mit zunehmendem Alter (Huttenlocher und Dabholkar (1997); Albert et al. (1999); Hecox (1975)). Man kann also davon ausgehen, dass Neugeborene sicherlich in der Lage sind, Unterschiede bei Lauten festzustellen, dass aber ihre phonetische Sensibilität noch nicht soweit gereift ist, unveränderbare Kategorien anzulegen. Diesen Hintergrund sollte man bei der Beurteilung der vorstehenden Experimente berücksichtigen.

Auf eine kategoriale Wahrnehmung weisen die überwiegende Zahl der experimentellen Befunde hin (vgl. Jusczyk (1997, 46ff); Jusczyk et al. (1998)). Houston et al. (2004) weisen auf die erstaunlichen Fähigkeiten von wenige Monate alten Babys hin, die sie zur Kategorisierung von Vokalen und Konsonanten sehr schnell entwickeln. Miller und Eimas (1983) legten eine Studie vor, indem sie zeigen konnten, dass die bereits gebildeten Kategorien von drei bis vier Monate alten Babys, denen von Erwachsenen in hohem Maße ähneln. Gerade in der Frühphase der Sprachperzeption geht es aber vielmehr um die Unterschiede als um die spezifischen Vergleiche zwischen phonetischen Eigenschaften.

In order to have access to contrasts, rather than sounds, infants must have some way of knowing that speech segments vary along underlying perceptual continua. To do that, infants must extract and represent detailed segmental information as opposed to syllabic or other high-order information. (Houston et al., 2004, 423)

Es muss also eine erste rudimentäre Vorstellung von segmentähnlichen Einheiten vorhanden sein. Diese könnte auch kontinuierlich sein und durch den Vergleich von unterschiedlichen Inputdaten, analog zu den Stimmtönen eines Instruments, eigene Kategorien bilden. Man muss aber davon ausgehen, dass sich deren Form grundlegend von den Repräsentationen der schon einige Monate älteren Babys unterscheidet (vgl. Kuhl et al., 2005). Möglicherweise werden, einem neuronalen Netzwerk gleich, zunächst alle Inputdaten häufigkeitsabhängig nach Frequenzen geordnet, wodurch sich schließlich Kategorien mit gewissen Toleranzbereichen bilden (Kluender (1994, 201ff u. 205); vgl. Elman (1989); Pierrehumbert (2003)). Die Beobachtung des Forschers ist dann, die folgende:

Infants are able to discriminate nearly every phonemic contrast on which they have been

die Laute innerhalb eines Prototyps beziehungsweise einer Kategorie als gleich und zwischen den Kategorien als unterschiedlich wahrgenommen werden können, ganz gleich nach welchen Prozessen dies abläuft.

4.1 Wahrnehmung von sprachlichen Einheiten

tested, but they typically fail to discriminate equal-sized acoustic differences that specify two variations of the same phoneme. (Werker, 1995, 88)

Nachdem Trehub (1973) in ihren Experimenten zeigen konnte, dass [a]-[i] und [i]-[u] unterschieden werden, testete Kuhl (1980) in einer modifizierten Variante der *Head-Turn-Preference-Technik* fünfeinhalb bis sechseinhalb Monate alte Babys hinsichtlich ihrer Fähigkeit, sprachliche Kategorien zwischen Vokalverläufen zu bilden. Im Ergebnis fasst sie zusammen:

The data do, in fact, demonstrate that infants are capable of forming categories that correspond to the perceptual categories defined by adult listeners of English. (Kuhl, 1980, 59)

Mit ungefähr sechs Monaten sind Säuglinge also im Stande, die Vokale kategorial zu unterscheiden. Bereits mit dem siebten Monat entwickelt sich die Wahrnehmung für muttersprachliche Laute sensibler als die Wahrnehmung von fremdsprachlichen Lauten (Kuhl et al., 2005, 251). Mindestens ab dieser Zeit muss sich der Sprachlerner über grundlegende phonetische Merkmale im Klaren sein und sie von einander abgrenzen können.

Jusczyk und Aslin (1995) berichten über drei Experimente mit siebeneinhalb und sechs Monate alten Babys, bei denen die älteren Probanden sehr detaillierte phonetische Informationen erkennen konnten (wie sie in *cup* und *tup* kontrastiert werden). Wie auch später noch in Jusczyk (1999), sieht der Wissenschaftler darin einen Hinweis für sehr detaillierte Repräsentationen und nicht nur für das bloße Erkennen offensichtlicher akustischer Eigenschaften. Diese scheinen sich nun kontinuierlich weiter zu verfeinern. Bei zehneinhalb jährigen Kindern sind die Repräsentationen als phonetische Einheiten dann schon gut nachweisbar.

Swingley und Aslin (2000) manipulierten Wörter in einem Segment, sodass sich ein Nichtwort ähnlich einem Aussprachefehler bildete, und betteten diese Einheiten in Sätze. Beim Vorspielen dieser Sequenzen wurden zwei Bilder gezeigt, eines davon der Referent zu dem Satz. Unabhängig vom Bild wurden die Sequenzen erkannt, wenn auch schlechter bei den Sätzen mit Aussprachefehlern. So enthüllt sich abermals eine recht hohe Detailtreue für segmentale Repräsentationen, wobei weiterhin unklar bleibt, ob diese überhaupt notwendig ist. Mit den Worten der Autoren lässt sich der Beitrag der Studie aber doch erkennen.

However, without more conclusive evidence that phonemic specification is necessary to account for word recognition, it would be premature to consider the present results as evidence for (or against) phonemic specification in 1-year-olds' lexical representations. We may say only that our results contradict any view holding both that (a) segmental representations are criteria for the differentiation of minimal pairs, and that (b) children of 18-23 month do not represent speech in terms of segments. (Swingley und Aslin, 2000, 161)

Als Zusammenfassung ergibt sich ein ähnliches Fazit wie bei der allgemeinen Betrachtung der Wahrnehmung linguistischer Einheiten (Kap. 4.1.1). Eine universale Wahrnehmung von Silben oder Segmenten kann man im Lichte der angestellten Exegese verneinen. Die verwendeten Repräsentationseinheiten richten sich nach sprachabhängigen

4 Randbedingungen der Simulation

Einflüssen. Im Alter zwischen sechs und acht Monaten verfügen die Sprachlerner über das notwendige phonetische Wissen, zumindest einfache phonemische Kategorien zu bilden und diese auch zu erkennen. Auf diese Weise verstärkt sich die Tendenz, das Phonem als Perzeptionseinheit für die Simulation zu verwenden.

4.1.3 Kritische Betrachtung der experimentellen Befunde

Bislang wurde die Forschung der vergangenen 30 Jahre über die kleinsten linguistischen Einheiten bei Erwachsenen und Säuglingen resümierend vorgestellt. Dieses Wissen kann nunmehr mit weiteren Erkenntnissen aus neurobiologischer und selbstorganisatorischer Sicht ergänzt, schließlich in einem neuen Licht betrachtet werden. Dies geschieht natürlich immer in Hinblick auf das Segmentierungsproblem.

Ein erster Gedanke sei zur Idee einer universal, wohlmöglich angeborenen Fähigkeit zur Erkennung von Silben und silbischen Strukturen angestellt, wie in den Experimenten von Bijeljic-Babic et al. (1993) versucht wurde nachzuweisen.³⁶ Gómez (2007, 604) legt dar, dass Sprachen sehr stark in ihrer Silbenstruktur variieren. Im Englischen sind beispielsweise V, CV, CVC, CCCVCCC und dazwischen liegende Kombinationen möglich. Diese Variationen müssen allerdings gelernt werden. Ein experimenteller Befund dazu wurde von Saffran und Thiessen (2003) erbracht. Eine kurze Konditionierung mit der entsprechenden Silbenstruktur ist ausreichend, um bei neun Monate alten Babys einen messbaren Lerneffekt herbeizuführen.

In diesem Zusammenhang stellt sich die Frage, warum man nun von der angeborenen Fähigkeit, eine CV-Silbe zu erkennen, ausgehen soll, wenn diese auch sehr schnell erlernt werden kann? Basieren die Fähigkeiten im neunten Monat auf der anfänglichen CV-Silbe oder ist sie nur eine willkommene, aber nicht gut begründete, Lösung zum Segmentierungsproblem? Eine mögliche Antwort versucht Blevins (1995, 217) zu geben. Sie listet die Silbentypen in den Sprachen der Welt auf. Dabei bemerkt sie, dass diese Typen in höchst unterschiedlichen Mustern vorkommen, die CV-Silbe aber überall enthalten ist. Auch Macken (1995, 689) und Bertoncini und Mehler (1981, 249) stellen die CV-Silbe als universell heraus. Allerdings geben sie keine Lösung für Kunjen, eine Sprache der australischen Ureinwohner.³⁷ Diese Sprache erlaubt nur Silben mit anfänglichen Vokalen (Sommer, 1981). Die CV-Silbe könnte demnach nicht für das Kunjen zur Seg-

³⁶Ein genereller Kritikpunkt an diesem Experiment ist, dass ausgerechnet nur französische Neugeborene getestet wurden. Wie in den Ausführungen des ersten Kapitels 1.2 bereits benannt, ist es weitestgehend unstrittig, dass die prosodischen Eigenschaften einer Sprache auch schon im Mutterleib erkannt werden können. Interessant wäre es, das Experiment mit Neugeborenen anderer Sprachen durchzuführen, welche nicht, wie das Französische, eine eindeutige silbenstrukturierte Rhythmik aufweisen. Hier wäre nicht nur das Englische ein möglicher Kandidat, sondern vielleicht auch das Tschechische, welches auch Konsonanten als Silbengipfel in betonten Silben erlaubt, wie dem Beispiel: *'strtf 'rst 'skrs 'krk*. (Steck den Finger durch den Hals) zu entnehmen ist. Man bedenke ebenso die Silbifizierung von Tonsprachen wie dem Chinesischen (Zhang, 1996).

³⁷Blevins (1995, 230) diskutiert dieses Phänomen, wenn auch kritisch und unter Verweis auf fünf alternative Modelle einer internen Silbenstruktur (212), doch ohne Erklärung hinsichtlich der Universalität der CV-Silbe.

mentierung genutzt werden. Eine Theorie, die vorgibt, dass die anfängliche CV-Silbe in eine VC-Struktur überführt werde, klingt wenig überzeugend, sondern eher nach einem fehlenden Mechanismus, der erklärt, wie CV-Silben allgemein gelernt werden können.

Selbst wenn man eine universale CV-Silbe anerkennt, kann die Problematik der Wortsegmentierung nicht annähernd gelöst werden. Folgendes Beispiel kann dies verdeutlichen. Die Aussage *Where is it? Ohh there it is?* würde sich in folgendes Silbenmuster zerlegen lassen: CVC VC VC V CVC VC VC (*wer ɪz ɪt əv ðeɪt ɪz*). Bei Beseitigung der Lücken beziehungsweise Pausen erhält man: CVCVCVCVCVCVCVC. Dies entspricht auf abstrakter Ebene der Repräsentation des Signalstroms, den der Sprachenlerner hört. Wendet das Kind nun, losgelöst von anderen prosodischen Informationen, sein universales Silbenmuster an, resultieren folgende Silben: CV CV CV CV CV CV CV C. Diese Repräsentation bedeutet wiederum, dass die Repräsentationseinheiten mit denen eine Statistik letztlich erstellt wird, folgende Ausprägungen haben: *we rɪ zɪ təv ðe rɪ tɪ z*. Die Silben reichen über Wortgrenzen hinweg und verhindern so eine spätere Segmentierung.

In der Tat betrachtet man die silbische Struktur natürlich nicht separat von Suprasegmentalia. Die Verbindung von silbischer Struktur und Suprasegmentalia gibt aber für die Wortsegmentierung als solche zunächst keine zusätzliche Information. Die notwendige Information ist erst dann gegeben, wenn einige Wortkandidaten als separate Einheiten erkannt werden. Erst dann können die Grenzen des Wortes mit der silbischen Struktur und den Suprasegmentalia in Zusammenhang gebracht werden.

Das Beispiel zeigt deutlich, dass selbst die Annahme, eine universale CV-Silbe zu verwenden, keine Lösung sein kann. Die Verwendung einer universalen CV-Silbe würde sogar eine spätere Wortsegmentierung erschweren. Oft vorkommende Wörter können dann auch nicht mit den Übergangswahrscheinlichkeiten lokalisiert werden. Unterstellt man nun weiter, dass auch andere komplexere Silbenmuster auftreten können, bleibt weiterhin die Frage offen, wann eine CVC und wann eine CV oder VC-Struktur in einem Lautstrom anzuwenden sei. Auch hierfür braucht man zumindest eine Reihe von Prototypen, an denen man eine Hypothese über eine Regel aufstellen kann. Eine derart komplexe Regel wie sie Gambell und Yang (2005) aufstellen, unterstreicht ihren Standpunkt, dass, ganz gleich wie das Resultat der Berechnungen auch ausgehen mag, die anfängliche Silbenstruktur, müsse doch in einem *Language Device* kodiert sein, stimmen wird. Von dieser Denkweise muss man sich jedoch trennen, wenn man zunächst einmal herausfinden will, wie viel Information aus dem Sprachsignal, denn ohne angeborenes Wissen zu gewinnen ist.³⁸

³⁸Gambell und Yang (2005) benutzen die Silbe als grundlegende linguistische Einheit in ihrer Simulation zur Wortsegmentierung. Selbst, wenn man eine Silbenstruktur als perzeptuelle Einheit anerkennt, ist es doch fraglich, ob der Sprachinput des hier verwendeten Korpus so segmentiert werden kann. Hier wird (offensichtlich wohl im Geiste der universalgrammatischen Idee) das Prinzip *Maximize Onset* (Gambell und Yang, 2005, 10) angeführt, welches unter einer speziellen Regel des Englischen den Text in eine nahezu perfekte Silbenstruktur verwandelt. Hier liegen neben den speziellen prosodischen noch phonotaktische Regeln zu Grunde, die als sprachspezifische Regeln erst aus einer Reihe von isolierten Ereignissen gelernt werden müssten. Man erkennt also das gleiche Zirkelschlussproblem, welches durch die Anwendung der bedingten Wahrscheinlichkeiten ja gerade vermieden werden sollte. Gerade dies ist natürlich auch das Argument gegen den statistischen Ansatz, welches schon zuvor von Yang (2004) angeführt wurde.

4 Randbedingungen der Simulation

Geht man im Gegensatz dazu von allgemeinen kognitiven Fähigkeiten zur Erkennung von derartigen, womöglich silbischen Strukturen aus, kann man logisch richtig schließen, dass die bedingten Wahrscheinlichkeiten Voraussetzung für alle anderen Wortsegmentierungsstrategien sein müssen.³⁹ Alle anderen setzen ja immer schon eine real nicht existente Zahl von Wörtern voraus, aus denen eine Regel (Phonotaktik, Prosodien, Allophone) deduziert wird. Unterstellt man, dass der Sprachlerner eine silbische Struktur bereits zur Verfügung hat, muss ja schließlich die gleiche Logik Anwendung finden, die auch zur Prejorisierung der bedingten Wahrscheinlichkeiten geführt hat, nämlich: Wie konnte diese Silbenstruktur entstehen ohne, dass es ein Datenmaterial gibt, nach dem eine Silbenbildung expliziert werden kann?

Wo liegt eine alternative Lösung? Ein sehr viel einfacheres Modell als die komplizierten Silbenstrukturen in unterschiedlichen Sprachen in einer universalen Theorie zu vereinen, ist eine zwei dimensionale Merkmalsmatrix wie sie von Clements und Hume (1995) vorgeschlagen wird. Dies setzt voraus, dass bisherige segmentale Aufteilungen zu Gunsten von Merkmalsreihen ohne eine innere Struktur aufgegeben werden (vgl. Clements und Hume, 1995, 246). In der hier beschriebenen Matrix definieren sich Phoneme nach einer Reihe zutreffender auditiver Merkmale.⁴⁰ Aus neurowissenschaftlicher Perspektive bestätigen Caplan et al. (1999, 1496) diese Sichtweise:

Phonological Theory holds that the 'underlying' distinctive features and phonemes that phonologists assign to words are not necessarily the phonemes that are produced and heard. Theories of phonology assume that the abstract representation of the phonemes of a word conveys just enough information about distinctive features to allow each phoneme to be uniquely determined by the application of the general phonological rules of the language.

Die Autoren argumentieren also auch hier dafür, das binäre System der distinctiven Merkmale als Instrument der Beschreibung der realen Vorgänge abzuschaffen. Als theoretische Vereinfachung zur Erkennung von Regelmäßigkeiten in der sprachlichen Struktur könne sie weiterhin jedoch nützlich sein (vgl. Caplan et al., 1999, 1494). Folgt man

First, how do infants know to pay attention to syllables? ... Second, where do the syllables come from? (Yang, 2004, 452).

Allerdings bezieht er sich bei seiner Plausibilitätsbegründung (452) lediglich auf Bijeljac-Babic et al. (1993) als Quelle für eine angebohrte Silbenstruktur. Er lässt aber außer Acht, dass das Französische als *syllable-timed language* eine andere Grundlage bietet als das Englische, welches als *stress-timed* eingeordnet wird. Mit einer einfachen Silbenstruktur der Form Konsonant-Vokal oder Konsonant-Vokal-Konsonant, wie sie in *syllable detection tasks* experimentell nachgewiesen wurden, kann ein englischer Text nicht zufrieden stellend in Silben aufgeteilt werden.

³⁹Lalonde und Werker (1995) ziehen eine entwicklungspsychologische Parallele zwischen der Sprachaufnahme und anderen kognitiven Fähigkeiten des Kindes. Das Ergebnis hier zeigt, dass die Entwicklungsstufen der Sprachaufnahme mit anderen Entwicklungsmechanismen zusammentreffen. Dadurch kann man auch die Möglichkeit einer gegenseitigen Beeinflussung von kognitiven Fähigkeiten und Sprachaufnahme nicht ausschließen.

⁴⁰Hinweise, die eine derartige Lösung sinnvoll erscheinen lassen, finden sich bei Levitt et al. (1988) für die Änderung von labiodentalen und interdentalen Frikativen, bei Kuhl (1980) und bei Darwin (1984) für die Stimmhaftigkeit von Vokalen, den Artikulationsort und -art. Sie zeigen, dass die Lautumgebung entscheidenden Einfluss auf die Erkennung von unterschiedlichen Lautsegmenten hat (vgl. auch Anderson (1992, 43); Pierrehumbert (1990); Rischel (1990, 402).)

dieser Herangehensweise kann man die in einer Sprache vorkommenden Phoneme mit allgemeinen Lernprinzipien der Habituation und Konditionierung stimmig erklären.

Die folgende Überlegung dient dem Verständnis der eben gemachten Aussage. Die Gesamtheit aller Phone der Sprachen dieser Welt sei durch einen Merkmalsvektor eindeutig darstellbar. Die Eigenschaften des Vektors werden begrenzt durch:

- a die physikalischen Eigenschaften des akustischen Signals und
- b die rein physischen Eigenschaften der menschlichen Sinnesorgane (Hör- und Sprechapparat).

Die Informationsmenge, welche sich aus der Schnittmenge dieser beiden Eigenschaftsmengen ergibt, definiert die einzelnen Merkmale des Vektors. Im Falle der natürlichen Sprache weisen die Signale, die abgegeben und aufgenommen werden, die gleichen Eigenschaften auf. Die Mengen sind damit gleich.⁴¹

Die Anzahl der phonetischen Merkmale des Vektors und deren Qualität sind äußerlich vom Signal gegeben und werden weiter eingegrenzt (innerlich) durch die Eigenschaften, welche die Natur den Messinstrumenten (Sinnesorgane) angedeihen ließ. So mag die Schwingungsbreite des Trommelfelles einen bestimmten Frequenzbereich zuverlässig aufnehmen. Die Härchen lassen ebenso nur eine bestimmte Informationsdichte als Übertragung zu. Unsere Stimmbänder sind auf gleiche Weise zur Produktion in diesem Frequenzbereich vorgesehen. Theoretisch mögen zwar mehr Informationen im Sprachsignal vorhanden sein (beispielsweise Eigenschaften des zwischen Hörer und Sprecher gelegenen Mediums oder unbeabsichtigte Wellenüberlagerungen (Interferenzen); diese können aber aufgrund der Physiologie des Menschen nicht weiter genutzt werden.

Welche Kombinationen dieser Merkmale als diskrete Phoneme in das Lautinventar einer Sprache eingehen, wird im Verlauf der Sprachrevolution bestimmt und wird letztendlich aus der Sprachumwelt (z.B. Mutter) an das lernende Kind weitergegeben. Das zu Grunde liegende Prinzip ist die Häufigkeit der Verwendung bestimmter Merkmalskombinationen (Phoneme) mit genau den Merkmalsausprägungen (phonetische Merkmale), die in einer bestimmten Sprache vorherrschen. „[M]ost neurons in the auditory system ... are frequency selective“ (Carlyon (2004, 466); siehe auch Kuhl et al. (2005, 240 u. 251); Gummer und Zenner (1996)). Damit kann man vernünftigerweise annehmen, dass sich so synaptische Verbindungen für genau die Stimuli herausbilden, die von der Sprachumwelt häufiger gegeben werden (Breitenstein und Knecht, 2003, 139f).⁴² Das Kind

⁴¹Dies mag eine Annäherung sein – eine Diskussion, ob man mehr physikalische Informationen mit dem Signal erzeugen als man selbst aufnehmen kann, liegt außerhalb der bewussten Wahrnehmung. Wenn auch mit präzisen Geräten messbar, bringt uns diese Diskussion an die Randbereiche dieses Forschungsbereichs und kann damit an die Philosophie delegiert werden. (siehe hierzu beispielsweise auch Soltis (2004) und anschließende Diskussionen 459ff, interessant ist auch der Aufsatz von Kemp (1978)). In dieser Arbeit können wir davon ausgehen, dass die beiden Mengen identisch sind.

⁴²Also anders als Eimas (1975) annahm, basieren die neuralen Mechanismen nicht auf erblichen Strukturen, welche die möglichen Sprachlaute schon genau spezifizieren, sondern darauf, dass bestimmte Neuronen von bestimmten Frequenzen und andere Neuronen von weiteren Eigenschaften eines bestimmten Signals (in diesem Fall Sprachsignal) angeregt werden. Erbliche Voreinstellungen beschränken sich nach dieser Theorie auf die Frage, warum die Neuronen auf genau die Eigenschaften

4 Randbedingungen der Simulation

wird also für ganz bestimmte Frequenzbereiche eines Signals sensibilisiert und für andere nicht (habituiert).⁴³ Brown (1999, 799) spricht in diesem Zusammenhang bei der auditiven Wahrnehmung von *monotopic mapping*, welches die Repräsentation von lautlichen Frequenzen zu einem Neuronenmuster beschreibt. Tritt eine Eigenschaft eines Signals mit bestimmten Frequenzeigenschaften häufiger auf, werden sich (als Funktion genau dieser Häufigkeit) die Dendriten der Neuronenpaare, welche für diese Eigenschaft zuständig sind, stärker synaptisch verbinden als für Signalmerkmale, die nicht mit dieser Häufigkeit auftreten. Der dadurch erzielte Verstärkungseffekt führt zu einem schnelleren Feuern der betroffenen Neurone und damit zur Sonderstellung der Sprachlaute in der akustischen Umwelt des Menschen. Auf diese Weise gruppieren sich einzelne Merkmale, die auf Dauer und im Durchschnitt oft zusammen auftreten, zu einem Informationspool oder im linguistischen Sinne zu einer Phonemeinheit zusammen. Da die Häufigkeit die Bildung von synaptischen Verbindungen bestimmt, werden auch immer Toleranzbereiche und gleitende Übergänge zwischen den Merkmalsausprägungen bestehen bleiben. Dieser Ansatz ist im Übrigen mit den Beobachtungen von Werker und Desjardins (1995, 79) und Kuhl (2000, 11852) stimmig.

Auch Erwachsene können durch intensives Training und Sensibilisierungstechniken lautliche Unterscheidungen wieder treffen, die in ihrer Sprache normalerweise irrelevant sind. Dieser Umstand wird mit Reorganisation umschrieben und drückt aus neurobiologischer Sicht nichts weiter aus, als dass hier neue Synapsen zwischen den entsprechenden Neuronenpaaren gebildet werden beziehungsweise vorhandene schwächere Synapsen verstärkt werden. Aufgrund des vorhandenen synaptischen Netzwerkes unterscheidet sich dieser Lernprozess von dem eines Kindes im Säuglingsalter, in dem es noch keine oder nur wenige synaptische Verbindungen gibt (vgl. Elman, 1999, 17f). Dieser Überlegung liegt die Anwendung der allgemeinen Lerntheorie auf das Erlernen von Phonemen zu Grun-

des Sprachsignals reagieren und nicht auf andere Signale, die keine sprachlichen Qualitäten aufweisen (vgl. Dietrich (2007, 219) basierend auf Bendor und Wang (2005) oder Zatorre (2005)). Dem könnte man entgegen, dass die Neuronen natürlich auf alle akustischen Signale reagieren. Da aber nur die Sprachlaute in der Häufigkeit und Genauigkeit vorkommen, können sich für die vereinzelt anderen Signale keine starken synaptischen Verbindungen entwickeln wie die Sprachlaute.

⁴³Zenner (1994, 129):

Im Tierexperiment scheinen manche Zellen ausschließlich dann zu reagieren, wenn ein ursprünglich komplexes Schallsignal frequenzmodelliert wurde. . . . Andere reagieren nur auf komplexe Phoneme, wie bestimmte Tierlaute. Es ist jedoch nicht anzunehmen, daß die betreffenden Zellen spezifische Detektoren für den gesamten komplexen Stimulus sind (z.B. das gesamte Phonem oder den gesamten Tierlaut). Vielmehr muß man annehmen, dass sie auf grundlegende akustische Elemente der Stimuli reagieren, welche den Stimuli ihre charakteristischen Merkmale geben. Werden diese charakteristischen Merkmale durch die vorangegangenen Stationen des Hirnstammes extrahiert, so können sie im Kortex empfunden werden. Diese Empfindung trägt zur Gesamtwahrnehmung des Gesamtstimulus bei. Der auditorische Teil der Wahrnehmung setzt sich damit aus verschiedenen extrahierten, charakteristischen Merkmalen der Schallinformation zusammen.

de.⁴⁴ Auch psycholinguistische Studien (Maye et al., 2002) oder neuere Entwicklungen der Emergenz (MacWhinney, 2001, 451) belegen die Plausibilität dieses Ansatzes.

Ein weiteres Argument, welches für das Phonem als Segmentierungseinheit spricht, ist die Sparsamkeit (*parsimony*) bei der Ausführung kognitiver Prozesse. Unter der Prämisse der Sparsamkeit kognitiver Prozesse erhält das Phonem einen entscheidenden Vorteil. Unterstellt man, dass Phoneme oder deren Merkmalsvektoren (wie oben beschrieben) durch sich verändernde raum-zeitliche Neuronensembles repräsentiert werden und dass die daran beteiligten frequenzsensiblen Neuronen synchron, also zeitgleich feuern, dann lässt sich der Reduktionismus der Häufigkeit als grundlegende Gesetzmäßigkeit abermals anwenden. Hieraus ergibt sich auf theoretisch logischer Ebene, dass es keine neuronale Repräsentationseinheit gibt, die so oft gemeinsam feuert wie die Neuronen, die ein Phonem repräsentieren. Dies gilt sowohl hinsichtlich einzelner phonetischer Eigenschaften als auch für Einheiten höherer Organisation, wie Silben oder Wörter und lässt sich mathematisch einfach zeigen. Für die phonetischen Eigenschaften kann man eine mengentheoretische Überlegung anstellen. Zur Menge eines Phonems gehört eine Anzahl an phonetischen Eigenschaften, die durch ein Neuronenmuster repräsentiert sind. Es gilt somit für ein beliebiges Neuronenmuster des Phonems A:

$$A = \left\{ \sum_{i=1}^k \chi_i \right\}, \quad (4.1)$$

wobei χ die phonetische Eigenschaft ist und k die Anzahl der phonetischen Merkmale, die zur Bestimmung des Phonems genügen.⁴⁵ Unter der Bedingung, dass Neuronen, die einzelne phonetische Merkmale kodieren, nicht einzeln feuern können, sondern immer zeitgleich mit anderen Neuronen, weil sie auch zeitlich zusammen im Signal auftreten, lässt sich die Einheit eines Phonems begründen. Ein einzelnes χ_i kann zwar öfter auftreten als A, aber immer nur innerhalb eines bestimmten Neuronenmusters, welches wiederum ein anderes Phonem bestimmt. Beispielsweise könnte eine Neuronengruppe für einen bestimmten Frequenzbereich der Stimmhaftigkeit sensibilisiert sein. Diese phonetische Eigenschaft tritt bei mehreren Phonemen auf. So würde das entsprechende Neuronenmuster sowohl bei /b/ als auch bei /d/ feuern, aber nicht allein, sondern als Teilmenge in dem Gesamtmuster des /b/ oder des /d/. Es gibt so kein χ_i , das isoliert öfter feuert als A. Grafisch lässt sich der Zusammenhang wie in Abbildung 4.1 veranschaulichen.

Nähert man sich der Frage dann aus kybernetischer Perspektive, kommt man unweigerlich zu dem Schluss, dass die kleinste Ordnungseinheit, genau die Einheit sein muss, welche eine Veränderung des Systems bewirkt und die in keine Untereinheit geteilt werden kann, die allein eine Veränderung des Systems hervorrufen könnte. Übertragen auf das System der Sprache, nehmen die Phoneme diese Rolle ein.⁴⁶ Eine Änderung eines

⁴⁴vgl. etwa Spitzer (2002, 62ff, 69, 79ff) und Spitzer (2000, 19ff, 42ff, 229ff); Brown (1999) oder Blake-more und Frith (2006, 61f).

⁴⁵ i ist nur aus Gründen der Vereinfachung mit 1 initialisiert. Phoneme besitzen natürlich immer mehr als ein phonetisches Merkmal.

⁴⁶Man bedenke, dass sich Phoneme oft nur in einem relevanten phonetischen Merkmal unterscheiden.

4 Randbedingungen der Simulation

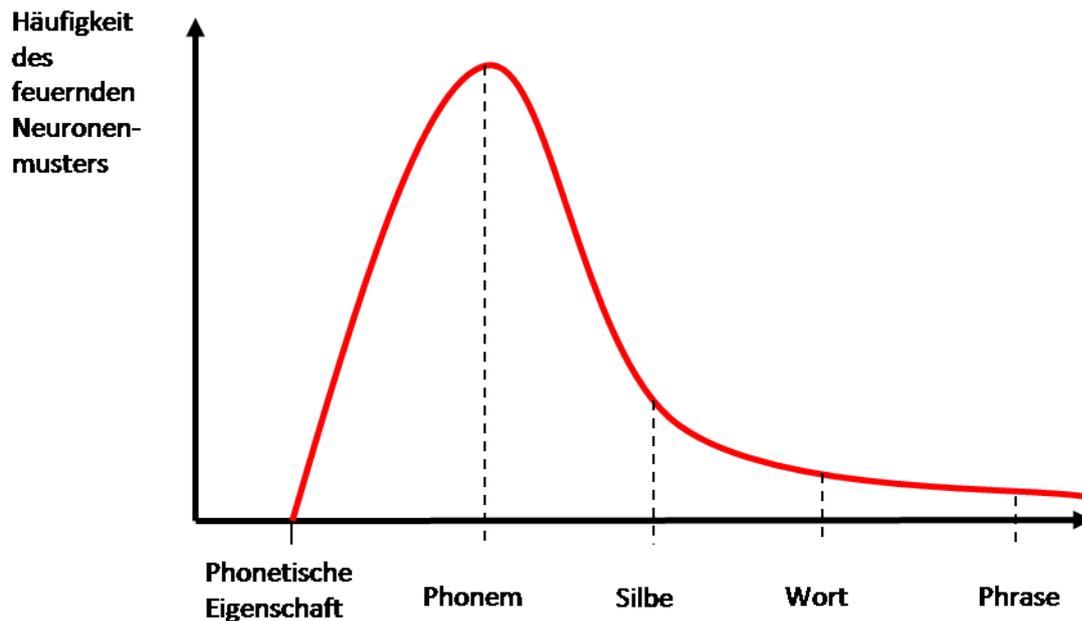


Abbildung 4.1: Hypothetische Annahme einer häufigkeitsabhängigen Repräsentation von Wahrnehmungseinheiten, Quelle: Eigene Darstellung

für eine Sprache relevanten phonetischen Merkmals ändert dieses in ein anderes Phonem um. Ist das Merkmal nicht bedeutungsunterscheidend, werden zwar Variationen als Allophone auftreten, eine Änderung auf der nächsten Ordnungsebene der Morpheme wird aber nicht zu beobachten sein. Genauso verhalten sich kleinere Variationen im Frequenzbereich des Signals. Erst wenn diese Variationen einen dehnbaren Toleranzbereich überschreiten, erfolgt eine Änderung im Output.

Definitionsgemäß wäre hier die Silbe nicht als kleinste Ordnungseinheit aufzufassen, da sie in kleinere Einheiten zerlegt werden kann, die allein eine Änderung des Outputs bewirken. Genau dies ist aber bei Phonemen nicht der Fall, da sich hier die einzelnen phonetischen Eigenschaften nicht voneinander isolieren lassen.⁴⁷ So müsste man den Phonemen auf theoretischer Ebene den Status der kleinsten Sprachbausteine zusprechen und somit sollten auch die statistischen Verteilungen der Phoneme Grundlage der anstehenden Simulation sein.

Diese systemtheoretische Lösung spiegelt sich aber nicht eindeutig in der experimen-

So könnte man zu dem Schluss gelangen, dass die einzelnen phonetischen Merkmale die systembestimmenden Einheiten sind, die sich ähnlich wie die Phoneme in Wörtern, zu höheren Einheiten zusammenschließen könnten. Dieser Einwand ist nicht ganz unberechtigt, aber aus der hier geschilderten Perspektive unbedenklich, weil die einzelnen phonetischen Merkmale nicht allein auftreten können, sondern immer nur zusammen mit anderen Merkmalen eine bedeutungsunterscheidende Einheit in einem Verbund bilden.

⁴⁷vgl. Jusczyk und Luce (2002, 4) für eine konzise Zusammenfassung dieses Gedankens, der schon am Anfang der Debatte um die kleinsten linguistischen Einheiten geäußert wurde.

tellen Forschung für die Sprachwahrnehmung von Kleinkindern und Erwachsenen wider, weil das Signal nicht die hier unterstellte strenge Linearität aufweist.⁴⁸ Man kann dagegen einwenden, dass auch die Phoneme im Sprachsignal nicht als einzelne Einheiten auftreten, sondern auf gleiche Weise, wie die einzelnen phonetischen Merkmale, das Frequenzspektrum nur im Verbund gestalten.⁴⁹ Das ist natürlich nicht bestreitbar, aber im Falle von Phonemen lässt sich, wenn auch nicht wie in der generativen Phonologie behauptet, streng linear, aber doch eine zeitliche Abfolge der Phoneme verzeichnen, die durch Assimilationen und Koartikulationen mehr oder weniger stark kenntlich ist (Remez, 1994, 148). Genau diese Abfolge lässt sich bei den phonetischen Merkmalen nicht feststellen, weil sie zusammen erst eine akustische Schwingung erzeugen. Die Frequenzzusammensetzung der Sprache ist jedoch derart kompliziert, dass eine genauere Aufschlüsselung zunächst nicht möglich ist (Gegenfurtner, 2005, 59).

Für eine ausführliche Begründung der Wahl einer Segmentierungseinheit kann es hilfreich sein, sich die verschiedenen Positionen der führenden Wissenschaftlern auf diesem Gebiet vor Augen zu führen. Saffran et al. (2006) versuchen die Argumente zu vereinen.

[S]peakers of different languages employ different units as their primary unit of segmentation. Speakers of syllable-timed languages (e.g. French, Spanish, Catalan, & Portuguese) show a processing advantage for the syllable . . . , speakers of stress-timed languages such as English and Dutch show greater access to the phoneme . . . , and Japanese adults use the mora as the primary unit of segmentation. (Saffran et al., 2006, 77)

Den Bezug zum sprachlernenden Kind schildern die Autoren nur wenige Seiten zuvor.

By 6 months of age, infants probably have adultlike representations of speech and other complex sounds. However, this is not to say that their perception of complex sounds is adultlike. It is clear that infants do not attend to the information within complex sounds in the same way that adults do. (Saffran et al., 2006, 71)

Sie beziehen sich offensichtlich auf die Arbeit Cutlers, welche sie mit eigenen Erkenntnissen über die Fähigkeiten von Säuglingen in Beziehung setzen und damit die eingangs genannte (Kap. 4.1) Entwicklungsrichtung skizzieren. Der Unterschied zwischen Kleinkindern und Erwachsenen wird in der Wahrnehmung, aber nicht in der vorhandenen Repräsentation, gesehen.

Cutler et al. (2001) beziehen einen recht deutlichen Standpunkt und argumentieren gegen die Silbe als *unit of perception*. Sie stellen eine kleine Kollektion an Argumenten zusammen, welche die Silbe, wie sie von Mehler (1981) so rigoros als universale Wahrnehmungseinheit angepriesen wird, logisch widerlegen könnten. Sie gestehen der Silbe eine mögliche Ersatzrolle zu und kehren damit den Vorschlag Mehlers um. Die Silbe spiele in

⁴⁸Marcus (1984) beschreibt eine Computersimulation, in der die Linearität der Einheiten aufgehoben wird. Der unbekannt Input wird mit kontextsensitiven Repräsentationen abgeglichen. Diese Repräsentationen sind mögliche Lautkombinationen der Sprache, welche, wenn auch nur zum Teil im Input enthalten, zugewiesen werden. Diese Simulation zeigt, dass eine strenge Linearität der Phonemabfolgen im Input nicht unbedingt vorhanden sein muss und trotzdem intern eine lineare Repräsentation entsteht.

⁴⁹Hockett (1955, 210) verwendet hier eine berühmte Analogie zu Ostereiern, welche wenn zusammen geworfen im Nachhinein mittels Analyse ihres Eigelbs und der Schale wieder zu ihrer ursprünglichen Form identifiziert werden sollen.

4 Randbedingungen der Simulation

speziellen Sprachen sicherlich eine größere Rolle als in anderen Sprachen. Daraus ließe sich jedoch keine Universalität ableiten.

A theory in which the syllable is the universal „unit of perception“ cannot account for this crosslinguistic variability. (Cutler et al., 2001, 189)

Wenn die Worterkennung auf Silben basiert, wären Wörter, die in keine Silben aufgeteilt werden können, schwieriger zu erkennen sein. Dafür gibt es jedoch keine Anhaltspunkte. Silben erleichtern die Segmentierung möglicherweise, aber diese Eigenschaft kann sie nicht in den Status einer universalen Einheit heben.

Syllables are not, as Mehler (1981) suggested units of perception into which listeners classify spoken input prior to lexical access. Instead, the syllable appears to be the measuring stick against which viable and unviable parses of continuous speech are judged. Syllables form acceptable chunks in the ongoing lexical parse of the speech stream, nonsyllabic sequences do not. (Cutler et al., 2001, 191)

Gerken und Aslin (2005) geben in ihrer Festschrift an Jusczyk einen Abriss über seine Forschungsphasen der Sprachperzeption. Die letzte Phase ist von einer Erkenntnis geleitet:

[I]nfants are keeping track of the distributional properties of their native-language input, although at the level of diphones rather than individual phonetic segments. (Gerken und Aslin, 2005, 7)

Diese Idee genießt derzeit recht große Anerkennung, weil sie viele experimentelle Befunde plausibel erklären kann und einen Kompromiss zwischen Silben und Phonemen bietet. Einerseits muss der abstrakte Charakter der Phoneme aufgegeben und die notwendigen Informationen einem variablen akustischen Input zugesprochen werden. Andererseits erfüllt dieser Input noch die Aufgaben eines Segments. Wie weit dieses phonetische Segment sich von einer abstrakten Einheit unterscheidet, unterliegt einer Definition, welche noch nicht in der notwendigen Genauigkeit dargelegt werden kann.

Auch Macken (1995) lässt eine klare Tendenz für Segmente erkennen:

we reasonably infer that the child „segments“ the input and represents words in terms of a linear sequence of segment-sized or equivalent-length units. (Macken, 1995, 688)

Dazu muss man wissen, dass in der Literatur zu diesem Themengebiet auf die Realitäten der Phoneme verwiesen wird, wenn man von Segmenten spricht. Die genaue Form dieser Segmente bleibt jedoch vage.

The format of these beginning representations cannot be decided at this point. (Macken, 1995, 688)

Vihman (1996) in Vorwegnahme der Aussage von Saffran et al. (2006) gibt zum damaligen Wissenstand ein Remis bekannt.

these results ... provide no clear evidence that syllables, rather than smaller or larger units, provide a first basis for perceptual processing. (Vihman, 1996, 69)

Dennoch sieht Jusczyk grundsätzlich Anlass, die Silbenstruktur als Repräsentationseinheit vorzuziehen (Jusczyk, 1997, 112ff), aber er räumt ein, dass die Methodik trotz aller Modifikation nicht unfehlbare Ergebnisse liefern kann. Zudem grenzt er das Alter für die Erkennung einer Silbenstruktur auf das erste halbe Jahr nach der Geburt ein. Unter

Berücksichtigung der Forschung Hillenbrands (Hillenbrand (1983) u. Hillenbrand (1984)) gesteht er weiterhin zu:⁵⁰

One possible implication of these results is that 6-month-olds are beginning to perceive some segmental and/or featural organization within syllables. (Jusczyk, 1997, 115)

Aus dem dargebotenen Material ist nun doch eine Tendenz für einzelne Segmente in Form der Phoneme entstanden. Die Silbe, wenn als Analyseeinheit betrachtet, bringt eine Reihe erheblicher Schwierigkeiten mit sich. Speziell hinsichtlich eines Simulationsdesigns ist schon die Mehrdeutigkeit der Silbe problematisch.⁵¹ Die Sprecher einer Sprache mögen eine oft genaue, aber doch intuitive Wahrnehmung von der Silbe haben. Soll sie jedoch genau definiert werden, treten Schwierigkeiten auf (*balcon* vs. *bacon*).

Wie in der oben geführten Diskussion schon angeführt wurde, ist es nicht möglich, einen englischen Text in Silben automatisiert aufzuteilen, ohne dass man recht spezifische prosodische Informationen einbezieht. Wenn man aber genau diese Informationen doch ausblenden möchte, um den Beitrag der rein statistischen Verteilungen zu messen, führt die Verwendung eines silbifizierten Textes zu einem Zirkelschlussproblem.⁵² Weil dieser Zirkelschluss vermieden werden soll und bislang kein Forschungsergebnis zeigen kann, dass die Silbe oder das Phonem die grundlegende Einheit ist, wäre schon allein deshalb das Phonem die bessere Wahl für ein Simulationsmodell. Dies hat zudem den Vorteil, dass Diphone, wie sie von Gerken und Aslin (2005) vorgeschlagen wurden, auch einbezogen werden könnten, allerdings in abstrahierter Form.

Andererseits kann man auch von der rigiden Vorstellung Abstand nehmen, dass Kleinkinder nur eine Art der Wahrnehmungseinheit haben können. Auf diese Weise nähert man sich den Ergebnissen Jusczyks an, wenn man annimmt, dass Phoneme nicht die kleinsten Repräsentationseinheiten sein müssen, aber dass sie dann als trennbare Einheiten erkannt werden.⁵³ Dies entspricht dann der Vorstellung Jusczyks (Jusczyk, 1997), die Phoneme übernehmen eine strukturorganisierende Rolle innerhalb der Silbe. Diese Eigenschaft wäre für zwei Monate alte Babys noch nicht nachzuweisen (Jusczyk und

⁵⁰Hillenbrand (1984) wies die Fähigkeit bei sechs Monate alten Babys nach, den Unterschied des labialen und alveolaren Artikulationsortes zu erkennen. Diese Untersuchung implizierte aber auch, dass die Lautumgebung der Vokale, keinen Einfluss auf die Erkennungsleistung ausübte. Wenn aber die Vokale keine Information zur Erkennung des Artikulationsortes beitragen, kann man auch davon ausgehen, dass die Merkmale der beiden Nasale allein identifiziert werden könnten und somit eigene Repräsentationseinheiten bilden.

⁵¹Vgl. McQueen (1998, 37):

If syllabification is indeterminate in some sequences of speech sounds, then a speech segmentation procedure requiring the classification of all speech input into syllables will not work reliably.

⁵²Die nativistische Antwort lautet dann, dass man schon allein für die Silbenstrukturierung auf die Universalgrammatik zurückgreifen muss (Yang, 2004).

⁵³Aus neurowissenschaftlicher Perspektive schlagen Hickok und Poeppel (2004) eine duale Verarbeitung vor. Sie beobachteten zwei Verarbeitungssysteme nach der akustisch-phonetischen Analyse des Inputs. Das dorsale Bündel übernimmt die explizite Segmentierung in phonologische oder akustische Einheiten und das ventrale System übernimmt eine Bedeutungszuordnung mit nicht näher spezifizierbaren Einheiten.

Derrah (1987); Bertocini (1988)), schon aber für Säuglinge doppelten Alters (Miller und Eimas (1979); vgl. Houston et al. (2004, 423)).

4.2 Statistische Berechnungen und logisches Schließen

Eine zusammenfassende Beschreibung der frühkindlichen Fähigkeit zur Erkennung von statistischen Zusammenhängen wurde bereits in Kapitel 3.2.4 verfasst. Eine Wiederholung der experimentellen Befunde erscheint somit unzweckmäßig. Stattdessen sollen weitergehende Ausführungen die Verwendung des statistischen Lernens als Randbedingung der Simulation rechtfertigen.

Eine erste theoretische Untersuchung für das Vorhandensein statistischer Mechanismen haben Hayes und Clark (1970) für Erwachsene erbracht. Nach einer anfänglichen Diskussion über weitere Möglichkeiten zur Segmentierung, stellen sie zwei Experimente unter Verwendung einer Kunstsprache an. Im ersten Experiment hörten die Probanden einem Lautstrom 45 Minuten lang zu und sollten dann entscheiden, wo sich mögliche Wortgrenzen einfügen ließen. Obwohl das Stimulusmaterial so gewählt wurde, dass keine Hinweise auf Pausen oder rhythmische Figuren vorhanden sein konnten, erkannten die Probanden mögliche Cluster in der Kunstsprache. Diese Gruppen entsprachen natürlich den von der *finite state grammar* erzeugten Wörtern. Im zweiten Experiment konnte zudem gezeigt werden, dass diese Erkennungsleistung von der Anzahl der Phoneme abhängt. Bei einer Sprache mit nur vier Phonemen können keine Muster mehr erkannt werden, wohl aber schon mit sechs.

Hasher und Chromiak (1977) gehen beim Aufnehmen und Verarbeiten von allgemeinen Fähigkeiten der Kognition aus, welche zwar auch auf die Sprache anwendbar, aber nicht darauf begrenzt bleiben. Eine mögliche allgemeine Lernstrategie ist das Benutzen von Häufigkeiten von Ereignissen der Welt oder auch von Wörtern. Dieser Prozess der Wahrscheinlichkeitsschätzung läuft unterbewusst ab.⁵⁴ Zwei Experimente sollten die dort ablaufenden Gesetzmäßigkeiten besser verstehen helfen. Im ersten Experiment wurden Probanden verschiedenen Alters (Schulkinder und Studenten) mit Wortlisten konfrontiert. Dabei wurde eine Gruppe zu Beginn über ihre Aufgabe informiert, während die andere Gruppe die Anweisung erst nach der Präsentation der Wortliste erhielt. Es zeigte sich, dass weder das Alter noch der bewusste Prozess des Zählens Auswirkungen auf die Leistung hatten. Im zweiten Experiment konnte dann noch gezeigt werden, dass auch mehrfaches Üben mit und ohne Feedback die Leistung nicht verbessert.

frequency counting or tagging is something which the organism engages in as an essential component of his processing of the world; or in the case of this study, in the processing of a list of words. (Hasher und Chromiak, 1977, 182)

Einige Jahre später wurden diese Ergebnisse *under truly memory conditions* Hasher et al. (1987, 69) repliziert. Vier Experimente bestätigten die aufgestellten Annahmen (*criteria*):

⁵⁴vgl. Hasher und Zacks (1984) für eine ausführliche Herleitung und Begründung.

4.2 Statistische Berechnungen und logisches Schließen

automatically encoded information is stored in memory in the absence of conscious intentions to do so ... the quality of this encoding is uninfluenced by the voluntary goals or strategies of an individual. (Hasher et al., 1987, 69)⁵⁵

Die Experimentiermethodik folgte einem ähnlichen Muster. Um die unbewusste Aufnahme der Worthäufigkeit, die eigentliche Worterinnerungsaufgabe (*memory test*) messen zu können, lösten die Probanden zur Ablenkung Aufgaben, wie das Erkennen von Anagrammen (Experiment 1), wie die Satzvervollständigung (Experiment 2) oder wie Bild-Wort Abgleiche (*stroop interference task* in den Experimenten 3 und 4).⁵⁶ Im Ergebnis bleibt auch hier festzuhalten:

... people can reliably discriminate the frequencies of such naturally occurring events as single letters, letter pairs, surnames, occupations, and diseases. (Hasher et al., 1987, 87)

Die Versuchspersonen können sich also unterbewusst die Häufigkeit zu den Lautsequenzen gut merken und diese auch für die weitere kognitive Verarbeitung verwenden.

McQueen und Pitt (1996) zeigten schließlich, dass Erwachsene Sprecher die Übergangswahrscheinlichkeiten immer dann benutzen, wenn die einzelnen Segmente schwieriger zu erkennen sind. Diese Fähigkeit wird sowohl zur Wortsegmentierung als auch zur Erkennung einzelner Sprachlaute bei gesprochener Sprache abgerufen. Die Verwendung dieser statistischen Informationen in Form der Übergangswahrscheinlichkeiten findet auch bei Lexikonabgleichen zwischen lexikalischer und vorlexikalischer Ebene in einem *Top-down*-Verfahren statt (Pitt und McQueen, 1998). Wenn Erwachsene einem Sprachstrom unbekannter Einheiten segmentieren sollen, verwenden sie ebenso die Übergangswahrscheinlichkeiten, nicht aber wenn es zu strukturellen Musterabgleichen kommt. Hier greifen die Hörer auf eine algebraische Lösung zurück (Peña et al., 2002).⁵⁷

Treffen diese Fähigkeiten aber auch auf wenige Monate alte Babys zu? Goodsitt et al. (1993) wiesen eine Art *Clustering*-Strategie für sieben und acht Monate alte Babys nach. Vorbild waren die Experimente von Hayes und Clark (1970), die auf die Erstspracherwerbsforschung angepasst wurden. Nach einer Trainingphase mit Silben, sollte die Fähigkeit der Säuglinge geprüft werden, ob sie diese auch in drei verschiedenen Umgebungen (*invariant, redundant, variable*) wieder erkennen würden. Nach 20 Durchläufen stellte sich für alle drei Szenarien die gleiche Erkennungsleistung heraus. Nach 20 weiteren Versuchen verbesserte sich jedoch lediglich die invariante Umgebung, in der die Silbenanordnung so gewählt wurde, dass sie eine mit Übergangswahrscheinlichkeiten kodierte Struktur offenlegte.

Allerdings erbrachten erst Saffran et al. (1996a) den entscheidenden Beweis (siehe Kap. 3.2.4). Die sich daraus entwickelnde Diskussion hat sich insofern als fruchtbar er-

⁵⁵Berns et al. (1997) vereinen die Erkenntnisse aus Hayes und Clark (1970) und Hasher et al. (1987) insofern sie die Gehirnregionen für beide Beobachtungen angeben können. Dabei wird nicht nur „... sequential information [in Form einer endlichen Grammatik] ... learned and used in the absence of awareness“, sondern auch von Kleinkindern genauso wie von Erwachsenen verarbeitet.

⁵⁶Im vierten Experiment wurden die Probanden direkt instruiert, sich die Häufigkeit bestimmter Wörter zu merken. Ansonsten war die Experimentieranordnung mit Experiment 3 identisch.

⁵⁷Für eine kritische Bemerkung der Studie hinsichtlich der algebraischen Regelableitung siehe Seidenberg et al. (2002).

4 Randbedingungen der Simulation

wiesen, dass sie zu einer kleinen Explosion an relevanten entwicklungspsychologischen Experimenten und Publikationen führte. Aslin et al. (1999) bezogen diese Fähigkeiten auf sprachliche und außersprachliche Phänomene. Dieses Ziel verfolgte auch Saffran et al. (1999) und Saffran (2003c) mit einer Untersuchung zur Anwendbarkeit des statistischen Lernens in der mentalen Verarbeitung von Musik und Tonhöhen. Saffran et al. (1999) und Saffran (2003c) berichteten zwar über unterschiedliche Strategien bei Erwachsenen und Kleinkindern, aber die Fähigkeit, Mechanismen des statistischen Lernens zu nutzen sind gleichermaßen vorhanden. Bereits in Saffran et al. (1997b) zeigten zwei Experimente, dass Erwachsene und Kinder gleichermaßen sprachliche Informationen auf Grundlage von statistischen Verteilungen beiläufig aufnehmen. Gómez und Gerken (1999) weiteten den Ansatz des statistischen Lernens auf komplexere Grammatiken aus. In ihrem vierten Experiment zeigten sie, dass 12 Monate alte Babys eine recht komplexe endliche Grammatik aus dem präsentierten Material erkennen, sodass man nicht nur von *first order dependencies*, wie der Wortsegmentierung ausgehen sollte, sondern auch höhere Ordnungsstrukturen in den Ansatz des statistischen Lernens aufnehmen darf. Sie unterstellten hier eine Art statistischen *Child Parser* (Saffran et al., 2001, 12874). Eine ähnliche Frage adressierten Saffran (2001a) und Saffran und Wilson (2003). Sie untersuchten den Einfluss des statistischen Lernens in einfachen Phrasenstrukturen bei Kindern⁵⁸ und Erwachsenen. Auch ihre zwei Experimente wiesen in die gleiche Richtung.⁵⁹ Die Mechanismen des statischen Lernens finden bei Erwachsenen und Kindern in verschiedenen kognitiven Domänen Anwendung. Sprache und das abstrakte syntaktische Regelsystem sind nur eine kognitive Dömane, in der statistische Zusammenhänge die wichtigen Hinweise für eine Regeldeduktion geben können. Saffran (2001a) bleiben jedoch vorsichtig in ihrer Formulierung:

humans possess statistical learning mechanisms which may assist in the acquisition of this abstract component of natural language. (Saffran, 2001a, 493)

Sie erachten die statistischen Lernmechanismen zwar als unbestreitbaren Bestandteil der menschlichen Kognition, legen sich aber nicht fest, ob diese Mechanismen tatsächlich für das Erlernen einer abstrakten Grammatik verantwortlich sind. Denn dabei stellt sich natürlich die Frage, wie die Sprachlerner wissen, von welchen Strukturen sie Statistiken bilden müssen. Dafür hat man aber noch keine eindeutige Lösung gefunden. Die Möglichkeiten der Strukturen sind unbegrenzt. Schließlich könnte man immer das erste Wort, die drittletzte Silbe, das siebente Phonem, etc. einer Lautkette zählen und daraus eine Häufigkeitsverteilung erstellen. Die Lösung soll das altbekannte Konzept der Randbedingungen (*Constraints*) bringen (Jusczyk (1998); Newport und Aslin (2000)), womit man sich jedoch wieder in die Nähe eines Sprachmoduls begibt. Auch hier muss man sich die Frage gefallen lassen, wie die schon recht speziellen Randbedingungen entstanden sein

⁵⁸vgl. auch Morgan (1996a) für einen kurzen Bericht über die Frage, ob kindgerichtete Sprache ein ausreichender Input für das Erkennen einer Phrasenstruktur ist. Nach Meinung des Autors ist dies der Fall.

⁵⁹Das statistische Lernen in linguistischen Domänen ist nicht auf die Bereiche der Wortsegmentierung und des Syntaxerwerbs beschränkt. Gómez (2007) weitet mit einer zusammenfassenden Darstellung des phonologischen Lernens das Einzugsgebiet weiter auf die Kategorisierung von Sprachlauten und Generalisierung von Silbenmustern aus.

können. Bleiben die Randbedingungen jedoch allgemein und für das gesamte kognitive System übertragbar, ließen sich die Besonderheiten des Spracherwerbs auch mit allgemein gültigen kognitiven Fähigkeiten erklären. Saffran (2002) führte dazu eine Reihe von Experimenten durch, in welchen sie *predictive dependencies* als Randbedingungen zur Musterextraktion einer Grammatik unterstellte.⁶⁰ Die vorhersagbaren Strukturen halfen natürlich beim Erkennen der Grammatikstruktur. Derselbe Mechanismus konnte dann auch bei nicht sprachlichem tonalem und visuellem Input nachgewiesen werden. Auf diese Weise wurde klar, dass das statistische Lernen zwar zielgerichtet geleitet sein könnte, aber auf jeden Fall einen Mechanismus verwendet, der für das gesamte kognitive System relevant ist und nicht nur auf die Sprache als Modul beschränkt bleibt.⁶¹

4.3 Erinnerungsvermögen von Lautsequenzen

Eine wichtige Fähigkeit, die Kleinkinder besitzen müssen, um die Annahmen des hier entwickelten Algorithmus zu erfüllen, ist, die häufigsten extrahierten Phonemketten langfristig in Erinnerung zu behalten. Sobald die Einheiten mit den Übergangswahrscheinlichkeiten im Lautstrom erkannt werden können, müssen sie in der ein oder anderen Form repräsentiert werden. Nur dann können diese Einheiten zur erweiterten Segmentierung auch genutzt werden. Nimmt man die Fähigkeit einer häufigkeitsbasierten Erinnerung hin, wie implizit in Jusczyk et al. (1994) oder explizit in Shi et al. (2006) gezeigt werden konnte, kristallisieren sich zwei weitere Experimente als zentral heraus:⁶² Jusczyk und Aslin (1995) und Jusczyk und Hohne (1997).

Shi et al. (2006) haben den bekannten Frequenzeffekt (Broadbent, 1967) auf das Segmentierungsproblem bei Säuglingen bezogen. Im Spezifischen untersuchen sie den Einfluss der hochfrequenten Funktoren des Englischen, wie dem Artikel *the*, auf die Erkennungsleistung der umgebenden lexikalischen Einheiten. Acht Monate alte Babys sind dazu in der Lage, wenn auch noch in einer eher unterspezifizierten Repräsentation, weil nicht nur das *the*, sondern auch das seltene *kuh*, wohl durch dessen besondere akustische Wirkung, zur Segmentierung genutzt wurden. Bei elf Monate alten Babys war der Sonderstatus des *kuh* bereits verschwunden.

Jusczyk und Aslin (1995) verwenden die *Head Turn Preference Procedure*, um zu zeigen, dass siebeneinhalb Monate alte Babys einzelne, ihnen bekannte Wörter in einer fremden Lautumgebung wiedererkennen (Experiment 1). Sechs Monate alte Babys verfügen noch nicht über diese Fähigkeit (Experiment 2). Dafür werden die Probanden mit einsilbigen, isoliert auftretenden Wörtern konditioniert und diese werden in der Testphase mit fremden Wörtern umgeben. Im dritten Experiment wird nur ein phonetisches

⁶⁰vgl. auch die Zusammenfassung in Saffran (2003b).

⁶¹Diese Ergebnisse entsprechen im Grunde der Ansicht von Langacker (1999, 91ff u. 129) und Langacker (1990, 1ff).

⁶²In Jusczyk et al. (1995) konnte schon gezeigt werden, dass zwei und drei Monate alte Babys Informationen über ein phonetisches Merkmal nicht nur zum Vergleich mit neuen Input zur Verfügung steht, sondern auch über längere Zeiträume behalten werden kann (Jusczyk et al., 1995, 38).

4 Randbedingungen der Simulation

Merkmal des Stimulusmaterials geändert und die Erkennungsleistung ist immer noch signifikant. Dies weist natürlich auch auf die hohe Detailliertheit der zu verarbeitenden Informationen hin. Im vierten Experiment wird das Trainingsmaterial sogar so weit verändert, dass nun die Zielwörter in Sätzen vorgetragen werden. Auch hierfür konnte eine positive Evidenz gefunden werden. Zusammenfassend lässt sich damit sagen:

Information about such regularities in the input ... is apt to prove useful in identifying sequences of sounds that are likely to be an important component in forming and building up a lexicon. (Jusczyk und Aslin, 1995, 17)

Für die Zwecke der Simulation darf eine Vorstufe des Lexikons, ein Paralexikon, unterstellt werden, weil noch keine Bedeutungen mit den Phonemketten verbunden sind. Die so gezeigten Fähigkeiten setzen natürlich ein Erinnerungsvermögen voraus, weil zwischen der Trainingsphase und der Testphase immer Zeit vergeht. Diese Zeitspannen wären sicher zu kurz, um eine realistische Entsprechung für die Prozesse der Simulation zu haben. Deshalb muss eine weitere Experimentenreihe hinzugezogen werden.

Diese Experimentenreihe von Jusczyk und Hohne (1997) unterstützt die Aussage, dass sich acht Monate alte Babys Lautmuster über zwei Wochen merken können, ohne sie mit einer Bedeutung zu verbinden. Dies ist natürlich auch eine wichtige Voraussetzung zur Bildung des Lexikons. Die Worte der Autoren geben die Bedeutung des Experiments abermals sehr treffend wieder:

The fact that the initial exposure to the words occurred in fluent speech contexts and subsequently led to listening preferences for citation versions of the words suggests that infants not only attended to and stored the sound patterns of words that recurred in the stories, but also that they have some means representing these patterns and compensating for different acoustic-phonetic forms of the same words. (Jusczyk und Hohne, 1997, 1985)

Ergänzend zu diesen Befunden könnte auch noch die Fähigkeit nützlich sein, sich längere Wortgruppen zu behalten und damit nicht nur einsilbige oder zweisilbige Einheiten in der Simulation zuzulassen. Nazzi et al. (2000b) und eine leichte Verbesserung in der Experimentieranordnung davon in Soderstrom et al (2005) zeigen, dass sich sechs Monate alte Babys längere Satzpassagen und einige prosodische Merkmale davon merken können und auch in Textpassagen wieder erkennen. Dabei gibt allerdings die Intonation wesentliche Anhaltspunkte, ob die bekannte Einheit über mehrere Sätze hinwegreicht (z.B. ... *rabbits eat. Leafy vegetables* versus ... *Rabbits eat leafy vegetables.*).⁶³

4.4 Verwendung und Form der segmentierten Einheiten

Christophe et al. (1994) geben Anlass über die Verwendung und Form von wortähnlichen Einheiten nachzudenken. Ihr Experiment ging zwar einer anderen Fragestellung nach, aber die daraus gewonnenen Erkenntnisse sollte man auch der letzten Simulationsrandbedingung nicht vorenthalten. Die Experimentatoren bestätigten, dass drei Tage alte französische Babys einen Unterschied zwischen *mati* in *mathématicien* und *pyjama tissé* erkennen können. Ob das Experiment als Beweis für die Wortgrenzenerkennung bei drei

⁶³Siehe hierzu die detaillierten Ausführungen in Kapitel 3.2.1.

Tage alten Säuglingen taugt, mag dahingestellt sein, weil die Präsenz eines Unterschieds noch keine Informationen über das Konzept der Wortgrenze enthält. Im Französischen kann der prosodische Unterschied der zwei Formen von *mati* möglicherweise eine Wortgrenze definieren. Das darf aber nicht heißen, dass damit auch schon die Grenzen der Einheit bekannt sind. *mati* könnte genauso als eine Art Minimalpaar gespeichert werden, wobei das unterschiedliche Betonungsmuster das unterscheidbare Merkmal des Paares wäre. Bedeutsam für die hiesige Simulation ist nun, dass Babys schon sehr frühzeitig in der Lage sind, Einbettungen kleinerer Einheiten in längeren Konstrukten zu erkennen. Dabei spielen anscheinend schon sehr feine prosodische Unterschiede eine wichtige Rolle.

Saffran (2001b) macht sich Gedanken, ob die mit den Übergangswahrscheinlichkeiten segmentierten Einheiten überhaupt weiter verwendet werden und in welcher Form sie als Repräsentationen zur Verfügung stehen. Würden die Kleinkinder den Output der statistischen Berechnungen als mögliche Wörter tatsächlich anerkennen, so müssten sie Wörter, die konform mit der statistischen Segmentierung sind, von Einheiten, welche die Wortgrenzen überschreiten (*part-word targets*) in englischen Sätzen und Kunstsätzen unterschiedlich behandeln. Nach einer entsprechenden Konditionierung sollte man erwarten, dass die statistisch korrekt segmentierten Wörter im englischen Satz bevorzugt werden, nicht aber im Kunstsatz. Diese Annahme konnte durch das erste Experiment bestätigt werden. In der Tat wurden bei den Kunstsätzen die Teilwörter besser erkannt. Dieser Unterschied ist allerdings nicht statistisch relevant. Das zweite Experiment kontrollierte die Fähigkeit der Probanden zwischen den Englischen Sätzen und Kunstsätzen unterscheiden zu können, was auch eindeutig der Fall war. Im dritten Experiment sollte schließlich noch eine weitere mögliche Informationsquelle, das Klangbild, kontrolliert werden. Analog zum zweiten Experiment stellte sich dafür keine Signifikanz heraus, so dass man getrost annehmen kann, dass die kindlichen statistischen Lernmechanismen wortähnliche Repräsentationen zur weiteren Verarbeitung in ihrer Muttersprache generieren (vgl. Saffran, 2001b, 166).

Diesen Befund kann man nun auch mit einem späteren Experiment in Verbindung bringen. Bortfeld et al. (2005) beschreiben drei Experimente, bei denen sechs Monate alte Babys mit der *Head-Turn-Preference-Procedure* auf ihre Fähigkeit getestet wurden, ein neues Wort zu erkennen, wenn es lediglich durch das Umfeld eines bekannten Wortes definiert ist. Dazu wurden zwei Szenarien konstruiert: Ein unbekanntes Wort tritt vor einem bekannten oder einem unbekanntem Namen auf. Das unbekannte Wort konnte nur dann erkannt werden, wenn es vom bekannten Namen gefolgt wurde. Im dritten Experiment sollte ausgeschlossen werden, dass die Segmentierung durch die Übergangswahrscheinlichkeiten geschehen ist. Wie sich zeigte, konnte man diese ausschließen. Offensichtlich mussten hier *Top-down*-Prozesse in Form von Abgleichen eine Rolle spielen.

4.5 Zusammenfassung

Das vierte Kapitel der Arbeit verfolgte zwei Ziele. Erstens griff es eine der am Anfang skizzierten Problemstellungen auf (Kap. 1.1) und versuchte die Qualität der im Input

4 Randbedingungen der Simulation

zugänglichen Informationen zu beurteilen. Zweitens wurden die Randbedingungen der Simulation auf die frühkindlichen Sprachperzeptionsfähigkeiten bezogen. Damit erfolgte eine Begründung der Annahmen des Modells und ihrer psychologischen Plausibilität.

Die Wahrnehmung sprachlicher Einheiten wurde in angemessener Tiefe diskutiert. Die Frage nach der qualitativen Repräsentation, ob Silbe oder Phonem als grundlegende Verarbeitungseinheit verwendet wird, konnte zwar nicht abschließend eindeutig, aber der Problematik angemessen, beantwortet werden. Dafür wurden psycholinguistische Experimente und deren Dispute in der Forschungsliteratur herangezogen. Dies geschah zudem in Hinblick auf Erwachsene und auf Babys unter einem Jahr.

Der hier dargelegte Wissensstand über die Perzeptionsfähigkeiten kann hinsichtlich der Modellentwicklung wie folgt zusammengefasst werden. Die kleinsten linguistischen Repräsentationseinheiten scheinen bei Erwachsenen einzelsprachabhängig unterschiedliche Ausprägungen anzunehmen. Bei Neugeborenen lässt sich dieser Umstand nur in wenigen Einzelfällen zeigen, jedoch nicht in einen direkten Zusammenhang miteinander oder mit dem zeitlichen Verlauf bringen, sodass sich die Entwicklung von einer universalen Einheit zu den einzelsprachlich beeinflussten Einheiten nicht nachweisen lässt. Im Grunde besteht auch hier einmal mehr das Problem, funktionale Einheiten aus dem Sprachstrom herauszulösen.

Aus der angestellten Diskussion folgt jedoch, dass Phoneme bedeutsame Vorteile einer internen Verarbeitung, gerade aus der Sichtweise einer möglicherweise universal gültigen und mit generellen Prinzipien zurückführbaren Theorie, aufweisen. Das Phonem, unabhängig vom Grad der Abstraktheit, besitzt den Vorteil, dass es auf jeden Fall wahrgenommen wird. Möglicherweise geschieht dies nur durch die umgebende Silbe.⁶⁴ Für die Silbe kann dies aber nicht gelten. Wie die Experimente von Anne Cutler und Kollegen (Kap. 4.1.2) belegen, werden die Silben eben nicht immer als Einheiten erkannt. Damit war klar, dass das Phonem als Repräsentationseinheit in der Simulation verwendet werden würde.

Die Fähigkeit für statistische Berechnungen, das logische Schließen und Erkennen von häufigen Phonemsequenzen, deren Erinnerung und Weiterverarbeitung als weitere Randbedingungen für die Simulation konnten in den dann folgenden Kapiteln begründet werden. Da die experimentellen Ergebnisse hier weniger Interpretationsspielraum ließen als für die sehr viel schwierigere Frage nach den kleinsten linguistischen Repräsentationseinheiten, blieben auch die wissenschaftlichen Kontroversen bislang eher beschränkt.

⁶⁴Pallier (1998) und Pallier et al. (1993) sprechen von einer Art Strukturierungsgitter, welches für das Französische zutreffend sein mag.

5 Modellierung einer stochastischen Sprachsegmentierung

5.1 Einleitende Bemerkungen

Die These dieser Arbeit wird mittels einer Computersimulation überprüft. Dazu ist es notwendig, die Architektur der Simulation aufzuzeigen und auf den Spracherwerb zu beziehen. Die Berechnungen, welche bei der Simulation anfallen, sind natürlich nur in einem abstrakt vereinfachten Sinn zu verstehen und auf den Spracherwerbsprozess zu übertragen.

... by using the term computation, we do not mean, of course, that infants are consciously performing a mathematical calculation, but rather that they might be sensitive to and able to store quantitative aspects of distributional information about a language corpus. (Aslin et al., 1998, 321)

Abbildung 5.1 zeigt, wie die Phasen des Algorithmus auf den Spracherwerbsprozess bezogen werden können. Diese Phasen sind stark abstrahiert und spiegeln die Wirklichkeit modellhaft wider. Die hohe Abstraktionsstufe rechtfertigt sich durch das Erkenntnisziel. Danach liegt das Interesse darin, den Einfluss von immer wiederkehrenden Lautstrukturen auf die Wortsegmentierung aufzudecken. Es ist notwendig, diesen Prozess einer stochastischen Informationsverarbeitung zu vereinfachen, also von anderen Sprachverarbeitungsprozessen zu isolieren, damit man den Einfluss der statistischen Informationen überhaupt eindeutig quantifizieren kann. Es sei an dieser Stelle nicht geleugnet, dass weitere kognitive Prozesse zur Wortsegmentierung eventuell schon parallel und in Beziehung zu den stochastisch orientierten Prozessen ablaufen können. Sie werden aber hier bewusst ausgelassen, um Überschneidungseffekte zu vermeiden. Da der Beitrag der stochastischen Informationen gemessen werden soll, müssen andere Einflussgrößen, wie Wort- und Satzmelodien, ausgeblendet werden.

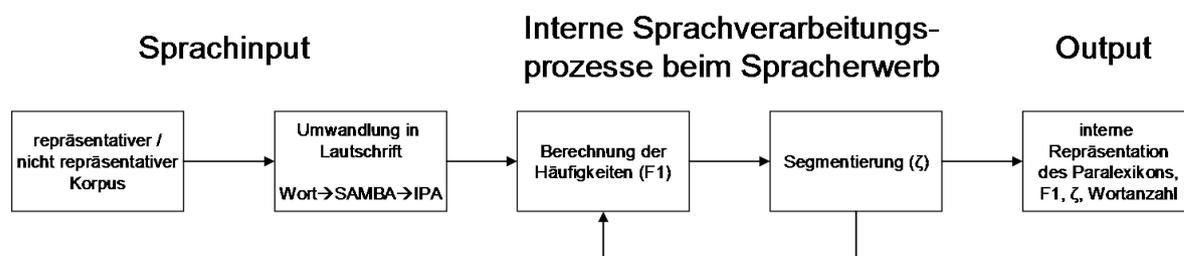


Abbildung 5.1: Abbildung der Simulation auf den Spracherwerbsprozess

5 Modellierung einer stochastischen Sprachsegmentierung

Aus Abbildung 5.1 ist die für die Informationsverarbeitung typische Dreiteilung von Input – Verarbeitung – Output zu entnehmen. Dabei interessiert das Verhältnis zwischen Input und Verarbeitung, dem internen Sprachverarbeitungsprozess. Der Output gibt dazu die notwendigen Informationen aus und ist nicht im Sinne eines realen Sprachoutputs zu verstehen.¹ Die genaue Form der dort vorliegenden Informationen ist unbekannt; sie werden als interne Repräsentationen lediglich betitelt. Die Sprachverarbeitung wird in Abbildung 5.3 noch detaillierter aufgeschlüsselt. Zunächst geht es an dieser Stelle darum, die Wirkung des Sprachinputs auf den Sprachverarbeitungsprozess zu verdeutlichen. Es gilt also, eine funktionale Beziehung zwischen der Sprachverarbeitung und dem Sprachinput in der Form:

$$v(i)$$

zu finden, wobei v für Sprachverarbeitung und die unabhängige Variable i für Sprachinput steht. Der Sprachinput wird durch einen Textkorpus repräsentiert. Den Sprachinput als manipulierbare Variable des Sprachverarbeitungsprozesses zu betrachten hat den Vorteil, dass die Wirkung des Sprachinputs eindeutig bestimmt werden kann. Das bedeutet also, dass jede Veränderung von i eine Änderung von $v(i)$ bewirkt und damit:

$$v(i_1) - v(i_2) \neq 0 \quad (5.1)$$

ist. 5.1 liefert zugleich ein quantitatives Maß zum Vergleich verschiedener Inputs. Gilt jedoch:

$$v(i_1) - v(i_2) = 0 \quad (5.2)$$

ist klar, dass der Sprachinput keine Auswirkung auf die Sprachverarbeitung hat. So wird sich beispielsweise feststellen lassen, ob repräsentative Korpora überhaupt für die Verteilung von Phonemen wichtig sind oder ob auch nicht-repräsentative Korpora diese Aufgabe erfüllen.

Der Begriff der Sprachverarbeitung selbst wird hier lediglich auf die distributionelle Verarbeitung der Phoneme zueinander verstanden. Dem Autor ist bewusst, dass die Sprachverarbeitung nicht auf die Berechnung von Häufigkeiten beschränkt ist, sondern dies nur eine Teilmenge von einer Vielzahl anderer Verarbeitungsprozesse darstellt. Für das Erkenntnisziel dieser Arbeit genügt es jedoch, den Sprachverarbeitungsprozess auf die Berechnung von Häufigkeiten und auf dieser Basis aufbauende rekursive Strukturen zu reduzieren.

Damit nur die rein verteilungsbedingten Eigenschaften der Phoneme wirken, wird die Informationsbreite des Sprachinputs eingeschränkt. Wichtige Segmentierungshilfen, die von Sprachlernern benutzt werden, aber nicht zur Abgrenzung eines Phonems gehören, sind Suprasegmentalia wie Akzente und Intonationen, aber auch Sprechtempi, Pausen

¹Die Begrifflichkeit wird hier analog zur chomskyschen Lerntheorie $LT(O, B)$, wobei O der Organismus und B ein bestimmter Bereich ist, der Input in der Analyse von B durch O besteht und

Der „Output“ (der natürlich intern und nicht offen und äußerlich repräsentiert ist) besteht in einer kognitiven Struktur einer gewissen Art. Diese kognitive Struktur ist eines der Elemente des durch O erreichten kognitiven Zustands. (Chomsky, 1977, 23f)

und Sprechrhythmen.² Diese Segmentierungsinstrumente werden in ihrer Wirkung ausgeschaltet, indem ihre Notation im Korpus überschrieben wird. Somit ist der Einfluss der distributionellen Eigenschaften der Phoneme auf den Sprachverarbeitungsprozess sicher gestellt.

In Abbildung 5.1 sind der Dreiteilung in Sprachinput, Sprachverarbeitung und Output die dazugehörigen Phasen des Algorithmus zugeordnet. Diese Phasen dienen zur Veranschaulichung der programmiertechnischen Abfolge des Algorithmus. Zum Sprachinput gehört ein repräsentativer Korpus, der in seinen Eigenschaften dem Sprachinput eines Säuglings während des ersten Lebensjahres entsprechen sollte. Unter Sprachverarbeitung subsumieren sich schließlich die in Kapitel 4 beschriebenen Randbedingungen: die Berechnung von Häufigkeitsverteilungen aus dem gegebenen Sprachmaterial und die Implementierung der rekursiven Struktur. Das Ergebnis der Berechnungen wird in der letzten Phase ausgegeben. Darunter fallen die Extraktion des Paralexikons, sowie die Ausgabe der Maßzahlen, Wortanzahl und der Häufigkeitsfunktion. In einem realen Sprachverarbeitungsprozess würden diese Repräsentationen nicht sichtbar werden, sondern weiteren kognitiven Prozessen unterzogen sein. Die hier ausgegebenen Repräsentationen sind, wie bereits ausgeführt, nur interne und modellhafte Repräsentationen und sollen nicht als bewusster Output des Kindes verstanden werden. Aus erkenntnistheoretischen Gründen werden sie vorzeitig visualisiert.

5.2 Bestimmung der Parameter und Maßzahlen

Um am Ende die aufgestellte Hypothese beweisen zu können, müssen objektive Kriterien zur Quantifizierung der Ergebnisse herangezogen werden. Die Ergebnisse lassen sich erst dann aussagekräftig interpretieren, wenn eine Vergleichsbasis vorhanden ist. Damit erfolgt eine kritische Abwägung auf zwei Ebenen. Erstens erfordert die Beantwortung der Frage, ob das Sprachsignal genügend Informationen zur Wortsegmentierung enthält, ein quantifizierbares Ergebnis, das sich nicht in eine tautologische Unantastbarkeit flüchten kann, sondern anhand eines quantifizierbaren empirischen Nachweises angreifbar und eventuell widerlegbar bleibt. Das ist für die Fragestellung dieser Arbeit zwingend. Gerade hier müssen objektive Kriterien angelegt werden, weil die Gefahr besteht, dass Modell derart anzupassen, dass es zwar die Hypothese bestätigt, aber an einer mangelnden Allgemeingültigkeit leidet. Zweitens ermöglichen objektive Messkriterien, insofern sie denn von anderen Wissenschaftlern gleichermaßen verwendet werden, einen Vergleich zu anderen Arbeiten dieser Problemstellung.³ Damit verschafft sich die Forschung an diesem Thema Zugang zu einer breit angelegten und fruchtbaren Diskussion. Der wissenschaftliche Dialog erlangt an dieser Stelle seinen höchsten Wert. Eine kritische Abwägung

²Der Einfluss dieser Prosodien wurde von Thiessen und Saffran (2003), Thiessen und Saffran (2004), Saffran und Thiessen (2003), Saffran (2003d) experimentell nachgewiesen. Jusczyk (1997) gibt eine zusammenfassende Darstellung der suprasegmentalen Zusammenhänge. In Jusczyk (1999) findet sich eine stark verkürzte Darstellung der Einflussgrößen auf die Wortsegmentierung.

³Gambell und Yang (2005, 9) äußern leichte Kritik, dass einige Autoren wohl nicht grundlos einem objektiven Vergleich ausweichen.

findet also einmal innerhalb der Arbeit statt (Belegen die Ergebnisse die Hypothese?), aber auch außerhalb im Vergleich mit anderen Forschungsergebnissen (Wie weichen die Ergebnisse voneinander ab, wenn Annahmen, Vorgehensweisen/Algorithmen verändert werden?).

5.2.1 Allgemeine Leistungsmessung in der Computerlinguistik

In der Computerlinguistik im Allgemeinen (Carstensen et al. (2004, 346f); Jurafsky und Martin (2000, 578)) werden übereinstimmend Präzision (*precision*) und Vollständigkeit (*recall*) als objektive Maße verwendet. Diese Maße fanden ursprünglich lediglich im Bereich des *Information Retrieval* Anwendung (Carstensen et al. (2004, 469 u. 487); Jurafsky und Martin (2000, 652); Mitkov (2003, 531)). Ihr Aussagegehalt ist aber für fast alle Bereiche der Computerlinguistik gegeben (vgl. Mitkov, 2003, 416ff).⁴ Allgemeine Definitionen der Termini werden wie folgt übernommen (Mitkov, 2003):

- *precision*: The number of correct responses divided by the total number of actual responses. (750)
- *recall*: The number of correct responses divided by the total number of possibly correct responses. (752)

Die Grundannahmen einer Bewertungsarchitektur sind in den computerlinguistischen Fragen sehr ähnlich und können deshalb problemlos auf die meisten Teilbereiche dieser Disziplin angewandt werden. So geht es im Prinzip immer um eine bewertbare Aussage, wie viel richtig und wie viel falsch hinsichtlich eines Absolutpunktes errechnet werden kann. Im *Information Retrieval* wird der Absolutpunkt mit der Anzahl aller passenden Dokumente definiert. Davon können von einer Suchmaschine einige Dokumente gefunden werden, die zu der definierten Menge gehören; andere mögen aber nicht Teil der definierten Menge sein. Dieser Umstand ist allgemein gültig und deshalb ist er leicht auf andere Teilbereiche in der Computerlinguistik übertragbar. In der Grammatik- oder Textkorrektur wird es ebenso Teilmengen einer definierten Menge geben wie auch Mengen außerhalb dieses Bereichs. Analoges gilt natürlich auch in den Bereichen der Textzusammenfassungen und Klassifikationsaufgaben oder etwa bei der Leistungserfassung von Lexikographiesystemen. Im Lichte dieser Erkenntnis kann nun eine allgemeine Formel dargelegt werden. Dazu eignet sich die mengentheoretische Beschreibung, weil hier das Abstraktionsvermögen der Mathematik die allgemeine Anwendbarkeit garantiert. Gegeben seien dazu folgende Mengendefinitionen:

- $A = \{\text{Alle Vorkommnisse des zu untersuchenden Phänomens}\};$
- $B = \{\text{Alle Ergebnisse der Simulation}\}$

Der Definitionsbereich der Menge A setzt also den Absolutpunkt fest, an dem gemessen werden soll. Die Elemente dieser Menge wurden mit einer allgemeinen Beschreibung versehen und sind ebendaher auf alle computerlinguistischen Bereiche anwendbar. Im spezi-

⁴Fallweise wird im Bereich der *Word Sense Disambiguation* nur die Präzision verwandt wie in Jurafsky und Martin (2000, 639) dargelegt wird.

ellen Fall können dies beispielsweise alle Syntaxfehler eines bestimmten Textes sein, alle Internetseiten mit dem Inhalt des Erstspracherwerbs oder alle gebundenen Morpheme eines Dokuments. In der Menge B sind alle Vorkommnisse enthalten, welche nach dem Algorithmus eines Programms kalkuliert werden konnten. Mit $A \cap B$ erhält man genau die Vorkommnisse, welche sowohl in A als auch in B vorhanden sind. Die Schnittmenge von A und B muss also alle richtig erfassten Phänomene als ihre Elemente beinhalten. Diese Schnittmenge allein ist aber als Maßzahl nicht aussagekräftig, weil sie in keiner Relation zu etwas steht. Eine Verbindung zur Mächtigkeit beider Mengen stellt ein solches Verhältnis her. Daraus ergeben sich auf einfache Weise die Maßzahlen für Präzision und Vollständigkeit in allgemeiner Form:

$$\text{Präzision} = \frac{\#(A \cap B)}{\#B} \quad (5.3)$$

$$\text{Vollständigkeit} = \frac{\#(A \cap B)}{\#A} \quad (5.4)$$

Die Präzision in (5.3) gibt also an, in welchem Verhältnis die richtig errechneten Ergebnisse eines Programms zu deren Gesamtzahl steht. Implizit ist darin $B \setminus A$ enthalten. Nimmt die Präzision sehr kleine Werte an, ist klar, dass $\#(B \setminus A) \ll \#B$ gilt, also das Verhältnis von richtigen und falschen Ergebnissen sehr groß und die Effizienz des Algorithmus entsprechend gering ist. Analog dazu ist die Vollständigkeit in (5.4) zu interpretieren. Das Verhältnis der richtig errechneten Ergebnisse bezieht sich nun aber direkt auf den definierten Absolutpunkt, nämlich alle objektiv richtigen Vorkommnisse des untersuchten Phänomens.

Die Aussagekraft der Vollständigkeit orientiert sich demnach an dem, was maximal noch zu erreichen wäre und die Präzision an dem Einfluss der tatsächlichen Fehler. Demgemäß eignet sich aber keines der Maße allein als ein objektives Maß. Beide Maßzahlen müssen zusammen betrachtet werden, um die Leistung eines Phänomens richtig zu bewerten. Da $\#(A \cap B) \geq \#A$ und $\#(A \cap B) \geq \#B$ liegt der Wertebereich von Präzision und Vollständigkeit immer zwischen 0 und 1. Sie können dementsprechend als Prozentzahlen dargestellt werden, wobei 100 % bei $(A \cap B) = (B \cap A)^5$ oder $A = B$ und 0 % bei $(A \cap B) = \{\}$ liegen.

Um die Datenfülle, die unweigerlich bei repräsentativen Studien aufkommt, anschaulich und übersichtlich zu gestalten, ist es sinnvoll, Präzision und Vollständigkeit in einem Maß zu vereinen. Bei Gleichgewichtung beider Maße eignet sich dazu das Harmonische Mittel, welches in der Computerlinguistik oft als F_1 oder $F_{0.5}$ -Maß⁶ bezeichnet wird. Es leitet sich aus der Formel

$$F_\alpha = \frac{1}{\alpha \frac{1}{p} + (1 - \alpha) \frac{1}{r}}; \text{ für } 0 \leq \alpha \leq 1, \quad (5.5)$$

⁵beziehungsweise formal korrekt: $A \supset B$ und $B \supset A$

⁶Ersteres findet sich beispielsweise mehrheitlich in der angelsächsischen Literatur wie bei Gambell und Yang (2005, 10), aber nicht ausschließlich wie Jurafsky und Martin (2000, 578) zeigen und das zweite vorrangig im deutschsprachigen Raum etwa bei Carstensen et al. (2004, 488).

5 Modellierung einer stochastischen Sprachsegmentierung

beziehungsweise

$$F_\beta = \frac{1 + \beta^2}{\frac{\beta^2}{r} + \frac{1}{p}}; \text{ für } \beta \geq 0, \quad (5.6)$$

her, wobei p für die Präzision (*precision*) und r für die Vollständigkeit (*recall*) steht. Aus beiden mathematischen Spielarten resultieren letztlich die unterschiedlich verwendeten Bezeichnungen für F_1 und $F_{0.5}$, da für eine Gleichgewichtung $\alpha = 0.5$ und $\beta = 1$ gewählt werden muss und dies an der Maßzahl kenntlich gemacht wird. Beim Einsetzen der entsprechenden Werte 1 und 0.5 lassen sich die Formeln vereinfachen und das Harmonische Mittel wird in seiner Grundstruktur sichtbar:

$$\bar{\chi}_{harm} = \frac{n}{\sum_{i=1}^n \frac{1}{\chi_i}}. \quad (5.7)$$

Da $n = 2$ und χ_i in (5.7) nur die Werte von p und r annimmt, lässt sich eine weitere Vereinfachung vornehmen:

$$F_\alpha = \bar{\chi}_{harm} = 2 \frac{pr}{p+r}. \quad (5.8)$$

Die unterschiedlichen Signaturen des F -Maßes in (5.5) und (5.6) sind natürlich problematisch. Im Falle von $\alpha = 1$ handelt es sich um eine Maßzahl, die die Vollständigkeit überhaupt nicht berücksichtigt. Mit $\beta = 1$ werden bei (5.6) Präzision und Vollständigkeit gleich gewichtet. Beide Varianten werden aber mit F_1 ausgewiesen, obwohl sie zwei völlig unterschiedliche Aussagen treffen. Folglich ist bei einem Vergleich in der Literatur mit gegebenen F -Werten Vorsicht geboten. In der vorliegenden Arbeit wird F_1 im Sinne einer Gleichgewichtung von Präzision und Vollständigkeit wie in Formel (5.6) verstanden.

5.2.2 Anwendung auf den Bereich der Wortsegmentierung

Für den Anwendungsfall der Wortsegmentierung können die entsprechenden Maßzahlen leicht eingepasst werden. Im Grunde genügt es, eine Menge als absolutes Kriterium zu bestimmen und die Folgeschritte ergeben sich schließlich aus der Deduktion der oben beschriebenen allgemeinen Form. Diese Menge ist mit der Problemstellung dieser Arbeit bereits vorgegeben. Dementsprechend lässt sie sich als die Menge der Lücken eines Textes beziehungsweise eines transkribierten Lautstroms eindeutig formulieren:

$$O = \{\text{alle Lücken eines Korpus}\}. \quad (5.9)$$

Die zweite Menge umfasst schließlich:

$$P = \{\text{alle Lücken des Korpus, die vom Programm markiert wurden}\} \quad (5.10)$$

Dabei gelten $\#O$ und $\#P$ jeweils als Divisoren der Maßzahlen:

$$\text{Gesamtzahl der Lücken des Korpus} = \text{Anzahl der Worttoken des Korpus} - 1. \quad (5.11)$$

(5.11) gibt die Mächtigkeit der Menge O in (5.9) beziehungsweise für den vom Algorithmus segmentierten Korpus für die Menge P (5.10) wieder. Die Mächtigkeit geht zwar letztendlich als ein Term in den Maßzahlen rechnerisch ein, allerdings sagt sie aber nur etwas über die Quantität, nichts aber über die Qualität der Elemente aus. Eine qualitative Aussage ist aber für die Kontrolle, ob die Lücken richtig markiert wurden, zwingend erforderlich. Die einzelnen Elemente werden programmintern mittels ihrer Position im Text kodiert. Aus algorithmischer Sicht bedeuten die Lücken des Korpus eine Reihe unterschiedlicher Elemente einer Menge, die aufgrund ihrer einmaligen Position im Text definiert werden, also von jeder anderen Lücke dieses Textes verschieden sind. Diese interne Repräsentation ist deshalb zwingend, weil nur so ein Vergleich mit den vom Programm ermittelten Lücken erfolgen kann. Richtig und falsch gesetzte Lücken werden immer nur in Hinsicht auf die Menge O identifiziert, also wenn die internen Repräsentationen in O und P gleich sind. Formal ausgedrückt, sind die richtig gesetzten Lücken gleich der Schnittmenge zwischen O und P . Durch einfaches Abzählen der Elemente in $O \cap P$ erhält man jeweils den Dividenten der beiden Maßzahlen. Schlussendlich können die Maßzahlen für die Wortsegmentierung zusammengefasst werden:

$$\text{Präzision} = \frac{\text{Anzahl der Lücken, die vom Algorithmus richtig gesetzt werden}}{\text{Anzahl der Lücken, die vom Algorithmus gesetzt werden}} \quad (5.12)$$

$$\text{Vollständigkeit} = \frac{\text{Anzahl der Lücken, die vom Algorithmus richtig gesetzt werden}}{\text{Anzahl der Lücken des Korpus}} \quad (5.13)$$

In den Mengennotationen von (5.9) und (5.10) entspricht (5.12) $\frac{\#(O \cap P)}{\#P}$ und (5.13) $\frac{\#(O \cap P)}{\#O}$. Ein Vergleich mit Brent (1999a, 94) und Gambell und Yang (2005, 9) zeigt, dass trotz der terminologischen Diffizilität, die Maße auf gleiche Weise berechnet werden.

5.2.3 Berechnung der Übergangswahrscheinlichkeiten und Lexikongröße

Die Übergangswahrscheinlichkeit im Allgemeinen ist ihrer mathematischen Natur nach eine bedingte Wahrscheinlichkeit und wird für zwei Ereignisse nach der Gleichung (5.14) berechnet.

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (5.14)$$

$P(A | B)$ ist die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung von B . $P(A \cap B)$ ist die Verbundwahrscheinlichkeit und sagt aus, mit welcher Wahrscheinlichkeit die Ereignisse A und B gemeinsam auftreten. Bezieht man die bedingten Wahrscheinlichkeiten nun auf den Anwendungsfall der Wortsegmentierung, bei der eine bestimmte Lautfolge (Ereignis A) unter einer bestimmten Wahrscheinlichkeit ($P(A | B)$) einer anderen Lautfolge (Ereignis B) folgt, dann spezifiziert man in der Regel auch die Terminologie und spricht nun von der Übergangswahrscheinlichkeit. Es ist demgemäß die Wahrscheinlichkeit, mit der eine Lautfolge einer anderen folgt. Bei der Wortsegmentierung sind die Ereignisse und deren Schnittmengen durch einfaches abzählen leicht

ermittelbar. Offensichtlich handelt es sich beim Abzählen um eine einfache Häufigkeitsverteilung, sodass man zur Berechnung der Übergangswahrscheinlichkeit abermals die Mächtigkeit der Mengen heranziehen kann. Das bedeutet formal, dass die Übergangswahrscheinlichkeit dem Verhältnis der beiden Häufigkeitsverteilungen der Ereignisse entspricht, wobei ein Ereignis die Anzahl einer bestimmten Lautfolge ist.

$$P(A | B) = \frac{\#(A \cap B)}{\#B}. \quad (5.15)$$

Ein weiteres Maß ist die Anzahl der Wörter, die ins Paralexikon geschrieben werden. Die Wortanzahl ergibt sich durch einfaches Abzählen der Einträge. Dabei ist die absolute Wortanzahl von der echten Wortanzahl zu unterscheiden. Die absolute Wortanzahl zählt jeden Eintrag ins Paralexikon ohne einen qualitativen Abgleich mit den vorhandenen Einträgen vorzunehmen. Die echte Wortanzahl zählt nur die neuen Zugänge und ist damit das bevorzugte Leistungsmaß, welches vorrangig angeführt werden wird. Beide Formen der Wortanzahl beinhalten auch alle *Token*, die Instanzen der Worttypen.

5.2.4 Spezifizierung der Parameter

Das Ziel dieser Arbeit ist, Aussagen über das Verhältnis zwischen Sprachinput und deren Verarbeitung, also über die Funktion $v(i)$, zu treffen. Aus diesem Grund müssen eine Reihe von Parametern (Variablen) eingeführt werden, mit deren Hilfe die Kontrolle des Sprachinputs i unter bestimmten Voraussetzungen von v ermöglicht wird. Zu den Parametern zählen Sensibilität (σ), Phonemkettenlänge (λ), Lexikonfunktion (ζ), Korpusgröße (ψ) und Anzahl der Perioden (ω). Diese Parameter wurden in primäre und sekundäre Variablen unterteilt. Primäre Variablen sind die Einflusskomponenten, welche die interne Struktur der Sprachverarbeitung (v) betreffen: Sensibilität, Phonemkettenlänge und Lexikonfunktion. Dabei ist noch wichtig anzumerken, dass die primären Variablen nicht im Sinne von Parametern in einer Universalgrammatik zu verstehen sind und sprachspezifische Informationen enthalten, die das Kleinkind in die richtige Richtung beim Sprachenlernen weisen. Die Länge der Phonemketten ist in der Mathematik der Übergangswahrscheinlichkeiten enthalten und deren Beobachtung bedeutet einzig und allein, dass ihr Einfluss auf das Ergebnis der Gleichung untersucht wird. Die Sensibilität beinhaltet ebenso wenig sprachspezifische Informationen. Sie zeigt nur, wie das Ergebnis ausgelegt werden müsste, um bei der Wortsegmentierung erfolgreich zu sein. Die Auslegung oder Interpretation beinhaltet dabei, ob eine Phonemkette zwei oder vielleicht fünf Mal in einer bestimmten Umgebung vorkommen muss, um als zusammenhängend erkannt zu werden. Die Lexikonfunktion gibt letztlich an, wie viele Einträge (Lautsequenzen) sich ein Kind merken würde. In diesen Fällen handelt es sich um Variablen, die bei anderen kognitiven Prozessen, bei denen derartige Berechnungen anfallen, ebenso Verwendung finden würden. Hier werden sie zunächst als Variablen verwendet, um ihre optimale Kombination zu ermitteln. Wenn diese ermittelt ist, verlieren sie ihren variablen Charakter und erhalten die optimalen Werte als ihre Grundeinstellungen. Darum werden sie dann nur noch als Parameter bezeichnet.

Die sekundären Variablen betreffen den Sprachinput i . Hierzu gehören die Korpusgröße (ψ) und die Anzahl der Perioden (ω). Wie noch zu erläutern sein wird, konnten die Art und der Inhalt des Korpus (Register und Zeit) aus methodischen Gründen nicht als kontrollierbare Variable des Inputs integriert werden.

Die funktionale Beziehung zwischen Sprachinput und Sprachverarbeitung kann nun semiformal erfolgen. Da i lediglich in ψ und ω definiert ist, diese sich aber auf σ , λ und ζ auswirken, lässt sich dieser Zusammenhang als eine Komposition von σ , λ , ζ und ψ , ω , die als interne Repräsentationen über ξ zu einer dritten Funktion (Output) ausgegeben werden können, darstellen.

$$\xi(\psi, \omega) := (v \circ i)(\psi, \omega) := v(i(\psi, \omega)) \quad (5.16)$$

Somit gilt $(\sigma, \lambda, \zeta) \circ i(\psi, \omega)$ und lässt sich aufschlüsseln als

$$\sigma(\psi, \omega)$$

$$\lambda(\psi, \omega)$$

$$\zeta(\psi, \omega)$$

Bei der Durchführung der Simulation und der Auswertung der ersten Ergebnisse konnte der Einfluss einer weiteren Variable entdeckt werden, die sich im Nachhinein nur schwer in das ursprüngliche Design einordnen ließ. Dieser Einflussfaktor, im folgenden mit μ bezeichnet, definiert den Zeitpunkt, ab wann oder ab welcher Periode das Paralexikon genutzt werden kann, also in die rekursive Struktur integriert wird beziehungsweise den Beginn der rekursiven Struktur mit zusätzlichen Informationen einleitet. Wie sich erst bei der Simulation gezeigt hat, beeinflusst diese Variable das Ergebnis entscheidend. μ gehört seinen Eigenschaften nach zu den primären Variablen, weil es eine Einstellung in v spezifiziert, andererseits ist sie aber vom Input (ψ, ω) unabhängig. Sie selbst nimmt Einfluss auf die Anzahl der Wörter im Paralexikon, sodass sie sich in dieser Hinsicht wie eine Inputvariable verhält. Aus diesem Grund sei μ als unabhängige Drittvariable klassifiziert, welche die Qualität des Paralexikons mitbestimmt. Nun seien die sekundären und anschließend die primären Variablen genauer vorgestellt.

5.2.4.1 Die sekundären Variablen

Die sekundären Variablen sind die unabhängigen Variablen und werden von außen vorgegeben. Die Korpusgröße ψ variiert zwischen Zeichenketten der Länge 25 bis 450.000 Zeichen. Die Anzahl der Durchläufe ω liegt zwischen 1 und 32 Perioden. Diese Werte sind aufgrund der zur Verfügung stehenden Korpora in dieser Höhe entstanden (siehe Kap. 5.3).

Nach Kenntnis des Autors gibt es nur eine große Sammlung für kindgerichtete Sprache, welche den Ansprüchen der Repräsentativität einigermaßen gerecht wird. Es handelt sich um das CHILDES-Projekt von MacWhinney (1995). Unter Benutzung des CLAN Tools sind die Daten zu Forschungszwecken im Internet frei zugänglich. Dort finden sich unter den Namen von Wissenschaftlern, die ihre Daten zur Verfügung stellen, Korpora mit

5 Modellierung einer stochastischen Sprachsegmentierung

Morgan	Mirinda	Dillon	Brooklyn, Henry, Xavier
c1-0901	d1-0904	f1-0828	f2-0827
c1-0917	d1-0925	f1-0910	v1-1115
c1-0930	d1-1026	f1-1013	v2-0903
c1-1027	d1-1106	f1-1022	
c1-1129	d1-1120	f1-1106	
c1-1207	d1-1224	f1-1318	
c1-1309	d1-1310	f1-1402	
c1-1329	d1-1324	f1-1424	
c1-1417	d1-1407	f1-1504	
	d1-1417		
	d1-1428		

Tabelle 5.1: Zuordnung der CHILDES-Korpora zu untersuchten Subjekten

kindgerichteter Sprache. Für die Zwecke der vorliegenden Arbeit wurde auf die Daten von Michael Brent zurückgegriffen, weil sie hier ab einem Alter von acht und neun Monaten nutzbar sind. In den Kategorien anderer Wissenschaftler ist das nur sehr vereinzelt der Fall, wohl deshalb weil deren Forschung in andere Zeitspannen fällt, vorrangig in die Zeit des aktiven Sprechens der Kinder und der sprachlichen Interaktion. Selbst die Sammlung Brents ist nicht erschöpfend. Dennoch stehen 15 Korpora für die benannte Zeitspanne zur Verfügung. Dieser Engpass hat dazu geführt, die Zeitspanne, wie weiter unten (Kap. 5.3) beschrieben, auszudehnen und auf Korpora auszuweiten, die auch die kindgerichtete Sprache an 15 Monate alte Babys beinhalten.⁷

Da die Quantität des Inputs als kontrollierbare Variable in die Untersuchung eingehen sollte, schieden weitere Korpora aufgrund ihrer geringen Größe aus. Als untere Grenze wurde eine Größe von 10.000 Zeichen festgesetzt, weil sich in anfänglichen Pilotversuchen gezeigt hatte, dass Korpora dieser Größe robust sind.⁸ Zudem entspricht diese Größe ungefähr dem Modalwert aller dort vorhandenen Korpora. Auf diese Weise wurde folgendes Datenmaterial zusammengestellt (Tab. 5.1).

Die Dateinamen kodieren das Alter des Kindes in der gewöhnlichen Notation in Monaten und Tagen. Vorrangig handelt es sich also um die Sprache, gerichtet an drei Kinder, welche ab dem achten Monat in Abständen von zwei Wochen bis einen Monat erhoben wurde.⁹ Jedes der Korpora umfasst eine zeitliche Dauer von etwa zwei Stunden

⁷Eine logistische Regression dieser Daten hat gezeigt, dass die Verteilung der Sprachlaute für die gesamte Zeitspanne gleich ist und somit die Daten als valide angesehen werden können.

⁸Die repräsentativen Ergebnisse der Simulation zeigten erst später, dass sich die Robustheit schon bei einer geringeren Größe beginnt einzustellen.

⁹Die rechte Spalte beinhaltet noch drei weitere Korpora von drei weiteren Kindern. Sie dienten ursprünglich lediglich dem Nachweis, dass die Lautverteilung innerhalb der Korpora für unterschiedliche Kinder unterschiedlichen Alters gleich sind. Im ursprünglichen Design sollte das Modell, das

Gesprächszeit. Um die Korpusgröße zu kontrollieren, wurde jeder Korpus auf 10.000 Zeichen gekürzt, d.h. das Ende des Dialogs wurde abgeschnitten. Leerzeichen, Punctuationen und Sonderzeichen wurden vorher gelöscht. Leichte Variationen traten dann auf, wenn die 10.000-Zeichen-Grenze in die Mitte eines Wortes fiel. In diesem Fall wurde das Wort bis zum Ende gezählt oder wenn die Distanz zum Wortanfang geringer war, das betreffende Wort komplett gelöscht.

Die Erstellung von Korpora größer als 10.000 Zeichen erfolgte durch Zusammenlegung der 10.000er Korpora. Für einen 20.000 Zeichen Korpus wurden also zwei, für einen 50.000 Korpus fünf der Ausgangskorpora in eine Textdatei kopiert. Für den 450.000-Zeichen Korpus sind die übrigen Korpora des „Brent-Ordners“ eingegangen, also auch Korpora mit weniger als 10.000 Zeichen.

Um Korpora kleiner als 10.000 Zeichen zu erhalten, konnte eine erweiterte Version des Algorithmus, welcher schon für die Standardisierung der 10.000-Zeichen Korpora Verwendung fand, genutzt werden. Die Erweiterung bestimmte das erste Wort des Korpus zufällig zwischen 1 und 10.000 („jeweilige Korpusgröße in Zeichen“ + 5). Danach fährt der Algorithmus wie oben beschrieben fort.

Korpora, für die $\psi < 5000$ gilt, sind den Korpora mit $\psi \geq 10.000$ zahlenmäßig überlegen, weil größere Veränderungen in der statistischen Verteilung der Phoneme bei kleineren Korpora eher zu erwarten sind als bei repräsentativen größeren Sammlungen. Durch die größere Anzahl der kleinen Korpora kann aber ein zuverlässiger Mittelwert errechnet werden, welcher die verteilungsbedingten Ausreißer glättet. So lassen sich die 17 untersuchten Korpusgrößen in Abhängigkeit von ihrer Zeichenanzahl mit einer streng konvexen Funktion (Abb. in Tab. 5.2) beschreiben. Da Korpora mit nur wenigen Zeichen nicht die gleiche Robustheit haben können, wie repräsentative Sammlungen, wurden die Messungen an verhältnismäßig vielen kleineren Korpora vorgenommen. Deshalb konnten die Verhaltensänderungen bei nur minimaler Zunahme der Korpusgrößen verfolgt und zusammenhängend dargestellt werden. Durch diese Vorgehensweise konnte auch die kritische Masse sehr viel genauer eingegrenzt werden, ab welchem Punkt die Quantität des Inputs ausreicht, um den Anforderungen einer robusten Segmentierung zu entsprechen.

Alle Korpora bis 10.000 Zeichen sind durch 32 einzelne Korpora (siehe Tab. 5.1) repräsentiert. Jeder dieser Korpora wird den Algorithmus, wie in Kapitel 5.4 noch zu beschreiben sein wird, durchlaufen und aus allen Ergebnissen werden schließlich durch einfache Mittelung der Messwerte die repräsentativen Werte dieser Gruppe errechnet. Analog dazu werden für größere Korpora repräsentative Messwerte kalkuliert. Die genaue Zahl der Korpusrepräsentanten ist in Tabelle 5.2 enthalten.

mit der kindgerichteten Sprache an Morgan, Miranda und Dillon „trainiert“ wurde, mit diesen drei Korpora getestet werden. In der Umsetzung geschah dies nach μ -Perioden. Das bedeutet, dass diese Korpora die Einfügungen des Paralexikons erhielten, um bei der Simulation mindestens 30 Korpora für die Bildung des Paralexikons zur Verfügung zu haben und zwei weitere um die Auswirkungen der Rekursion zu erkennen.

Korpusgröße in Zeichen	Korpusgröße in Wörtern (Spanne / Durchschnitt)	Stich- proben
25	3-8 / 5.5	32
50	8-14 / 11.7	32
100	20-29 / 25.2	32
150	23-44 / 35.9	32
250	51-72 / 61.5	32
500	116-141 / 125.4	32
1000	236-277 / 251.2	32
1500	331-399 / 370.7	32
2500	577-669 / 618.2	32
4000	926-1043 / 994.3	32
5000	1180-1315 / 1248.1	32
10.000	2376-2622 / 2501.4	32
20.000	4832-5032 / 4924.1	10
50.000	11760-12525 / 12183.3	4
100.000	22719	1
300.000	72426	1
450.000	112187	1

Anzahl Korpora

Korpusgröße

kleine Korpora

große Korpora

Tabelle 5.2: ψ , Anzahl der Stichproben und deren Zusammenhang

5.2.4.2 Die primären Variablen

Zur Herleitung der ersten primären Variable λ sei noch einmal auf Gleichung (5.15) verwiesen. Gleichung (5.15) beantwortet die Frage, wie groß die Wahrscheinlichkeit ist, dass sich eine Lautfolge A einer Lautfolge B direkt anschließt. Hier stellt sich zwangsweise die Frage nach den Eigenschaften der Lautfolge. Die Lautfolge ist zunächst ein unspezifizierter Begriff, der in einer Definition abgegrenzt werden muss. Insbesondere muss definiert werden, aus wie vielen diskreten Einheiten die Lautfolge besteht.¹⁰ Ziel ist es, aufgrund der Verteilung der diskreten Lauteinheiten, eine Gesetzmäßigkeit zu finden, welche die anderen Informationen im Lautstrom überhaupt erst nutzbar macht. Darum ist die einzig mögliche Ausprägung der Variablen die Quantität der enthaltenen diskreten Einheiten in ihrer gegebenen Reihenfolge oder genauer: die Anzahl der aufeinander folgenden Phoneme. Dieser Parameter muss also eine Vorstellung davon geben, wie die Anzahl der zusammenhängenden Phoneme (Länge einer Phonemkette) die Übergangswahrscheinlichkeiten beeinflusst. Der Gedanke dahinter ist natürlich der Informationsgehalt, der sich für den Sprachlerner in unterschiedlicher Höhe ergibt. Wie viele diskrete Lauteinheiten müsste das Kleinkind denn theoretisch zur Berechnung der Übergangswahrscheinlichkeiten heranziehen, damit es die maximale Informationsdichte in Gleichung (5.14) der Korpusstruktur erhält? Genügt es, wenn das Ereignis B aus einem einzigen Phonem besteht oder sollte es eine silbenähnliche Struktur mit zwei, drei oder mehr diskreten Einheiten besitzen? Die Antwort darauf kodiert die primäre Variable (λ , im Quellcode als *laenge* bezeichnet). In allen Arbeiten über dieses Thema wird im Vorfeld diese Variable entweder als Silbe (Gambell und Yang, 2005) oder als einzelnes Phonem (Elman (1990); Brent (1999b) oder Cairns et al. (1997)) festgelegt und entzieht sich somit jeglicher Kontrolle.

Die zweite primäre Variable (σ , im Quellcode mit *limit* bezeichnet) drückt einen Schwellenwert aus. Hinsichtlich der Wortsegmentierung geht man von der Beobachtung aus, dass sich in vielen Umgebungen die Übergangswahrscheinlichkeiten innerhalb eines Wortes von denen zwischen den Worten signifikant unterscheiden (Saffran et al., 1996b, 610).¹¹ Ab wann entscheidet sich aber ein Sprachlerner, ob es sich um eine Wortgrenze handelt oder nicht? Wie hoch oder niedrig müssen die Übergangswahrscheinlichkeiten also sein, damit sich das Kleinkind für eine Wortgrenze entscheidet? In der Regel werden hier die Minima der Wahrscheinlichkeitsfunktion automatisch im Algorithmus als Wortgrenze definiert. Das heißt, immer wenn

$$P(A_{n-1} | B_{n-1}) > P(A_n | B_n) < P(A_{n+1} | B_{n+1})$$

gilt, wird eine Lücke eingefügt. Dieser Vorgehensweise wird hier nicht gefolgt, da auch lokale Minima, die leicht als kleine Ausreißer beim Anblick der Gesamtfunktion erkennbar

¹⁰Wie bereits ausführlich dargelegt wurde, spielen die Gesamtheit der im Lautstrom enthaltenen Informationen, wie Prosodien, hier keine Rolle.

¹¹Strittig bleibt dabei aber, ob die Signifikanz groß genug ist, um als Segmentierungsstrategie tatsächlich von Kleinkindern genutzt zu werden. Die Kritiker werfen das Argument auf, dass die Abweichungen zu groß sind als das man daraus auf Wortgrenzen schließen könnte.

5 Modellierung einer stochastischen Sprachsegmentierung

sind, als eine Wortgrenze interpretiert werden. Denn das Hauptaugenmerk liegt wiederum, wie schon mehrfach angeführt, in einer kontrollierten Verwendung dieser Variablen.

Die Kontrolle über diese Variable ermöglicht, den Wert zu bestimmen, bei dem die Wortgrenzenerkennung maximal ist. Man kann auf eine analoge Erklärung zurückgreifen, wie dies auch schon bei der Länge der Phonemketten geschehen ist. σ entschlüsselt, wie oft die einzelnen Phonemketten im Verhältnis zueinander vorkommen müssen, damit eine Wortgrenze wahrscheinlich ist. Bei genauer Betrachtung ist diese zweite primäre Variable ein Maß für die Sensibilität hinsichtlich der Fähigkeiten der Mustererkennung des Kindes, denn sie sagt aus, wie oft ein Lautmuster in einer bestimmten Umgebung auftreten muss, um zu wissen, dass es zusammen (zu einem Wort) gehört oder nicht. Muss sich das Lautmuster in konstanter Umgebung sehr oft wiederholen ist die Sensibilität gering; reichen andererseits schon wenige Male aus, weist es doch auf eine hohe Sensibilität hin, da nur eine kleine Anzahl genügt, um einen Unterschied zu erkennen. Nach dieser Logik verhält sich die Sensibilität umgekehrt proportional zu den Übergangswahrscheinlichkeiten. Eine kleine Übergangswahrscheinlichkeit bedeutet eine hohe Sensibilität des Kindes gegenüber Veränderungen im Lautmuster und große Wahrscheinlichkeiten deuten auf eine kleine Sensibilität hin.

Die Sensibilität zeigt weiterhin, wie das Kind das Ergebnis am besten interpretieren sollte, um bei der Wortsegmentierung mit der Strategie der Übergangswahrscheinlichkeiten am erfolgreichsten zu sein.

Stellt man ein objektives Maß, wie das F_1 -Maß, in Abhängigkeit von σ und λ dar, kann mittels der partiellen Ableitungen ein Maximum berechnet werden, welches die ideale Kombination der beiden primären Variablen ausgibt und damit den Punkt zeigt, bei dem die maximale Information zur Wortsegmentierung auf Grundlage der Lautverteilung zur Verfügung steht. Diese Extrempunkte werden schließlich als Grundeinstellungen übernommen (siehe Kap. 6.3).

Es liegt in der mathematischen Natur der Übergangswahrscheinlichkeiten, dass die hier identifizierten primären Variablen die einzig sinnvollen Kontrollparameter für eine Periode ohne rekursive Strukturen sind. Die Länge und die Anzahl der Phonemketten sind die einzigen Größen, die das Ergebnis direkt beeinflussen können. Das sind die Eigenschaften von A und B in Gleichung (5.15). Dabei entspricht die Länge von B der Länge von A .¹² Die Anzahl der Phonemketten interessiert nicht absolut, sondern nur im Verhältnis zueinander. Dieses Verhältnis wird durch die Division mit der Mengen-

¹²Natürlich ist es möglich A und B unterschiedliche Längen zuzuordnen. Damit würden alle möglichen Kombinationen von aufeinander folgenden Phonemketten errechnet werden. In diesem Fall würde aber eine Anfangsbedingung nicht beachtet. In dieser Untersuchung sollen die in der Wissenschaft begründeten Fähigkeiten des Sprachlerner vorrangig sein. Die anstehende Komplexität der Berechnung aller Kombinationen wäre in der Tat für das Kind zeitlich nicht zu realisieren. Wie in der anfänglichen Betrachtung zur psychologischen Plausibilität ausführlich dargelegt wurde, muss diese Aufgabe in wenigen Wochen erfolgen, aber sich definitiv nicht auf mehrere Monate erstrecken. Aus diesem Grund erschien die Annahme, dass die Länge der Phonemketten von B gleich der Länge der Phonemketten von A ist, vertretbar.

mächtigkeit erreicht und schlägt sich so im Ergebnis des Leistungsmaßes nieder. Damit braucht die Anzahl nicht gesondert beobachtet oder kontrolliert zu werden.

Erweitert man das „Einperiodenmodell“ auf mehrere Durchläufe ω , indem die Vorgängerperiode auch immer relevante Informationen an die Folgeperiode weiterleitet, ist noch eine weitere primäre Kontrollvariable notwendig. Diese Variable muss angeben, wie viel Information in der Folgeperiode übernommen werden soll. Die Lexikonfunktion ζ übernimmt diese Manipulation. Sie kodiert die Anzahl der Einträge, welche sich ein Kind merken und in der Folgeperiode zur Segmentierung nutzen könnte. Dabei kann sie komplexe funktionale Verläufe in Abhängigkeit der Perioden annehmen. Beispielsweise kann sie mit jeder Periode mehr oder weniger Einträge übertragen beziehungsweise linear, exponentiell, logarithmisch oder trigonometrisch wachsen oder fallen. In Abhängigkeit von der Korpusgröße werden die Einträge ins Paralexikon sehr hoch oder eher gering (aufgrund zahlreicher Doppelzählungen) ausfallen. Die Ergebnisse der Simulation zeigten jedoch recht überraschend bei allen ψ gleichermaßen, dass die besten Segmentierungsergebnisse erzielt werden, wenn ζ konstant gehalten wird, d.h. wenn in jeder Periode gleich viele Einträge in das Paralexikon aufgenommen werden.

5.3 Zusammenstellung der Korpora und Probleme der Repräsentativität

Für die Modellierung des Sprachinputs muss ein Korpus nach bestimmten Gesichtspunkten zusammengestellt werden. Das theoretische Rüstzeug für dieses Vorgehen ist den Arbeiten von Biber (1993), Biber (2004), Church (1993), McEnery und Wilson (1996) und Kennedy (1999) ausführlich beschrieben.¹³ Danach steht am Anfang einer jeden korpuslinguistischen Untersuchung die Frage nach dem Verhältnis zwischen Untersuchungsgegenstand und Korpus. Hier ist der Untersuchungsgegenstand die Sprache, die das Kind im ersten Lebensjahr hört. Genauer könnte man das Sprachmaterial natürlich zwischen dem sechsten und dem achten Lebensmonat vorgeben, weil genau dieser Zeitraum vorrangig für die Berechnung von Übergangswahrscheinlichkeiten experimentell ermittelt wurde. Andererseits darf man nicht zwingend davon ausgehen, dass der Sprachlerner nur das Material dieser Zeitspanne zur Verfügung hat. Es ist möglich, dass bereits Sprachfragmente aus früheren Monaten zur mentalen Verarbeitung zur Verfügung stehen. Zudem würden starke methodische Probleme auftreten, weil repräsentative Daten für einen derart spezifischen Zeitraum nicht zugänglich sind. Aus diesem Grund definieren wir den Untersuchungsgegenstand etwas weitläufiger und setzen ihn mit dem Input gleich, den das Kind erreicht, bevor es seinen eigenen sprachlichen Output erzeugt.

Tabelle 5.2 auf Seite 124 bildet bereits das Ergebnis eines Prozesses ab, der von Biber (1993) zur Erstellung eines repräsentativen Korpus beschrieben wurde (Abb. 5.2). Diese Darstellung ist allerdings eher eine Illustration als eine endgültige Lösung und zeigt

¹³Als kurze Beschreibung für die hier gewählte Vorgehensweise siehe auch Paprotté (2005, 368).

5 Modellierung einer stochastischen Sprachsegmentierung

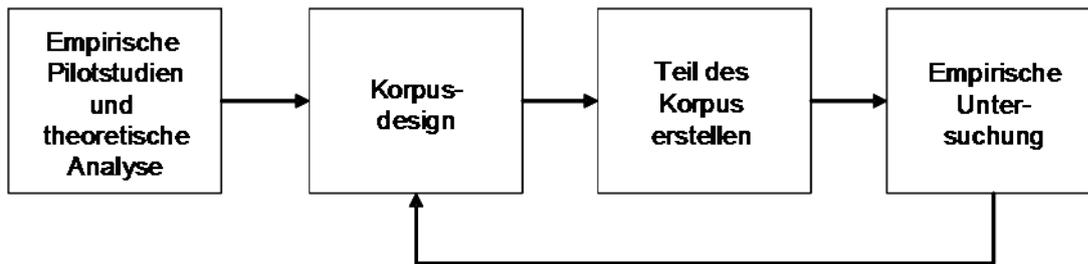


Abbildung 5.2: Entwicklung eines repräsentativen Korpus, Quelle: Biber (1993, 256)

wie ein Korpus idealerweise entworfen werden sollte (Biber, 1993, 244). Dabei wird die Repräsentativität eines Korpus in zwei Dimensionen gemessen (Biber, 1993, 243):

- a Anzahl der Texttypen in einer Sprache und
- b Anzahl von linguistischen Verteilungen einer Sprache.

Unter dem ersten Punkt werden die situationsbedingten Textkategorien (z. B. Roman, Sport, Nachrichten, etc.) verstanden. Unter dem zweiten Punkt subsumiert er die linguistisch definierten Texttypen. Letztere entstehen „on the basis of shared linguistic co-occurrence patterns“ (Biber, 1993, 245), sodass die Texte hinsichtlich ihrer linguistischen Eigenschaften große Gemeinsamkeiten aufweisen und danach klassifiziert werden können. Die linguistischen Texttypen basieren also auf internen Kriterien, während die situationsbedingten Texttypen beziehungsweise Textkategorien äußerlichen Kriterien gehorchen. Um mehr Klarheit in die begriffliche Vielfalt zu bringen, nennt Biber die situationsbedingten Textkategorien Register oder Genres.

Folgt man nun Bibers Vorschlag zur Entwicklung eines repräsentativen Korpus (Abb. 5.2), sind zuerst eine theoretische Analyse und erste explorative Untersuchungen notwendig. Erst dann geht man zur Zusammenstellung der Texte über. Dazu wird ein Teil des Korpus zusammengestellt und letztlich dann durch eine empirische Untersuchung überprüft. Weichen die empirischen Befunde zu stark von den Eigenschaften des erstellten Korpus ab, muss die Zusammenstellung des Korpus geändert werden. Dies geschieht so lange bis die Ergebnisse der Stichprobe nahe genug bei den empirischen Ergebnissen liegen. Weil die einzelnen Parameter eines repräsentativen Korpus nicht von Anfang an genau bestimmt werden kann, unterläuft die Arbeit am Korpus in Zyklen (Biber, 1993, 256). Mit fortschreitenden Zyklen erhöht sich dann die Repräsentativität des Korpus. Diese Vorgehensweise geschieht immer im Hinblick auf die oben genannten Dimensionen der Repräsentativität.

Die empirische Untersuchung umfasst den Abgleich der Korpusdaten mit der Realität, d.h. inwieweit das Korpus alle für die Untersuchung relevanten Eigenschaften beinhaltet. Ein Korpus, der beispielsweise die englische Schriftsprache repräsentiert, muss nicht nur die verschiedenen Genres, typische Satzkonstruktionen oder Wörter der gesamten englischen Schriftsprache berücksichtigen, sondern auch deren richtige Verteilung beziehungsweise Anzahl im Verhältnis zur Korpusgröße. Die empirische Überprüfung dient

also dazu, die Diskrepanz zwischen der Grundgesamtheit (englische Schriftsprache) und der „Stichprobe“ (Korpus) daraus zu verkleinern, sodass sich die Stichprobe der Repräsentativität nähert.

Da Neugeborenen von Geburt an kein sprachlicher Stimulus entzogen wird und diese Sprache sich somit nicht eingrenzen lässt, müsste man vernünftiger Weise die gesamte Sprache als Untersuchungsgegenstand akzeptieren.¹⁴ Sicherlich werden Neugeborene in höherem Maße der Mutter und deren direkter, meist veränderter Sprache, ausgesetzt. Dennoch lässt es sich kaum vermeiden, dass auch Sprachmaterial von außerhalb zum Kind gelangt.¹⁵ Selbst die Sprache der Mutter lässt sich nicht vollständig in ein Register einordnen, da sie nicht nur Kontaktperson zum Kind, sondern selbst in das soziale Umfeld der Familie und Freunde eingebettet ist und damit verschiedene „Sprachen“ in Gegenwart des Kindes annimmt (*code switching*).

Given this constant need to switch among registers, the acquisition of register characteristics is fundamentally important at all developmental stages: whether for a child learning that the language used with playmates is different from the language used with adults; or for an elementary school student discovering how written narratives are different from conversation; or a college student discovering how term papers in biology are different from written personal narratives; or for ESL students attempting to control the range of registers in their second language (Biber, 2004, 136).

An verschiedenen sozialen Schnittpunkten ist eine Vielzahl sprachlicher Stimuli aus unterschiedlichen Registern unvermeidlich. Diese Stimuli können keiner genauen Verteilungsgesetzmäßigkeit unterzogen werden, da die soziale Umgebung, in der Kinder aufwachsen, höchst unterschiedlich ist.

Dies kann zu Repräsentativitätsproblemen führen.

if a corpus does not represent the range of text types [registers] in a population, it will not represent the range of linguistic distributions. (Biber, 1993, 243)

Im Grunde sind lediglich die linguistischen Verteilungen von Lautfolgen in der vorliegenden Arbeit interessant. Die Repräsentativität dieser Lautfolgen könnte aber direkt von der richtigen Anzahl der verwendeten Register abhängen. Wie gerade erläutert wurde, sind der Einfluss unterschiedlicher Soziolekte und Sprachstile auf das Kind nicht eindeutig quantifizierbar. Man kann also nicht genau sagen, wie viel die einzelnen Register Teil der gehörten Sprache des Kindes ausmachen.

Im Zusammenhang damit tritt auch die große Masse des Datenmaterials als Problemquelle auf, da die Grundgesamtheit die komplette Sprache umfassen könnte. Die Repräsentativität ist für die Validität der Forschungsergebnisse von hoher Bedeutung.¹⁶

Für publizierte Texte sind Lösungsansätze im Brown- und LOB-Korpus bereits er-

¹⁴vgl. Fodor und Crowther (2002, 108).

¹⁵vgl. Biber (1994, 188ff).

¹⁶Biber (1993, 243):

a corpus must be 'representative' in order to be appropriately used as the basis for generalizations concerning a language as a whole

folgreich umgesetzt worden. Hier wurde eine zeitliche Grenze zu Grunde gelegt. Eine zeitliche Grenze wird auch für den hier verwendeten Korpus angesetzt. Sie kann aber nicht mit einem Datum absolut fixiert werden, wie bei Brown oder LOB, da nicht alle Kinder zum gleichen Zeitpunkt geboren werden. Die Zeitspanne beträgt 15 Monate (fix), die sich aber von einem absoluten Zeitpunkt beliebig entfernen kann.

Die Überprüfung mittels empirischer Daten (gemäß Abbildung 5.2) konnte hier nicht vollzogen werden. Aus den eben dargestellten Gründen gestaltet sich eine empirische Überprüfung äußerst schwierig. Sie hätte den Umfang eines gesonderten Forschungsprojekts. Derzeit muss davon ausgegangen werden, dass der linguistische Input die kindgerichtete Sprache der Mutter ist¹⁷ und von den Daten in CHILDES repräsentativ wiedergegeben wird. Die oben unter (a) spezifizierte Dimension nach Biber (1993) muss unberücksichtigt bleiben. Nach Jusczyk (1997) ist zwar bekannt, dass Säuglinge durchaus in der Lage sind, Sprachregister zu unterscheiden,¹⁸ jedoch sind die Auswirkungen fehlender Register unbekannt.

to the present time, there have been no investigations of whether changes in speech register might affect infants' recognition of particular sound patterns. (Jusczyk, 1997, 69)

Dieser Umstand wird auch hier vorausgesetzt.

Die zweite Korpusressource, welche für die Simulation benötigt wird, ist ein Lautsprachelexikon. Die Relation zwischen Schriftsprache und Aussprache ist im Englischen nicht vorhersagbar, d.h. es gibt keine Regel, die jedem Phonem eindeutig ein Graphem zuordnet. Weil $\exists z \mid \neg(x \rightarrow z, z \rightarrow y \Rightarrow x \leftrightarrow y)$, wobei $x = \text{Graphem}$ und $y = \text{Phonem}$, kann kein Algorithmus die Zuordnung von Phonem zum Graphem in jedem Umfeld eindeutig vollziehen. Eine Zuordnung auf Wortebene ist aus diesem Grund ebenso wenig möglich. Allerdings ist eine Abbildung mit Hilfe einer Zuordnungsvorschrift für jedes Wort oder Phonem möglich. Die Zuordnungsvorschrift ist nicht vorhersagbar oder als allgemeine Regel formulierbar; sie ist aber im Lautsprachlexikon als Zuordnung kodiert. Die *Carnegie Mellon University* stellt das *CMU-Pronouncing Dictionary* als Download bereit. Es beinhaltet 125.000 Einträge des Englischen mit den dazugehörigen Transkriptionen und unterscheidet sich von konventionellen Lexika darin, dass es auch alle Flexionsvarianten eines Wortes als eigene Lexeme auflistet. Die Ausspracheeinträge liegen in der ASCII lesbaren SAMBA Notation vor. Diese Notation unterscheidet 39 Phoneme des Englischen, davon zehn Vokale und fünf Diphthongnotationen (siehe Anhang C).

5.4 Beschreibung des Algorithmus als erklärenden Ansatz

Eine noch unspezifische Formulierung, wie ein Segmentierungsablauf von statten gehen könnte, findet sich überraschender Weise schon bei Pinker:

Perhaps segmentation, for example, can be accomplished in part by a combination of stochastic methods that tabulate syllable transition probabilities . . . Alternatively, (or perhaps in addition), the child could store auditory templates of words or short phrases heard

¹⁷Falk (vgl. 2004), Bortfeld (2004), Dominey und Dodane (2004), Kuhl (2000, 11854) und Thiessen et al. (2005) als Argumente für die Verwendung von kindgerichtete Sprache.

¹⁸Und dies auch in Verbindung mit visueller Information (Hollich et al., 2005).

in isolation, then match these templates against the input stream, recursively defining as new words those series of segments found in between the recognized portions of the stream. (Pinker, 1984, 28)

Pinker hat hier schon einige der experimentellen Ergebnisse gut zehn Jahre zuvor erahnt. Ihm fehlten seiner Zeit aber noch die genauen Spezifikationen, um seine Annahme schlüssig herleiten zu können. So blieb der doch recht einleuchtende Gedanke als bloße Behauptung stehen. Die Idee des Rekursionsmechanismus mit gespeicherten Wörtern oder Satzfragmenten ist auch in der vorliegenden Simulation vorhanden. Allerdings besteht der entscheidende Unterschied darin, einen Zusammenhang zwischen den Übergangswahrscheinlichkeiten und dem Rekursionsmechanismus zu finden und damit den Fehlschluss, wie in Kapitel 3.3 diskutiert – die frühkindlichen Sprachlerner würden isoliert auftretende Wörter als Ausgangspunkt benutzen – zu vermeiden.

5.4.1 Randbedingungen und schematischer Ablauf

Im Kapitel 4 wurden die kognitiven Fähigkeiten von Kleinkindern, die in die Simulation als Randbedingungen eingehen, untersucht. Aus der Gesamtheit der kognitiven Fähigkeiten, die einem Kleinkind von acht Monaten zur Verfügung stehen, werden fünf für die Arbeitsweise des Algorithmus verwendet. Das ist zugleich die kleinstmögliche Anzahl, die als Ausgangsbedingungen für die Funktionsweise des Algorithmus notwendig sind. Die primären Variablen, welche diese Fähigkeit kontrollieren, sind ebenfalls aufgeführt.

- a Die Phoneme (beziehungsweise eine ähnliche Einheit, die über lautliche Merkmale definiert wird) der Sprache können als diskrete Informationseinheiten aufgefasst werden (Brent (1999b); Jacobs (2003); Pisoni und Luce (1987); Swingley (2005); Kap. 4.1), λ
- b Die Übergangswahrscheinlichkeiten von einzelnen Phonemketten können berechnet werden (Saffran et al. (1996a); Kap. 4.2), σ
- c Die häufigsten Einheiten (immer wiederkehrende Lautketten) können erkannt werden (Shi et al. (2006); Jusczyk et al. (1994); Kap. 4.2 und 3.2.3), ζ
- d Die Lautsequenzen können über einen gewissen Zeitraum gespeichert und erinnert werden (Jusczyk und Aslin (1995); Jusczyk und Hohne (1997); Kap. 4.3), ζ und μ
- e Die lautlichen Repräsentationen können mit einem *Top-down*-Mechanismus im Sprachstrom eingebracht werden (Bortfeld et al. (2005); Saffran (2001b); Kap. 4.4), rekursive Struktur

Es ist wichtig zu erkennen, dass hinter diesen fünf Fähigkeiten generelle Heuristiken stehen, die hier speziell auf die Sprache Anwendung finden (vgl. Tomasello, 2003, 4). Sie bilden nun die Grundlage des Algorithmus, der vereinfacht im Folgenden (Kap. 5.4.2) beschrieben wird (die Buchstaben in Abb. 5.3 entsprechen den oben aufgelisteten Fähigkeiten; die Nummern der Ordnung der Programmabfolge). Die vier Phasen des Algorithmus werden zunächst in ihrem Zusammenhang und danach einzeln erklärt. Bezug

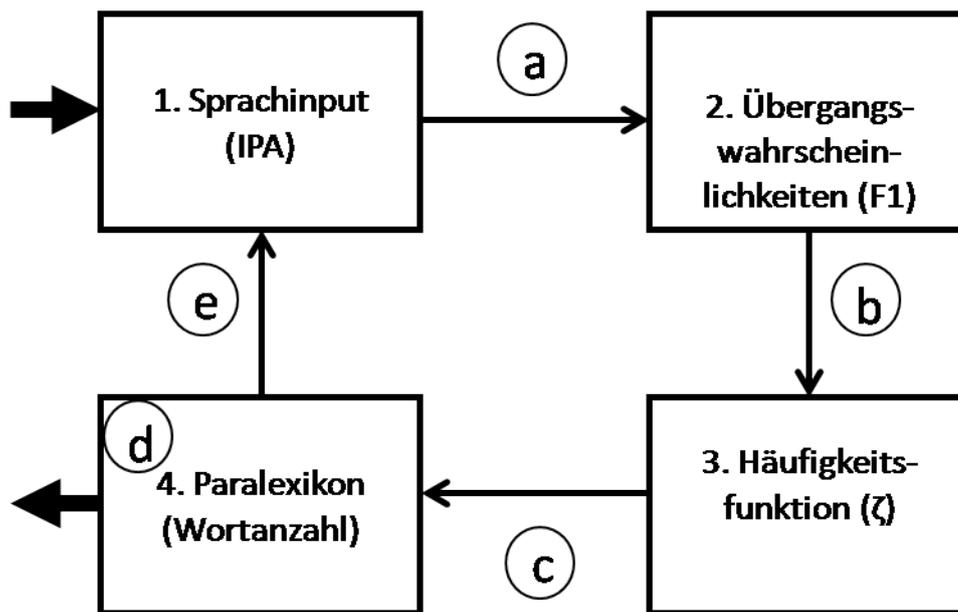


Abbildung 5.3: Allgemeine Funktionsweise des Algorithmus

nehmend auf die bisherige Forschung folgt letztlich eine Diskussion und eine kritische Abwägung zu rechentechnischen Einschränkungen.

5.4.2 Zusammenhängende Darstellung

Der sprachverarbeitende Prozess beginnt mit Phase 1, dem Sprachinput. In einer realen Umgebung trifft hier das Sprachsignal auf die Hörzellen, die es zum auditorischen Kortex weiterleiten und von dort in einer neuronalen Vernetzung repräsentieren. Die Vernetzung entspricht im verallgemeinerten Sinn dem, was aus Gründen der Vereinfachung und Veranschaulichung mit Phonem bezeichnet wird. Diese neuronale Repräsentation eines Phonems bezieht sich auf die Fähigkeit (a) des Kleinkindes, ein bestimmtes Lautkonstrukt neuronal abzubilden und als solches zu erkennen. Aus diesen Repräsentationen errechnet das Kind nun in der zweiten Phase die Wahrscheinlichkeit der Abfolge einzelner Laute oder Lautketten. Das Kind nutzt dazu die Fähigkeit (b). Im Grunde tut es nichts weiter als ungefähr zu schätzen, wie oft ein bestimmter Laut oder eine bestimmte Ansammlung von Lauten vorkommt. Im mathematischen Sinne kann man hier durch einfaches Abzählen von der Anzahl aller möglichen Laute oder von zusammenhängenden Lautkonstrukten sprechen.

Dadurch, dass das Kind das Vorkommen mehrerer möglicher Laute oder Lautkonstrukte kennt, weiß es auch, wie sie sich zueinander verhalten. Also kommt zum Beispiel die Konstruktion */ga/* viel öfter vor als die Konstruktion */xi/* oder etwa genauso oft wie */ba/*. Treffen nun Lautkombinationen aufeinander, die im Verhältnis zu anderen sehr selten sind, fällt dem Kind diese Besonderheit natürlich auf. Hier wäre ein erster Anhaltspunkt zur Lautsegmentierung vorhanden, nämlich anzunehmen, dass an dieser

„besonderen“ Stelle eine Einheit zu Ende ist und eine neue anfängt. Gruppiert das Kind nun die Einheiten nach diesem Kriterium, entstehen voneinander unabhängige Lautsegmente.

Es ist durchaus mit neurophysiologischen Erkenntnissen vereinbar, dass sich das Kind nach der Häufigkeit richtet, weil sich auch genau diese Häufigkeit in der neuronalen Repräsentation durch das gleichzeitige Feuern der Neuronen widerspiegelt (vgl. Yost, 2007, 229). Je öfter ein bestimmter Reiz auf ein Nervenzellenbündel trifft (und das sind hier die frequenzsensiblen Hörzellen), je stärker bilden sich die Synapsen mit den verbundenen Neuronen heraus und repräsentieren eben genau dieses Konstrukt bevorzugt. Das gleiche Prinzip der Häufigkeit unterliegt natürlich auch den Repräsentationen eines Sprachsignals. Im Modell wird dies in der dritten Phase mit der Fähigkeit (c) simuliert. Da nun unablässig verschiedene Sprachsignale vom Kleinkind aufgenommen werden, erhöhen sich auch die Repräsentationen der verschiedenen Lautkonstrukte, wovon wiederum nur die häufigsten eine entsprechende neuronale Repräsentation haben können. Bei allen Kombinationen, für welche die dafür sensiblen Nervenzellen keinen Reiz erhalten haben, können sich die Synapsen nicht derart verstärken wie bei Repräsentationen, deren Zellen ständig feuern. Genau diese Repräsentationen von neuronalen Mustern, werden hier als Paralexikon bezeichnet. Diese Repräsentationen auch über einen längeren Zeitraum aufrecht zu erhalten, ist hier mit der Fähigkeit (d) kodiert.

Schließlich können nun diese Informationen des Paralexikons mit der Fähigkeit (e) in einem unbekanntem Sprachsignal genutzt werden. Dadurch entstehen wiederum neue Informationen, die allein mit den Übergangswahrscheinlichkeiten nicht errechnet werden können. Im Prinzip handelt es sich hierbei um Informationen, die in einem *Top-down*-Verfahren gewonnen, selbst aber durch einen reinen *Bottom-up*-Prozess ausgelöst wurden. Die neuen Informationen können aus dem Paralexikon selbst erbracht werden, da es nun durch einfache Analogieschlüsse möglich ist Gesetzmäßigkeiten aufzustellen,¹⁹ etwa über die prosodischen Eigenschaften, die dann wiederum als zusätzlicher Anhaltspunkt bei der Segmentierung verwendet werden können. Nun entsteht aber schon bei einer einfachen Anwendung des Paralexikons wieder neue Information, welche für die Lautsegmentierung relevant ist, allein dadurch, dass durch die Einbringung von Lautsequenzen aus dem Paralexikon wieder neue Einheiten entstehen, die dann auch wieder dem Paralexikon zugeführt werden. So entwickelt sich ein rekursiver Prozess, solange bis die Informationsausbeute zu niedrig wird. Das Endprodukt ist nach n Durchläufen, dann eine lautliche Repräsentation von Worteinheiten, mit dem alle Sprachsignale segmentiert werden können und das dann schließlich auch in späteren Entwicklungsstufen mit semantischem Gehalt gefüttert, die symbolische Verknüpfung von Lautbild und Bedeutung, vollziehen kann.

¹⁹Möglicherweise auf Grundlage dergleichen Prinzipien, wie sie von Holyoak und Thagard (1995, 81ff) oder Clark (2004) und für etwas ältere Kinder (94ff) beziehungsweise in Gentner und Medina (1998, 183ff) oder Goswami (1992, 100ff) für die Begriffserkennung beschrieben werden.

5.4.3 Der Sprachinput

Die Korpora und das *CMU*-Aussprachelexikon wurden wie in den Kapiteln 5.2.4.1 und 5.3 beschrieben, als Textdateien hinterlegt und in entsprechende Verzeichnisse abgelegt. Die für den Sprachinput bedeutsamen Verarbeitungsschritte wurden auch dort vereinfacht bereits genannt: Korpus einlesen, Umwandlung in eine phonetische Repräsentation und Entfernen der Lücken, Satz- und Sonderzeichen sowie die Umwandlung der Diphtonge zu einer Einzeichenrepräsentation.

Programmiertechnisch sind dafür die Methoden der Klasse *IPA.java* bedeutsam. Diese Klasse enthält eine ganze Reihe statischer Methoden, die nicht alle für die Simulation allein notwendig sind. Einige Methoden sind beispielsweise für eine automatisierte Überprüfung der Ergebnisse, anfängliche Pilotstudien mit und ohne Lautsprache, zuständig. Die relevanten Methoden und deren Arbeitsweise sollen hier kurz vorgestellt werden. Der vollständige Quellcode befindet sich im Anhang A.

```

8  static String cmudict = "";
24 static String[] cmuChars = {"AA", "AE", "AH", "AO", "AW", "AY", "CH",
25                             "DH", "EH", "ER", "EY", "HH", "IH", "IY",
26                             "JH", "NG", "OW", "OY", "SH", "TH", "UH",
27                             "UW", "ZH", "B", "D", "F", "G", "K", "L",
28                             "M", "N", "P", "R", "S", "T", "V", "W",
29                             "Y", "Z"};
30 static char[] ipaChars = {0x252, 0x259, 0x28C, 0x254, 0x263A,
31                             0x263B, 0x2A7, 0x0F0, 0x065, 0x25C,
32                             0x263C, 0x068, 0x26A, 0x069, 0x2A4,
33                             0x14B, 0x2640, 0x2642, 0x283, 0x3B8,
34                             0x28A, 0x075, 0x292, 0x062, 0x064,
35                             0x066, 0x067, 0x06B, 0x06C, 0x06D,
36                             0x06E, 0x070, 0x072, 0x073, 0x074,
37                             0x076, 0x077, 0x06A, 0x07A};
38
39 static String thisData;

```

Quellcode 5.1: Variablenbelegung (Klasse *IPA.java*)

Nach der statischen Initialisierung der Klasse werden der Inhalt des *CMU Pronouncing Dictionary* der Stringvariable *cmudict* zugeordnet und weitere globale Variablen belegt. Der Stringvariable *cmuChars* wird die *SAMBA*-Notation zugeordnet. Für die Zuordnung der *IPA* Notation genügt hingegen eine Variable vom Typ *char*, weil diese Zeichen als Hexadezimal in *Unicode*format eindeutig definiert sind. Die Stringvariable *thisData* wird die in *IPA* umgewandelte Textdatei enthalten.

```

9  static
10  {
11      int cnt=0;

```

```

12   String line = "";
13   try
14   {
15       FileReader fr = new FileReader("cmudict");
16       BufferedReader br = new BufferedReader(fr);
17       char [] cmubuff = new char[3511906];
18       fr.read(cmubuff,0,3511906);
19       fr.close();
20       cmudict = String.valueOf(cmubuff);
21   } catch(IOException e){}
22 }

```

Quellcode 5.2: Zuordnung des Aussprachelexikons zu einer Stringvariable (Klasse IPA.java)

Die dann folgenden Methoden *isLetter* und *isIPAletter* (siehe Anhang A: Zeilen 41-65 und 67-75) überprüfen nur, ob es Schnittmengen zwischen den Variablen *cmuchars* und *ipaChars* gibt. Die erste Methode gibt ein „wahr“ zurück, wenn eines der IPA-Zeichen der Variable *ipaChars* auch ein Buchstabe des Schreibalphabets ist und die zweite Methode setzt eine boole'sche Variable (*isIPA*) auf „wahr“, wenn einer der Buchstaben des Alphabets auch einem IPA-Zeichen entspricht. Auf diese Methoden wird zurückgegriffen, um herauszufinden, ob es sich um Sonderzeichen und Lücken handelt. Immer dann, wenn die Methode ein falsch zurückgibt, wird das betreffende Zeichen gelöscht.

```

3   public class NextWord
4   {
5       int position;
6       String word;
7
8       NextWord(String s, int i)
9       {
10          word = s;
11          position = i;
12      }
13 }

```

Quellcode 5.3: Die „Setzerklasse“ NextWord

Auf diese Überprüfung muss bei der Methode *findWord* (Quellcode: 5.4) vom Typ der Klasse *NextWord* zurückgegriffen werden. Dabei sorgt *NextWord* (Quellcode: 5.3) dafür, dass jedes Wort eines Textes eindeutig definiert wird, nämlich durch seine Position im Text.²⁰ Da das Wort und die Position untrennbar zusammen gehören werden sie in einer

²⁰*NextWord* ordnet den Variablen *word* und *position* die Werte von *s* und *i* zu, welche es beispielsweise aus *findWord* bei der Instanziierung mit *NextWord n = new NextWord(s,pos)*; als *s* beziehungsweise *pos* bekommt. Das ist auch deshalb nützlich, weil *findWord* dann beide Werte als ein *n* zurückgeben kann.

5 Modellierung einer stochastischen Sprachsegmentierung

extra Klasse auch zusammengefasst. Die Methode *findWord* sucht das nächste Wort ab *pos* in der Stringvariable *thisData* und gibt dieses Wort zusammen mit der Position vom Typ *NextWord* aus.

```
77 static NextWord findWord(int pos)
78 {
79     while((pos<thisData.length())&&!isLetter(thisData.charAt(pos)))
80     {
81         pos++;
82     }
83     String s = "";
84     while((pos<thisData.length())&&isLetter(thisData.charAt(pos)))
85     {
86         s=s.concat(String.valueOf(thisData.charAt(pos)));
87         pos++;
88     }
89     NextWord n = new NextWord(s, pos);
90     return(n);
91 }
```

Quellcode 5.4: Die statische Methode *findWord* (Klasse *IPA.java*)

Die folgenden Methoden können als das Kernstück der Klasse *IPA.java* angesehen werden. Sie konvertieren einen Text in seine IPA-Transkription.

Zunächst wird der umzuwandelnde Text der globalen Stringvariable *thisData* zugeordnet. *dataOut* wird am Ende den IPA-Text enthalten und *pos* ist eine lokale Variable, welche die Position des Wortes kodiert, also die „Setzerklasse“ *NextWord* bedient. Eine Schleife wird für die Länge des Textes durchlaufen. Die Position des ersten Zeichens (*pos = 0*) wird an die Methode *findWord* gegeben. Diese erhöht *pos* jeweils um eins solange bis eine Lücke auftritt (Quellcode: 5.4, Zeile 79), also die Methode *isLetter* ein Falsch zurückgibt. Die zweite *while*-Schleife in *findWord* verbindet alle Zeichen, die nicht durch eine Lücke unterbrochen werden und ordnet sie der Variable *s* zu. *pos* und *s* können dann als eine Instanzvariable vom Typ *NextWord* an die Methode *convertToIPA* zurückgegeben werden. Dort wird die Zeichenkette in Kleinbuchstaben umgewandelt. Schließlich erfolgt die Umwandlung des Wortes in seine IPA-Notation durch Aufruf der Methode *convertToIPA* (Quellcode: 5.6).

```
93 static String convertTextToIPA(String data)
94 {
95     thisData = data;
96     String dataOut = "";
97     int pos = 0;
98     while(pos<data.length())
99     {
100         NextWord nextWord = findWord(pos);
101         nextWord.word = nextWord.word.toLowerCase();
```

```

102     String w = convertToIPA (nextWord . word );
103     if (w!=null)
104     {
105         nextWord . word = w;
106         dataOut+=nextWord . word ;
107     }
108     pos = nextWord . position ;
109 }
110 return ( dataOut );
111 }

```

Quellcode 5.5: Umwandlung eines Textes in seine IPA-Form (Klasse IPA.java)

Diese Methode sucht genau das Wort s beziehungsweise $word$ in der globalen $cmu\text{-}dict$ -Variable, in welcher ja die Aussprache in SAMBA-Notation gespeichert ist, und gibt die Position des Wortes an pos . Wenn das Wort im CMU -Lexikon vorhanden ist, wird an der Stelle pos die SAMBA-Notation der Variable $cmu\text{-}word$ zugeordnet.²¹ Weil die Intonationen nicht beachtet werden sollen, werden an dieser Stelle die Zahlen, die die Betonung im Lexikon enkodieren, beseitigt. Die nun folgende Schleife kann die SAMBA-Notation durch die entsprechenden IPA-Zeichen einfach ersetzen, weil sie bereits als globale Variablen $cmu\text{-}Chars$ und $ipa\text{-}Chars$ und als geordnete Paare vorliegen. Die Position i entspricht also in beiden Notationen dem gleichen Laut. Die IPA-Transkription wird als $cmu\text{-}word$ zur $convertTextToIPA$ -Methode (Quellcode: 5.5) zurückgegeben oder als Wert Null, wenn das Wort nicht im Lexikon enthalten ist. Dort wird nur noch die „Setzerklasse“ (Quellcode: 5.3) modifiziert und die Ausgabevariable $dataOut$ entsprechend ergänzt.²² Das neue Wort und seine Position liegt dort schließlich in IPA-Form vor. Sollte das Wort nicht im Aussprachelexikon enthalten gewesen sein, wird natürlich die Position angepasst.

```

135 static String convertToIPA (String word)
136 {
137     int pos = cmudict . indexOf (String . valueOf ((char)0x0a) .
138         concat (word . toUpperCase () . concat (" _ _ ")));
139     if (pos >=0)
140     {
141         String cmuword = cmudict . substring (pos+3+word . length () ,

```

²¹Die hier etwas eigenartig anmutende Definition der Stelle liegt am Design des CMU -Lexikons. In der SAMBA-Notation sind auch Lücken enthalten, welche die einzelnen Laute voneinander trennen, nicht jedoch die Worteinträge voneinander. Jeder Eintrag wird mit einem Umbruchzeichen begonnen. Darauf folgt das Wort, gefolgt von zwei Lücken und der SAMBA-Notation. So ergibt sich die Zahl 3 in $\dots substring(pos+3+word.length(), \dots$ pos steht an der Stelle des Umbruchs, dann folgt das Wort, (+1), dann die zwei Lücken (+2).

²²Dadurch, dass nach $word$ kein Leerzeichen eingefügt wird, erscheint bei Ausgabe des $Strings$ auch kein Leerzeichen. Die einfache Ergänzung +“ “ wie in der Methode $convertTextToIPAWhiteSpace$ (siehe Anhang A) gibt den kompletten Text als IPA-Transkription mit Lücken aus. Dieser ist für die Leistungsmessung (Berechnung der Fehler durch Vergleich) des Algorithmus wichtig.

5 Modellierung einer stochastischen Sprachsegmentierung

```
142         cmudict.indexOf( String.valueOf((char)0x0a),
143         pos+3+word.length()));
144     cmuword = cmuword.replaceAll("\\d", "");
145     for (int i=0; i<cmuChars.length; i++)
146     {
147         cmuword = cmuword.replace(cmuChars[i], String.
148         valueOf(ipaChars[i]));
149     }
150     cmuword = cmuword.replaceAll("\\s", "");
151     return (cmuword);
152 }
153 return (null);
154 }
```

Quellcode 5.6: Umwandlung in IPA Notation (Klasse IPA.java)

In der IPA.java finden sich noch einige weitere Methoden, die aber zum Verständnis des Algorithmus nicht beitragen. Hier sei lediglich ihr Zweck kurz erwähnt. Die Methode *IPAToHTML* (Zeilen:170-187 der Klasse IPA.java im Anhang) wandelt einen Text in IPA und anschließend in seine HTML-Form um. Die Methode *convertToPlain* (Zeilen 155-167) wandelt ein IPA transkribiertes Wort in seine schriftsprachliche Normalform um. Diese Methode fand beim automatisierten Abgleich der Fehlermessung des Paralexikons Anwendung. Dabei ist jedoch Vorsicht geboten, weil die Beziehung Lautsprache-Schriftalphabet nicht eindeutig ist. Schon bei der Graphem-Phonem-Umwandlung treten Fehler auf, wenn Homographe unterschiedlich ausgesprochen werden, also heteronym sind. Hier wird das Problem allerdings dadurch abgemildert, dass die Position im *CMU-Lexikon* die meistgebrauchte Form zuerst auflistet.²³ Die umgekehrte Richtung (Phonem-Graphem) liefert dann natürlich nicht das wahrscheinlichere Wort zurück sondern einfach einen Eintrag.²⁴ Die umgekehrte Transposition führt auch deshalb zu suboptimalen Ergebnissen, weil im *CMU Pronouncing Dictionary* circa doppelt so viele Homonyme mit unterschiedlicher graphemischer Repräsentation vorhanden sind als Heteronyme.²⁵

Die Methoden *substituteDiphthongs* (Zeilen 189-202) und *resubstituteDiphthongs* (Zeilen 204-217) wandeln die Diphthonge (au, ai, əv, ɔi, ei) in Einzeichenrepräsentationen (⊙, ⊕, ⊖, ♂, ☆) um beziehungsweise umgekehrt.²⁶ Diese Relation ist umkehrbar eindeutig, sodass jeder Diphthong stets genau eine Entsprechung hat und beide Formen austauschbar sind: Für die internen Berechnungen wird ein Zeichen verwendet, für die Visualisierung werden zwei Zeichen ausgegeben. Die Funktionsweisen der Methoden sind

²³Das *CMU Pronouncing Dictionary* definiert jeden homographen Eintrag mit einer Zahl in Klammern.

So könnte man zur Umwandlung eines Textes in seine Lautform die Zeitform im Korpus am Wort ebenfalls kodieren und hätte diese Fehlerquelle beseitigt.

²⁴Für die Leistungsmessung ist dies unerheblich, weil auf jeden Fall klar ist, dass ein solcher Eintrag vorhanden ist, also das Wort existiert und als Wort gezählt werden kann.

²⁵Ein Perl-Programm zur Überprüfung der Aussage findet sich im Anhang A. Folgendes Ergebnis wird bei der Durchführung ausgegeben: Homonyme: 13004; Heteronyme: 3840

²⁶siehe Anhang A für den Quellcode und Anhang C für die Zuordnungen.

denkbar einfach. Die Zeichenketten werden durch Aufruf von *replace* und Mitgabe der entsprechenden Hexadezimalkodierungen ersetzt.

Die Methoden *readFile* und *writeFile* (siehe Anhang A: Zeilen 219-236 und 238-251 der Klasse *IPA.java*) sind die Standardmethoden für die Ein- und Ausgabe von Texten. Sie greifen auf die voreingestellten *Streams* von *BufferedReader* beziehungsweise *Writer* zurück. Ihnen wird ein Wahrheitswert *unicode* dazugegeben, welcher, wenn wahr, den Text als 16-Bit kodierte Unicodezeichen einliest beziehungsweise in eine Datei schreibt. Die andere Besonderheit (Zeile 223) bei *readFile* ist die Beschränkung der einzulesenen Corpora auf 5 MB²⁷ durch einen *Array* des Typs *char* der Größe [5*1048576]. Das entspricht 220 Zeichen, also fünf Millionen Bytes.

5.4.4 Die Übergangswahrscheinlichkeiten

Die Berechnung der Übergangswahrscheinlichkeiten wird in der Klasse *TransProbAutoCompareLexiconDelayed.java* vorgenommen. In der Stringvariable *dataIn* ist die Korpusdatei enthalten. Bei Mitgabe des *ipa*-Arguments wird zunächst eine Datei ausgegeben, die den Text des Korpus in IPA-Transkription mit Leerzeichen enthält und in *dataIn* wird die IPA-Transkription ohne Leerzeichen geschrieben. In *limitStep* werden die Schrittweiten für die Sensibilitäten gespeichert. Diese werden dann entsprechend auf eine Kommastelle gerundet und durch die geschachtelte Schleife für jede Phonemkette wiederholt. Schließlich werden im *String outFile* noch Dateiname, Periode, Phonemkettenlänge (λ) und Sensibilität (σ) hinterlegt. Diese Zahlen werden später zur eindeutigen Identifizierung für Dateinamenergänzungen benötigt.

Der nächste Schritt ist die Initialisierung einer Vektordatenstruktur *phonemeChains1*, in der, wenn es sich mindestens um die zweite Periode handelt und die rekursive Struktur noch nicht begonnen hat (*if*-Bedingung in Zeile 59, Quellcode: 5.7), alle Wörter des dann bereits bestehenden Paralexikons²⁸ eingelesen werden. Auch hier werden die Diphtonge entsprechend angepasst. Weil im Paralexikon auch die Vorkommenshäufigkeiten der *Token* enthalten sind, diese aber nicht beim Einfügen in dem neuen Text zu fehlerhaften Einsetzungen führen sollen, werden in Zeile 68 (Quellcode: 5.7) alle Zeichen gelöscht, die nicht durch die Methode *ipaChars* in der Klasse *IPA.java* definiert sind. Die sich anschließenden *for*-Schleifen beinhalten ein Standardsortierverfahren (*bubblesort*),²⁹ welche alle Einträge im *Stringarray c* ordnet, sodass die längsten Einträge am Anfang stehen. Diese werden mit *addElement* in den Vektor gepackt. Für die erste Periode besteht noch kein Paralexikon; so bleibt der Vektor leer. Der so sortierte *Array* wird später helfen, eine einfache Implementierung des Kohortenmodells zu integrieren.

49 **double** limitStep = 1.0 / raster;

²⁷Zum Zeitpunkt der Programmierung war noch nicht klar, wie groß die Corpora maximal werden würden, sodass hier zunächst großzügig Speicher belegt wurde.

²⁸Wurde aus der Klasse *LexiconDelayed.java* als *arg[0]* mit übergeben.

²⁹Das Verfahren ist dadurch definiert, dass *i* und *j* jeweils alle möglichen Paarkombinationen durchgehen, wobei sie aber stets paarweise verschieden sind.

5 Modellierung einer stochastischen Sprachsegmentierung

```

50 for (int len=1;len<=maxlength;len++)
51 {
52     System.out.println("len_"+len);
53     for (double limit=1;limit>=0;limit=java.lang.Math.round(1000*
54         (limit-limitStep))/1000.0)
55     {
56         System.out.println("limit_"+limit);
57         String outFile = args[0]+".run"+"_"+len+"_"+limit;
58         Vector phonemeChains1 = new Vector();
59         if (run>1 && (run>delay))
60         {
61             String chainsfile1 = args[0]+".(run-1)+"_"+len+"_"+
62                 limit+".lexicon";
63             String data = IPA.resubstituteDiphthongs(IPA.readFile
64                 (chainsfile1,ipa));
65             System.out.println("====>"+chainsfile1+"_"+data.length());
66             String c[] = data.split("[^a-zA-Z"+
67                 String.valueOf(IPA.ipaChars)+" ]+");
68             for (int i=0;i<c.length;i++)
69             {
70                 for (int j=i+1;(i<c.length-1)&&(j<c.length);j++)
71                 {
72                     if (c[i].length()<c[j].length())
73                     {
74                         String tmp = c[i];
75                         c[i]=c[j];
76                         c[j]=tmp;
77                     }
78                 }
79                 phonemeChains1.addElement(c[i]);
80                 //System.out.println("1 "+i+c[i]);
81             }
82         }

```

Quellcode 5.7: Stringsortierung für beliebige ω und μ (Klasse TransProbAutoCompareLexiconDelayed.java)

Nun sind alle Daten vollständig,³⁰ um die wichtigste Methode *berechneTransProb* (Quellcode: 5.9) aufzurufen. Der Stringvariable *dataOut* wird das Ergebnis dieser Methode zugeordnet, d.h. der mit Übergangswahrscheinlichkeiten segmentierte Text wird übergeben (Zeile: 85 der Klasse TransProbAutoCompareLexiconDelayed.java im An-

³⁰Das bedeutet: Die Werte der ersten beiden *for*-Schleifen der Klasse Phonemkettenlänge (*len*) und die dazu gehörigen Sensibilitäten (*limit*), bestimmt durch *raster* (zweite *for*-Schleife), sind berechnet. Der Text ohne Lücken (*dataIn*) und das Paralexikon (*phonemeChains1*).

hang). Dieser wird gleich darauf in zwei Dateien ausgegeben,³¹ einmal lediglich mit den Daten von *outFile* und beim zweiten Mal mit der zusätzlichen Endung „substitute“.³² Schließlich wird ein zweiter Vektor angelegt, welcher das Ergebnis der Methode *Compare* enthalten wird. Diese Methode nimmt den Text des Korpus (*arg[0]*), die Referenz der segmentierten Datei (*outFile*) sowie die boolesche Variable (*ipa*) und gibt wiederum einen Vektor (*results*) zurück. Letzterer enthält einen *Array* und einen *Integer*: Der *Integer hit* enthält die richtig gesetzten Lücken, *fmiss[1]* enkodiert die falsch gesetzten Leerzeichen und *fmiss[0]* enthält die Gesamtzahl der Leerzeichen. Damit sind alle Zahlen vorhanden, die zur Berechnung der Präzision, der Vollständigkeit und damit auch F_1 später benötigt werden. Die *try*-Anweisungen (Zeilen: 97ff) legen nur die Struktur der *result*-Dateien an, so wie sie beispielhaft in Tabelle 5.3 als Text ausgegeben und von dort manuell in eine Excel-Datei beziehungsweise als Matlab-Input transferiert werden. Zuletzt erfolgt dort Berechnung und grafische Aufbereitung der Leistungsmaße. Der Vektor *compareResults* enthält nämlich die Rückgabe der Methode *Compare* und dies ist der Vektor *results*.

```

222 while ((( fpos [0] = nextLetter(0)) >=0) &&
223         ( fpos [1] = nextLetter(1)) >=0)
224 {
225     //System.out.println(""+fpos[0]+" "+fpos[1]+" "+
226     //f[0].charAt(fpos[0])+" "+f[1].charAt(fpos[1]));
227     if (( fpos[0]+1 < f[0].length()) && (fpos[1]+1 < f[1].length()))
228     {
229         if (isWhiteSpace(f[0].charAt(fpos[0]+1)))
230         {
231             if (isWhiteSpace(f[1].charAt(fpos[1]+1)))
232             {
233                 hit++;
234             }
235             else
236             {
237                 fmiss[0]++;
238             }
239         }
240         else
241         {
242             if (isWhiteSpace(f[1].charAt(fpos[1]+1)))
243             {
244                 fmiss[1]++;
245             }
246         }

```

³¹Die dazu notwendigen Dateieindungsschlüssel wurden zuvor in *outFile* geschrieben.

³²Diese wird dann nach Durchlauf dieser Klasse in der *LexiconDelayed.java* in „lexicon“ umgewandelt. In *WordCount.java* wird darauf schließlich die Lexikonfunktion angewendet.

5 Modellierung einer stochastischen Sprachsegmentierung

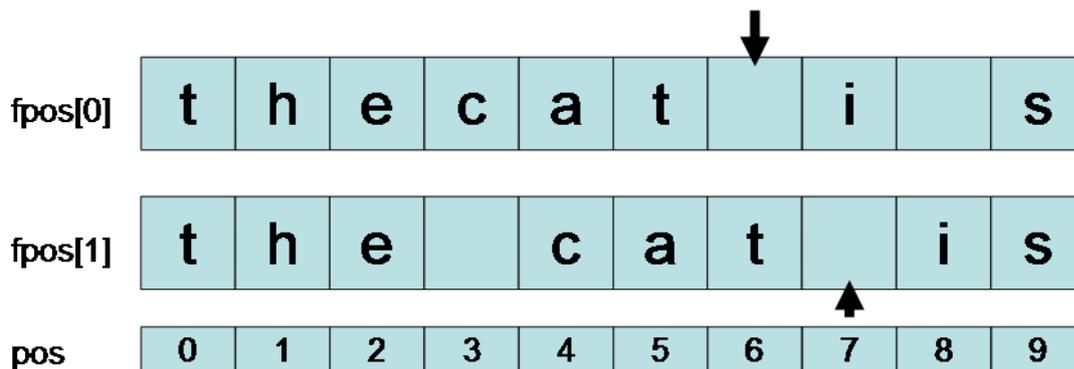


Abbildung 5.4: Problematik des Stringabgleichs

247
248

```
}
}
```

Quellcode 5.8: Datenstrukturen für Berechnung der Leistungsmaße (Klasse TransProbAutoCompareLexiconDelayed.java)

Mithin bleiben noch fünf Methoden einige klärende Gedanken schuldig. Zuerst sollen die drei einfachen Methoden erklärt werden. *removeWhiteSpace* entfernt die Lücken eines Textes. Sie wird eigentlich nur für Texte, die nicht in IPA transkribiert werden, zur Leerzeichenentfernung genutzt. Zuerst wird eine Charakterklasse mit regulären Ausdrücken definiert, die dann negatiert und mit der *replaceAll*-Methode aus der „java.lang.String“-Bibliothek als Argument übergeben wird. Das zweite Argument ist ein leerer *String*, der für alles, was nicht zur definierten Charakterklasse gehört, eingesetzt wird.

Die boolesche Methode *isWhiteSpace* (Zeilen 254ff der Klasse TransProbAutoCompareLexiconDelayed.java im Anhang) überprüft, ob ein Zeichen ein Leerzeichen ist. Dabei wird jedes Zeichen, welches nicht ein IPA-Zeichen oder ein Buchstabe des Schreibalphabets ist, als Leerzeichen definiert.³³

nextLetter (Zeilen 258ff der Klasse TransProbAutoCompareLexiconDelayed.java im Anhang) sucht ab einer bestimmten Stelle *i* das nächste Zeichen. Die Methode gibt die Position im Array zurück oder eine -1, welche als boolesche Variable benutzt wird. Damit fungiert die Methode zwar als Integer, kann aber auch als boolesche Methode gebraucht werden, wenn die *if*-Anweisung falsch wird.

Im Folgenden werden die zwei umfangreichen Methoden beschrieben. Beide haben bei der Klassenbeschreibung bereits Erwähnung gefunden. Die Methode *Compare* vom komplexen Datentyp *Vector* dient dem Vergleich zwischen dem Inputtext und dem segmentierten Quelltext. Die Herausforderung besteht hier darin, zwei *Arrays* zu vergleichen, deren Positionszeiger nicht synchron auf einer Position laufen, aber trotzdem verglichen

³³Dadurch werden auch Satzzeichen, Sonderzeichen oder Apostrophe als Leerzeichen angesehen. Dabei wird beispielsweise das *'s* wie bei *that's* dem Wort zugehörig definiert(*thats*), nicht jedoch als zwei Wörter wie *that is*.

5.4 Beschreibung des Algorithmus

Inhalt einer .result-Datei	Repräsentiert mit Variablen des Programms
laenge = 2	laenge = $\text{maxLength} - i_{\text{maxLength}-2}$, also: $(5-(5-2))=2$
1.0 177 109 2425	j0/raster a210 b210 c210
0.9 186 122 2416	j1/raster a29 b29 c29
0.8 255 140 2347	j2/raster a28 b28 c28
0.7 293 178 2309	j3/raster a27 b27 c27
0.6 326 297 2276	j4/raster a26 b26 c26
0.5 615 216 1987	j5/raster a25 b25 c25
0.4 675 349 1927	j6/raster a24 b24 c24
0.3 585 431 2017	j7/raster a23 b23 c23
0.2 1036 502 1566	j8/raster a22 b22 c22
0.1 1168 430 1434	j9/raster a21 b21 c21
0.0 0 0 2602	j10/raster a20 b20 c20
laenge = 3	laenge = $\text{maxLength} - i_{\text{maxLength}-2}$, also: $(5-(4-2))=3$
1.0 585 431 2017	j0/raster a310 b310 c310
0.9 585 434 2017	j1/raster a39 b39 c39
0.8 628 450 1974	j2/raster a38 b38 c38
0.7 697 453 1905	j3/raster a37 b37 c37
0.6 744 476 1858	j4/raster a36 b36 c36
0.5 906 465 1696	j5/raster a35 b35 c35
0.4 931 436 1671	j6/raster a34 b34 c34
0.3 973 373 1629	j7/raster a33 b33 c33
0.2 996 284 1606	j8/raster a32 b32 c32
0.1 779 98 1823	j9/raster a31 b31 c31
0.0 0 0 2602	j10/raster a30 b30 c30
laenge = 4	laenge = $\text{maxLength} - i_{\text{maxLength}-2}$, also: $(5-(3-2))=4$
1.0 675 349 1927	j0/raster a410 b410 c410
0.9 675 349 1927	j1/raster a49 b49 c49
0.8 682 352 1920	j2/raster a48 b48 c48
0.7 3727 345 1875	j3/raster a47 b47 c47
0.6 778 341 1824	j4/raster a46 b46 c46
0.5 775 239 1827	j5/raster a45 b45 c45
0.4 762 240 1840	j6/raster a44 b44 c44
0.3 674 200 1928	j7/raster a43 b43 c43
0.2 543 122 2059	j8/raster a42 b42 c42
0.1 351 37 2251	j9/raster a41 b41 c41
0.0 0 0 2602	j10/raster a40 b40 c40
laenge = 5	laenge = $\text{maxLength} - i_{\text{maxLength}-2}$, also: $(5-(2-2))=5$
1.0 615 216 1987	j0/raster a510 b510 c510
0.9 615 216 1987	j1/raster a59 b59 c59
0.8 621 213 1981	j2/raster a58 b58 c58
0.7 653 211 1949	j3/raster a57 b57 c57
0.6 670 217 1932	j4/raster a56 b56 c56
0.5 620 121 1982	j5/raster a55 b55 c55
0.4 603 123 1999	j6/raster a54 b54 c54
0.3 461 108 2141	j7/raster a53 b53 c53
0.2 298 53 2304	j8/raster a52 b52 c52
0.1 200 8 2402	j9/raster a51 b51 c51
0.0 0 0 2602	j10/raster a50 b50 c50

Tabelle 5.3: Form der Ausgabedatei ($\text{maxLength} = 5$, $\text{raster} = 10$)

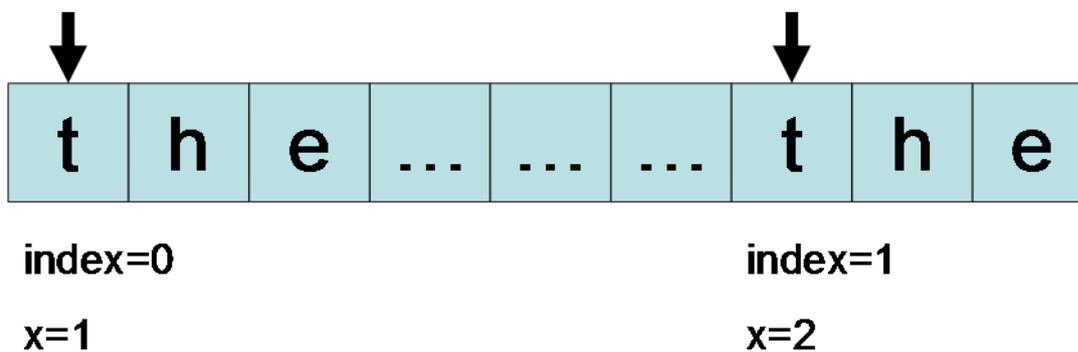


Abbildung 5.5: Abzählen von Phonemketten

werden müssen. Dies kommt deshalb zu Stande, weil einige Lücken an der falschen Stelle, andere gar nicht gesetzt werden (siehe Abb. 5.4). Zuerst werden drei globale *Arrays* deklariert:³⁴ Der *Stringarray* *f* enthält die beiden zu vergleichenden Texte, *fpos* gibt die Positionen im *Array* an, in dem gerade gesucht wird und in *fmiss* werden die falschen Leerzeichen und deren Gesamtzahl geschrieben.

Das Problem wird mit einer *while*-Schleife, einigen *if*-Anweisungen und den oben beschriebenen Methoden *isWhiteSpace* und *nextLetter* gelöst (Quellcode: 5.8). Die *while*-Schleife prüft, ob Position und Zeichen in *fpos* übereinstimmen und geht für den wahren Fall schrittweise den *Array* durch. Da *pos* in *nextLetter* für die Länge des *Arrays* immer kleiner ist als *f[i]*, liefert die Methode *nextLetter* erst dann die -1 zurück und tritt damit aus der *while*-Schleife aus, wenn die *Arrays* vollständig durchlaufen sind. Innerhalb der Schleife wird nun mit drei ineinander geschachtelten *if*-Anweisungen und unter Rückgriff auf *isWhiteSpace* (siehe oben) geprüft, ob $f[0] == f[1]$, $f[0] == \text{Leerzeichen}$ und $f[1] == \text{Leerzeichen}$. Wenn diese drei Bedingungen wahr sind, dann wird eine Zählvariable (*hit*) erhöht. Das sind die richtigen Treffer. Wenn letztere Bedingung nicht erfüllt ist ($f[1] \text{ Leerzeichen}$), dann wird $fmiss[0]$ um eins addiert, zwar erfolgte keine Lücke an der falschen Stelle, aber ein Leerzeichen wurde nicht erkannt. Ist die zweite Bedingung nicht erfüllt ($f[0] == \text{Leerzeichen}$), aber die dritte, dann ist klar, dass ein falsches Leerzeichen eingefügt wurde. $fmiss[1]$ muss also erhöht werden. Letztlich werden die drei *Arrays* dem *results*-Vektor zugeordnet.

Die wichtigste Methode dieser Klasse und eine der zentralen Kernpunkte der Arbeit ist die Methode *berechneTransProb*. Sie berechnet die Übergangswahrscheinlichkeiten für jede spezifizierte Phonemkettenlänge und Sensibilität. *berechneTransProb*, wenn so eingestellt, fügt auch das Paralexikon in den unsegmentierten Text ein. Die Methode erhält, wie in Zeile 85 (der Klasse *TransProbAutoCompareLexiconDelayed.java* im Anhang) spezifiziert, den lückenlosen Text, die Phonemkettenlänge, die Sensibilität und das Paralexikon.³⁵

³⁴Global deshalb, weil die Inhalte später in der Methode *nextLetter* benötigt werden.

³⁵Der Vektor, welcher das Paralexikon beinhaltet, kann natürlich auch Null sein wie im ersten Durchlauf.

```

126 static String berechneTransProb(String data,int len ,
127                               double limit , Vector phonemeChains1)
128 {
129     boolean marks [] = new boolean[data.length ()];
130     for (int i=0;i<=data.length()-2*len ;i++)
131     {
132         String sDouble = data.substring (i , i+2*len );
133         int index = 0;
134         int x = 0;
135         while ((index = data.indexOf(sDouble , index))>=0)
136         {
137             x++;
138             index++;
139         }
140         String sSingle = data.substring (i , i+len );
141         index = 0;
142         int y = 0;
143         while ((index = data.indexOf(sSingle , index))>=0)
144         {
145             y++;
146             index++;
147         }
148         if (x/(double)y<limit)
149         {
150             marks [ i+len -1]=true ;
151         }
152     }

```

Quellcode 5.9: Berechnung der Übergangswahrscheinlichkeiten (Klasse TransProbAuto-CompareLexiconDelayed.java)

Nach der Initialisierung eines boole'schen *Arrays*, welcher für jedes Zeichen des unsegmentierten Textes Speicherplatz reserviert, wird für jede mögliche Phonemkettenlänge eine *for*-Schleife gesetzt. Der Ausdruck *data.length()-2*len* (Quellcode: 5.9, Zeile 130) gibt genau die Zahl, die man für jede beliebige Länge *len* von der Gesamtzeichenzahl des Textes abziehen muss, um $P(A)$ oder $\#A$ zu erhalten.³⁶ Hier wird also festgelegt, über wie viele Phoneme gemittelt werden soll. *data.substring(i, i+2*len)* gibt an, wo sich die Phonemkette im *String* befindet. Mit jedem Durchlauf der Schleife wird *i* erhöht und so im *Array* ein Zeichen weiter gegangen. Die nun folgende *while*-Schleife gibt solange ein wahr, bis die Suche nach dem so definierten *substring* im unsegmentierten Text keine Treffer mehr findet. Für jeden Treffer wird *x* erhöht. Wichtig ist aber auch der *index*-Integer, der die Position im *Array* markiert und der nach *x* erhöht wird und ab dem wieder gesucht werden muss (Abbildung 5.5). Ansonsten würden Doppelzählungen

³⁶Siehe Kapitel 5.2.3, Formel 5.14 beziehungsweise 5.15 auf Seite 119.

5 Modellierung einer stochastischen Sprachsegmentierung

entstehen, weil nach jeder erfolgreichen Suche, der letzte *substring* wieder mitgezählt würde. Das „B“ in den Formeln (5.14) und (5.15) in Kapitel 5.2.3 auf Seite 119 entspricht natürlich der Länge *len* (also λ), welche mittels dem *i* der ersten *for*-Schleife eine eindeutige Position zugewiesen bekommt. Auch hier wird der gleiche Zählmechanismus verwendet, wie bei der Berechnung von „A“.

Nun wird der Quotient aus Gleichung (5.14), die Übergangswahrscheinlichkeit von einer Lautkette zu der Folgelautkette, aus *A* und *B* berechnet (Quellcode: 5.9, Zeile 48). Gleichzeitig wird geprüft, ob das Ergebnis kleiner als die Sensibilität ist. Ist dies nämlich der Fall, wird der eingangs definierte boole'sche *Array* genau an der Stelle (Mitte der Phonemkettenlänge von „A“) auf „wahr“ gesetzt (Zeile 150, Quellcode: 5.9). Damit kann die anfängliche *for*-Schleife geschlossen werden. An den Stellen des *Arrays*, wo ein wahr markiert wurde, sollen später die Leerzeichen eingefügt werden. Davor müssen aber noch alle Markierungen, welche aufeinander folgen, gelöscht werden. Diese Vorgehensweise hat folgenden Grund: Der Segmentierungsansatz der Übergangswahrscheinlichkeiten setzt voraus, dass sich die Verteilungen innerhalb eines Wortes von denen zwischen den Wörtern unterscheiden. Deshalb muss die *if*-Anweisung in Zeile 148 (Quellcode 5.9) ab einer bestimmten Sensibilität (*limit*) für alle Phonemübergänge innerhalb eines Wortes wahr sein. Genau am Wortende beziehungsweise -anfang ändert sich aber die Übergangswahrscheinlichkeit. Hier wird ein falscher Wert von einem wahren Wert im boole'schen *Array marks[]* gefolgt. Nur an dieser Stelle muss aber ein Leerzeichen eingefügt werden. Darum werden die anderen Wahrheitswerte wieder auf *Default* gesetzt.

```
168 String dataOut=" ";
169
170 for(int i=0;i<data.length();i++)
171 {
172     System.out.println(i);
173     int chains1 = -1;
174     for(int j=0;(j<phonemeChains1.size()) && (chains1<0);j++)
175     {
176         String chain = (String)phonemeChains1.elementAt(j);
177         if(data.regionMatches(i, chain, 0, chain.length()))
178         {
179             chains1=j;
180         }
181     }
182     if(chains1 >=0)
183     {
184         String chain = (String)phonemeChains1.elementAt(chains1);
185         System.out.println("1_" + chain);
186         if((dataOut.length()>0) && (dataOut.charAt(dataOut.length()
187             -1)!=32)) dataOut=dataOut.concat(" ");
188         dataOut=dataOut.concat(chain+" ");
189         i+=chain.length()-1;
```

```

190     }
191     else
192     {
193         dataOut=dataOut.concat(data.substring(i,i+1));
194         if(marks[i])
195         {
196             dataOut=dataOut.concat(" ");
197         }
198     }
199 }

```

Quellcode 5.10: Einfügen der Leerzeichen am Wortende und Kohortenmodell (Klasse TransProbAutoCompareLexiconDelayed.java)

Vor dem Einfügen des Leerzeichens wird aber erst noch geprüft, ob im Vektor *phoneChains1* (Paralexikon) Wörter vorhanden sind.³⁷ Diese werden, wenn wahr, eingefügt, wenn nicht wird zur *else*-Anweisung (Zeile 191, Quellcode: 5.10) gesprungen und mit *dataOut=dataOut.concat(data.substring(i,i+1))*³⁸ immer das *(i+1)*te Zeichen an *dataOut* angehängt. Wenn es aber markiert, also wahr, ist (*if(marks[i])*) wird ein Leerzeichen angehängt (*dataOut=dataOut.concat(" ")*). Zuletzt wird *dataOut* an *LexiconDelayed.java* zurückgegeben.

5.4.5 Häufigkeitsfunktion und Paralexikon

Die Häufigkeitsfunktion wird direkt in *LexiconDelayed.java* in der statischen Methode *grenzfunktion* als mathematischer Ausdruck (Zeile 59 der Klasse *LexiconDelayed.java* im Anhang) hineingeschrieben. Diese Funktion bestimmt, wie viele Einträge der Sprachlerner zur Verfügung hat. Die Häufigkeitsfunktion ζ beeinflusst auf diese Weise das Paralexikon. An dieser Stelle macht es Sinn, beide Simulationskomponenten, Häufigkeitsfunktion und Paralexikon, zusammen zu beschreiben, weil die Häufigkeitsfunktion simulativ durch den Vergleich mehrerer Durchläufe ermittelt wird und deshalb eher ihre Verarbeitung innerhalb der Wortliste des Paralexikons algorithmisch interessiert.

Zuerst sei aufgezeigt, wie die Wortliste mit einem bestehenden Text eingepasst werden kann. Im Kapitel 3.1 ist deutlich geworden, dass das Kohortenmodell (Marslen-Wilson und Welsh, 1978) eine einfache Implementierung verspricht und experimentell auch gut begründet ist. Der Abgleich mit Kohorten kann sehr leicht durch bereits vorhandene Datenstrukturen erzwungen werden. Im hiesigen Fall sind die möglichen Kohorten im Paralexikon enthalten, die mit dem Textinput abgeglichen werden müssen. Wie zu Beginn der Beschreibung der Klasse *TransProbAutoCompareLexiconDelayed.java* der Zeilen 59-82 (Quellcode 5.7) ausgeführt wurde, enthält die Vektordatenstruktur *phonemeChains1* das nach Länge geordnete Paralexikon in einer bestimmten Periode (*run*). Der

³⁷Dieser Programmteil soll aber aus methodischen Gründen im Unterkapitel Paralexikon erfolgen.

³⁸Das *i* ist Teil einer übergeordneten *for*-Schleife, welcher über die Länge von *data* läuft (siehe Quellcode 5.10).

Wettbewerb wird hier auf die Länge des Eintrags beschränkt.³⁹ Die programmiertechnische Umsetzung des Modells wird dadurch erreicht, dass die Kohorten in umgekehrter Reihenfolge abgearbeitet werden, d.h., das längste Wort des Paralexikons wird zuerst geprüft.⁴⁰ Wenn es bereits passt, hat es den Wettbewerb gewonnen und es kann schnell zum nächsten Eintrag übergegangen werden. Dies erfolgt in Zeile 174 (Quellcode 5.10). Dort wird geprüft, ob die Einträge des Vektors dem gerade aktuellen Wort entsprechen. Ist im Inputtext das Wort aus *phonemeChains1* vorhanden, werden das Wort und die entsprechende Lücke danach eingefügt. Ist das Wort nicht vorhanden (*else*-Anweisung ab Zeile 191 in Quellcode 5.10), wird das $(i+1)$ te Zeichen an *dataOut* angehängt, wenn es nicht markiert wurde.⁴¹ Im Falle einer Markierung, welche ja bedeutet, dass die Berechnung der Übergangswahrscheinlichkeiten hier zu einer Lückeneinfügung geführt hat, wird eine Lücke natürlich gesetzt.

Die Klasse *WordCount2.java* zählt die Wörter des Paralexikons, die pro Periode neu hinzukommen sowie die Gesamtzahl der Wörter des Lexikons pro Periode. Hier wird also gemessen wie groß das Paralexikon ist und wie schnell es wächst. Dies ist für die Kontrolle, simulative Beobachtungen und Grenzwertbetrachtungen, wie sie im Kapitel 6 ausführlich dargelegt werden, wichtig. Neben der Messung werden diese Angaben auch dafür benötigt, die häufigsten *Token* in das Paralexikon zu schreiben. Für Letzteres wird ein *Hashtable* als vorherrschende Datenstruktur verwendet. Er erlaubt die Speicherung von verschiedenen Variablentypen. Zu jedem Wort (*String*) wird dessen Anzahl (*integer*) zugeordnet. Die Strings werden nach diesem Integer geordnet und nur die x -häufigsten werden in die neue *lexicon*-Datei geschrieben.

Nach einer großzügigen Festsetzung einer maximalen Grenze von 1.000.000 Einträgen⁴² für das Paralexikon und der Initiierung der *Hashtable* wird zu Beginn der Klasse das Vokabular der Vorgängerperiode, falls vorhanden, aus genau jener Lexikodatei eingelesen, die den primären Parametereinstellungen (Phonemkettenlänge und Sensibilität) entspricht, und in eine neue Lexikodatei geschrieben. Die Lexikodatei wird entsprechend der aktuellen Periode (Phonemkettenlänge und Sensibilität) benannt. Das neue Vokabular der letzten Periode befindet sich in einer Datei gleichen Namens aber mit der Endung *substitute* oder im Programm in der Stringvariable *f*. Dies ist also der mit Übergangswahrscheinlichkeiten segmentierte Text, der möglicherweise in den Vorgängerperioden bereits mit Lexikoneinfügungen manipuliert wurde. Die *while*-Schleifen (ab Zeile: 45 der Klasse *WordCount2.java* im Anhang) suchen in bereits bekannter Weise darin Wörter und zählen diese. Mit *put* (Zeile 63 der Klasse *WordCount2.java* im Anhang) werden Wort und Zahl in einen der anfangs angelegten *Hashtable* geschrieben. Sollte das Wort bereits im *Hashtable* vorhanden sein, wird nur dessen Zahl um eins

³⁹Die Ausweitung des Wettbewerbs auf die Häufigkeit kann aber problemlos erweitert werden. Da diese Zahlen ebenfalls in *WordCount2.java* gezählt werden.

⁴⁰Dies entspricht zwar nicht der Beschreibung von Marslen-Wilson und Welsh (1978), aber das Ergebnis wäre gleich.

⁴¹Mit *marks[i+len-1]=true*; in Zeile 150, Quellcode: 5.9.

⁴²Dies ist der Maximalwert, für den Speicherplatz reserviert wird. Man kann wohl davon ausgehen, dass sich ein Kind auch unbewusst nicht mehr als eine Million Einträge merken wird.

erhöht. Der dichotome Inhalt des *Hashtables* wird mit Hilfe eines Vektors *v* wieder im *Bubblesort*-Stil sortiert (Zeilen 72-85 der Klasse `WordCount2.java` im Anhang), sodass die häufigsten Einträge am Anfang des Vektors stehen. Mit der sich anschließenden *for*-Schleife kann dann immer das letzte Element des Vektors entfernt werden bis nur noch die *x*-häufigsten Werte übrig sind.

In dem zweiten *Hashtable* (*oldWords*) wurde bereits am Anfang der Klasse `WordCount2.java` über die Methode `readVocabulary` (Zeilen 151-167 im Anhang) der Inhalt der Vorgängerlexikondatei geschrieben (Einträge und deren Anzahl). Diese Daten werden nun benötigt um die Zahl neuer Wörter zu bestimmen. Sie werden in der Variable *countNew* hinterlegt. Jedes Mal *if* (`oldWords.containsKey((String)v.elementAt(i))`) wird nur der Wert des gefundenen Wortes um eins erhöht. Wenn aber das Wort nicht gefunden werden kann, erhöht sich *countNew* entsprechend, weil es sich ja um einen neuen Eintrag handeln muss. Ebenso werden dann alle Wörter und deren Häufigkeiten der Vorgängerperiode in den aktuellen *Hashtable* übernommen (Zeile 106 der Klasse `WordCount.java` im Anhang). Die dann folgende *try*-Anweisung gibt das Ergebnis in der Datei *wordcount* aus. Letztlich wird die neue Liste noch einmal sortiert (Zeilen 121ff)⁴³ und an *dataOut* übergeben (Zeilen 136-141 der Klasse `WordCount.java` im Anhang).⁴⁴

5.4.6 Programmanfang

Die Klasse `LexiconDelayed.java` enthält die zu startende *main*-Methode, der die Argumente in der Eingabeaufforderung übergeben werden und welche die weiteren Klassen zur Berechnung von Übergangswahrscheinlichkeiten oder Häufigkeitsfunktionen aufruft. Die *switch*-Anweisung weist jedem Argument der Eingabe die entsprechende Variable im Programm zu.

```

16 for (int i = 1; i < args.length; i++)
17 {
18     switch(i)
19     {
20         case 1:
21             maxlength = (Integer.parseInt(args[i]));
22             break;
23         case 2:
24             raster = (Integer.parseInt(args[i]));
25             break;
26         case 3:
27             runs = (Integer.parseInt(args[i]));
28             break;
29         case 4:

```

⁴³Die Sortierung erfolgt durch die Übergabe aller „Schlüssel“ (`keySet()`) des *Hashtables* an den Vektor *v*. Ein „Schlüssel“ ist eine Nummer, die beispielsweise den Eintrag „you 20“ (Wort/Häufigkeit) enthält. Dieser Eintrag steht an dritter Stelle des *Hashtable*, weil wie bei *Arrays* bei Null begonnen wird.

⁴⁴In Zeile 143 wird abermals die Diphthongtranskription normalisiert.

5 Modellierung einer stochastischen Sprachsegmentierung

```
30         delay = (Integer.parseInt(args[i]));
31         break;
32     }
33 }
```

Quellcode 5.11: Einlesen der Variablen (Klasse LexiconDelayed.java)

Die dann folgende *for*-Schleife ist der Kern der Klasse. Sie wird für jede Periode (*run*) durchlaufen. Zuerst wird die erste Korpusdatei an die *copy*-Methode (Zeilen 64-74 der Klasse LexiconDelayed.java im Anhang) weitergegeben, deren Dateiname dort leicht geändert wird.⁴⁵ Dieser Dateiname (*file*), die Länge der Phonemkette (*maxlength*), die Sensibilität⁴⁶ (*raster*), die Anzahl der Perioden (*run*) und die Periodenverschiebung (*delay*) werden in ein *Stringarray* geschrieben, und dieses *Array* wird als Input an die *main*-Methode der Klasse TransProbAutoCompareLexiconDelayed.java gegeben. Danach werden alle Korpusdateien aus dem aktuellen Verzeichnis in ein weiteres *Stringarray subfile* geschrieben. Für jeden Eintrag in *subfile f* wird dann geprüft (Zeile 45, Quellcode: 5.12), ob dort ein Eintrag mit der Endung „.substitute“ vorhanden ist. Wenn dies der Fall ist, wird die Dateiendung durch „.lexicon“ ersetzt. Auch hier wird wieder ein *Stringarray* initialisiert, der den String *f*, die umbenannte Datei (*lexfile*), die Lexikonfunktion sowie einen Platzhalter „ipa“, der nur die Transkriptionsart angibt, zuordnet. Mit diesem String wird letztlich die *main*-Methode der Klasse WordCount2.java aufgerufen. Die Lexikonfunktion ζ wird in der Methode *grenzfunktion* definiert. Hier können alle funktionalen Verläufe direkt hineingeschrieben werden.

```
35 for (int run=1;run<=runs;run++)
36 {
37     copy(new File(file+"."+run),new File(file));
38     String ar[] = {file, String.valueOf(maxlength),
39         String.valueOf(raster), String.valueOf(run),
40         String.valueOf(delay), "ipa"};
41     TransProbAutoCompareLexiconDelayed.main(ar);
42     File dir = new File(".");
43     String[] subfile = dir.list();
44     System.out.println(file+"\\."+run+"_.*\\.substitute");
45     for (String f:subfile)
46     {
47         if (f.matches(file+"\\."+run+".*\\.substitute"))
48         {
49             String lexfile=f.replace(".substitute",".lexicon");
50             System.out.println(lexfile);
51             String ar2[] = {f, lexfile,
```

⁴⁵Die Endungen der Inputdateien enthalten jeweils eine Zahl als Endung. Diese Zahl wird gelöscht, damit die TransProbAutoCompareDelayed.java eine Datei mit gleichem Namen bekommen kann.

⁴⁶Hier wird das Raster oder die Feingliederung der Sensibilität weitergegeben, aus dem sich später die Sensibilitäten mit $1/raster$ errechnen lassen.

```

52         String.valueOf(grenzfunktion(run)), "ipa"};
53     WordCount2.main(ar2);
54     }
55 }
56 }

```

Quellcode 5.12: Erstellen der Dateiarchitektur (Klasse LexiconDelayed.java)

In Abbildung 5.6 wird die Funktionsweise des Algorithmus hinsichtlich des schematischen Programmablaufs übersichtlich dargeboten, sodass der rekursive Charakter deutlich wird. Die *delay*-Funktion μ sei hier aus Gründen der Übersichtlichkeit ausgeklammert. Linker Hand sind die primären und sekundären Variablen benannt, welche als Argumente an die einzelnen Klassen übergeben werden. Folgt man den Pfeilen, ausgehend vom ersten Aufruf der `TransProbAutoCompareLexiconDelayed.java`, gelangt man zum ersten Output, welcher dann gleich als Input der Klasse `WordCount2.java` dient. Ab diesem Punkt beginnt die Rekursion, was durch die dick markierten Pfeile illustriert werden soll.

5.4.7 Nachbetrachtung

Alle dem Autor bekannten Simulationsmodelle zur Lautstrom- oder Wortsegmentierung mittels der Übergangswahrscheinlichkeiten umfassen lediglich die ersten beiden Phasen des oben beschriebenen Algorithmus.⁴⁷ Nach Vollendung der zweiten Phase wird festgestellt, dass der hier erbrachte Output nicht ausreichend sein kann, um ein Sprachsignal zu dekodieren. Entweder wird daraus der Schluss gezogen, dass noch weitere Informationen im Sprachsignal genutzt werden müssten, deren Gesetzmäßigkeiten noch nicht erkannt wurden oder es wird das Zusammenspiel anderer sprachtypischer Anhaltspunkte (z.B. Prosodien, Allophone, Phonotaktik) betont, ohne aber sagen zu können, wie diese zusätzlichen Informationen genau interagieren. In der generativen Auffassung wird das Scheitern der rein induktiven Methode dagegen als weiterer Beleg für ein angeborenes Sprachmodul gewertet. Offensichtlich hatte man bislang nicht erkannt, dass eine mögliche Gesetzmäßigkeit generell auf Häufigkeiten beruht, zumal auch dies, vereinfacht ausgedrückt, die neurobiologische Erklärung des Lernens ist. Diese Erkenntnis trifft natürlich dann auch auf die Übergangswahrscheinlichkeiten zu. Erst damit hat man einen widerspruchsfreien Mechanismus entdeckt, der es ermöglicht, in einer rekursiven Vorgehensweise, genau das Datenmaterial zu sammeln, welches von allen anderen Modellen auf eigentümlichste Weise unterstellt wird.

Der hier gezeigte Algorithmus bietet eine alternative Sichtweise zu allen drei Argumenten. Er zeigt, dass es sehr wohl möglich ist; erstens, mit den bekannten Informationen, die induktiv ermittelt werden; zweitens, mit den generellen Fähigkeiten eines Kleinkinds und drittens, ohne Bezug auf ein Sprachmodul, einen Lautstrom zufrieden stellend zu segmentieren.

⁴⁷In Abb. 5.3 auf Seite 132 sind dies die Phasen (1) und (2).

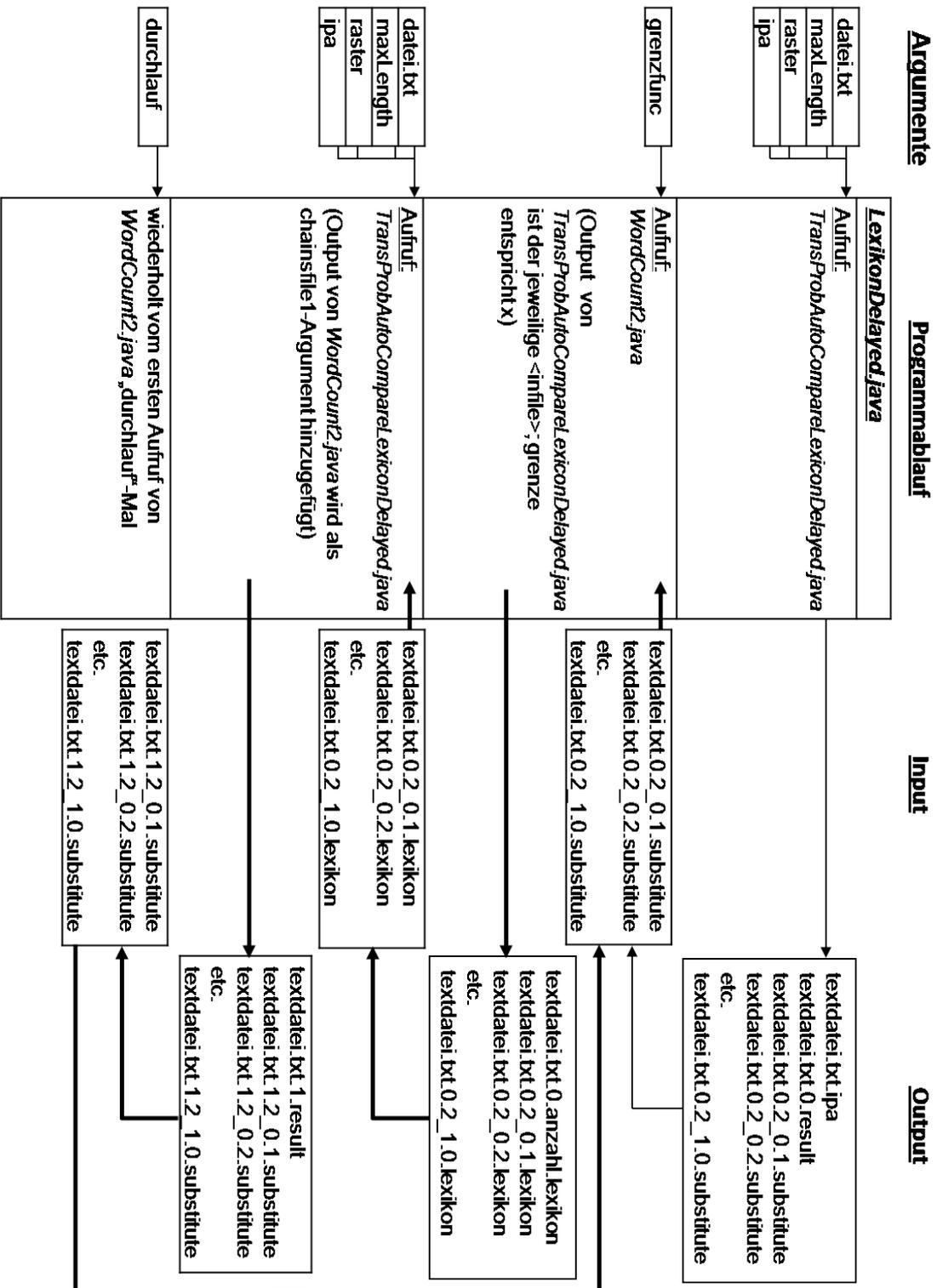


Abbildung 5.6: Informationsfluss der rekursiven Struktur des Algorithmus

Dabei muss natürlich erwähnt werden, dass der hier vorgeschlagene Ablauf empirisch nicht bewiesen ist, d.h. es gibt kein Experiment, welches beweist, dass die einzelnen kognitiven Fähigkeiten auch so interagieren, wie es im Modell simuliert wurde. Allerdings gibt es auch keinen Grund, etwas Gegenteiliges anzunehmen. Im Popper'schen Sinn gibt es derzeit noch keinen negativen Beweis, der die Funktionsweise des Modells in Frage stellen würde. Das Ziel einer Modellierung kann natürlich auch nicht eine Abbildung von Prozessen sein, die noch gar nicht erforscht sind. Auf diese Weise kann das Simulationsmodell als ein Vorschlag verstanden werden, einzelne bereits erwiesene Teilprozesse zu einem ganzheitlichen Prozess zu vereinen. Es sollte demnach zeigen, dass es theoretisch möglich wäre, einen Lautstrom unter Berücksichtigung der Fähigkeiten von Kleinkindern zu segmentieren. Natürlich stimmen die Ergebnisse optimistisch und es wäre wünschenswert, wenn der Algorithmus als Vorlage für künftige Experimente dienen könnte.

Letztlich sollen noch einige Bemerkungen zur Effizienz des Algorithmus erfolgen. Viele Autoren haben gezeigt (vgl. Brent, 1999b), dass sich in der hier als Phase 2 bezeichneten Modellstufe, komplexere Berechnungen und Vergleiche durchführen lassen (*Mutual Information, Minimum Description Length*). Es ist ebenso denkbar, dass *Strings* unterschiedlicher Länge miteinander verglichen werden und so womöglich eine abgeänderte Häufigkeitsverteilung entsteht. Diese Änderung ist ohne Weiteres möglich. Hier werden jetzt alle mathematisch möglichen Kombinationen berechnet. Mit kleineren Korpora wurden diese Berechnungen auch explorativ durchgeführt. Es konnten keine nennenswerten Effekte in der Leistungsmessung des Outputs erkannt werden. Der große Nachteil dieser Erweiterungen ist der enorme Rechenaufwand, der dadurch geleistet werden muss. Schon bei einem Satz mit 20 Zeichen, müssten bereits 220 Vergleiche durchgeführt werden. Bei repräsentativen Szenarien werden aber 10.000 *Token* verwendet. Schon allein die Darstellung der Auswertung wäre unverhältnismäßig, da man aufgrund von versteckten Minima beziehungsweise Maxima nicht einfach die größten Werte ausgeben könnte. Eine mögliche Lösung dieses Problems bestünde in der Einführung einer weiteren Zufallsverteilung. Damit wäre es möglich, eine begrenzte Auswahl zu treffen, welche auf Dauer und im Durchschnitt auch die Grundgesamtheit aller Kombinationen wiedergeben würde. Fraglich bleibt dabei aber, ob die daraus resultierenden Effekte nennenswert sind oder ob sie noch den Fähigkeiten der Kleinkinder entsprechen. Grundsätzlich wären derartige Erweiterungen für detailliertere Erkenntnisgewinne aber zu begrüßen.

5.5 Zusammenfassung

Das fünfte Kapitel dieser Arbeit zielte auf die stochastische Modellierung der Wortsegmentierung ab. Eingangs wurden einige Vorüberlegungen zur generellen Architektur angestellt und gezeigt, wie sich diese auf eine Simulation beziehen lassen. An dieser Stelle erfolgten eine kurze Benennung und eine Begründung der allgemeinen Annahmen und der Randbedingungen von Simulationen. Im Mittelpunkt dieser Ausführungen stand die Beziehung zwischen dem Input und der Sprachverarbeitung $v(i)$.

Im zweiten Unterkapitel (5.2) erfolgte die Herleitung der relevanten Maßzahlen und

5 Modellierung einer stochastischen Sprachsegmentierung

Parameter sowie die Formel zur Berechnung der Übergangswahrscheinlichkeiten. Aus Gründen der Vergleichbarkeit fiel die Wahl auf das F_1 -Maß, das Präzision und Vollständigkeit gleich gewichtet. Die Übergangswahrscheinlichkeiten wurden, wie üblich als bedingte Wahrscheinlichkeiten errechnet und die Größe des Paralexikons geschah durch das Abzählen von neuen Einträgen. Weiterhin wurde eine Unterscheidung in primäre und sekundäre Variablen vorgenommen. Die primären Variablen (σ , λ , ζ) kontrollierten die interne Sprachverarbeitung und die sekundären (ψ , ω) die Inputvariationen. μ wurde als Drittvariable angesehen. Genauer gaben ψ die Größe des Korpus, ω die Anzahl der Durchläufe, σ die Sensibilität, λ die Länge einer Phonemkette, ζ die Lexikonfunktion und μ den Zeitpunkt des Rekursionsbeginns an.

Im sich anschließenden Kapitel (5.3) wurde die Inputseite beschrieben. Dafür sorgte eine ausführliche Beschreibung über die Zusammensetzung der Korpora, deren Herkunft und die Frage der Repräsentativität für das notwendige Verständnis. Eine vollständige Repräsentativität der Korpora im Sinne Bibers (Biber, 1993) stellte sich im Rahmen dieser Arbeit als nicht realisierbar heraus.

Das letzte Kapitel (5.4) widmete sich ganz der Beschreibung des Algorithmus, also der Handhabung der primären Variablen. Zuerst wurden die Annahmen zusammenhängend mit einer einführenden Darstellung der Simulation als erklärenden Ansatz aufgeführt. Schließlich folgte eine detaillierte Beschreibung des Programms mit einer abschließenden kurzen Diskussion über die Neuerung des Algorithmus im Vergleich zu anderen Simulationen.

6 Simulationsergebnisse und Auswertung

6.1 Einleitende Bemerkungen

Die Visualisierung der Ergebnisse kann nur exemplarisch für eine Teilmenge der Ergebnisse vorgenommen werden. Diese Teilmenge repräsentiert aber alle wichtigen Eigenschaften und zeigt ebenso das Gesamtverhalten der Berechnungen und die Entwicklungen unter den vorgenommenen Parameteränderungen. Eine derartige Zusammenstellung ist im Sinne einer konzisen Beschreibung, bei welcher der Blick für das Wesentliche nicht verloren gehen darf, notwendig.¹

Die anschließende Auswertung der Simulationsergebnisse erfolgt in vier Unterkapiteln: Zuerst werden die optimalen Parameterausprägungen der drei primären Variablen bestimmt. Daraufhin wird die Wirkung der rekursiven Strukturen auf die Entwicklung des Paralexikons und deren Segmentierungsleistung näher untersucht. Damit geht auch eine Revision der anfänglichen Parametereinstellungen einher, weil sich hier herausstellte, dass die primären Variablen mit denen der Entwicklung des Paralexikons zusammenwirken. Das Vorhandensein eines Lexikons ändert die Segmentierungsmechanismen und so beeinflusst auch diese Synergie die Segmentierungsleistung. Im dritten Unterkapitel steht die Lexikonfunktion im Mittelpunkt. Ihr Einfluss auf die Größe des Paralexikons wird ebenso beleuchtet, wie die Auswirkungen der Perioden unter Verwendung verschiedener Korpora. Schließlich widmet sich das letzte Kapitel einer genauen Untersuchung des Paralexikons und seiner möglichen Eigenschaften. Zudem wird gezeigt, wie andere Mechanismen auf Grundlage des Paralexikons zur weiteren Segmentierung genutzt werden könnten. Die ersten zwei Unterkapitel sind in zwei Teile gegliedert. Im ersten Teil wird das Datenmaterial beschrieben und im zweiten werden Interpretationen als mögliche Erklärungen angeboten. Diese Aufteilung ist für die Unterkapitel drei und vier nicht notwendig, weil die Datenbestände klein und übersichtlich sind.

6.2 Bestimmung der genauen Parameterwerte

6.2.1 Beschreibung der Simulationsdaten

Die 17 untersuchten Korpusrepräsentanten wurden zunächst hinsichtlich der Anzahl der Phoneme und der Sensibilität untersucht. Die typische Verteilung der Punktmengen ist

¹Für eine vollständige Auflistung des Datenmaterials und der dazugehörigen Daten sei auf eine DVD-Sammlung verwiesen, welche am Fachgebiet Linguistik – Computerphilologie der Universität Kassel vorliegt. Dort lassen sich die im Folgenden beschriebenen Verläufe detaillierter nachvollziehen.

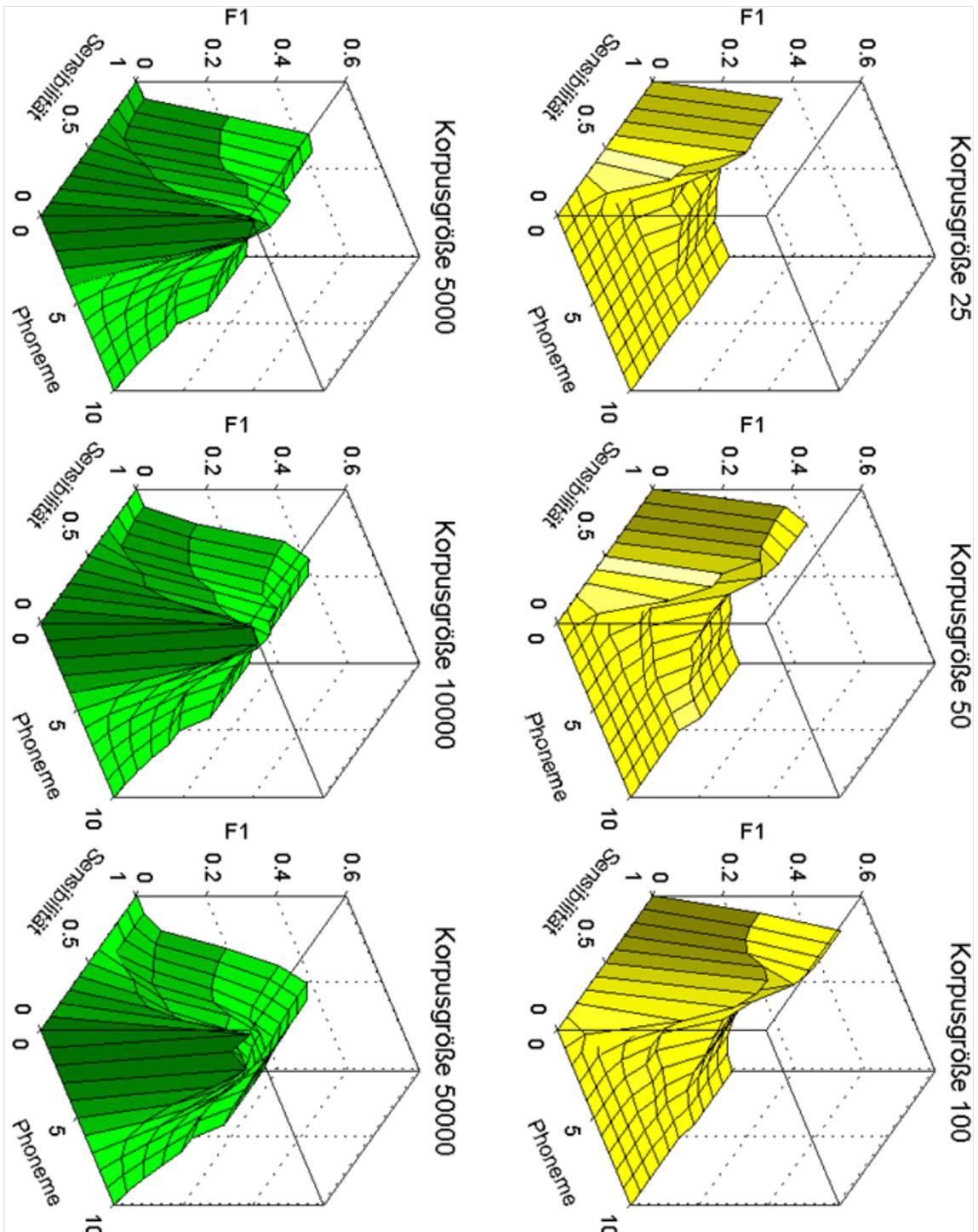


Abbildung 6.1: $F_1(\sigma, \lambda)$ für verschiedene ψ

aus Abbildung 6.1 zu entnehmen. Verhaltensunterschiede waren, wie erwartet, zwischen großen und kleinen Korpora augenfällig. Mit zunehmender Größe gleichen sich die funktionalen Verläufe aber einer typischen Gestalt an (wie in der unteren Reihe von Abbildung 6.1), die dann keinen relevanten Änderungen mehr unterliegen, unabhängig von der Größe des Korpus.² Ab einer Korpusgröße von 5000 Zeichen ist dieser Zustand erreicht. Korpora zwischen 1000 und 5000 Zeichen unterliegen nur noch minimalen Schwankungen mit wenigen Ausreißern. Korpora, die kleiner als 1000 Zeichen sind, zeigen dagegen sehr deutliche Veränderungen. Kleine Korpora können nur bei sehr hoher Sensibilität (> 0.5) und möglichst kleinen Vergleichseinheiten (Anzahl der Phoneme = 1) noch vernünftige Segmentierungen aufweisen. Werden die Phonemketten größer, läuft F_1 gegen Null, d.h., es werden keine richtigen Segmentierungen vollzogen. Recht überraschend ist jedoch das Ergebnis, dass selbst bei Korpora mit nur 25 Zeichen (weniger als 10 Wörter) F_1 ein gutes Drittel beträgt. Das entspricht ungefähr der Hälfte des Wertes von größeren Korpora. Schon bei einem Korpus von 25 bis 30 Wörtern (100 Zeichen) ist die Segmentierungsleistung mit der von großen Korpora vergleichbar, wenn auch hier nur für sehr bestimmte Parameterwerte (hohe Sensibilität und kurze Phonemketten). Mit zunehmender Korpusgröße verringert sich proportional die Sensibilität und die Länge der Phonemketten ist nicht mehr entscheidend.

Die in Abbildung 6.1 dargestellten Punktmengen können zum besseren Vergleich der Korpusgrößen in zweidimensionalen Schnitten übersichtlich visualisiert werden. Die Abbildungen 6.2-6.9 bilden diese Form für konstante Phonemlängen und konstante Sensibilitäten ab.

Zuerst sei die Anzahl der Phoneme konstant gehalten. Ein zweidimensionaler Schnitt erlaubt so, die Entwicklung der Sensibilität von unterschiedlichen Korpusgrößen vergleichend zu betrachten (Abb. 6.2, 6.3, 6.4 und 6.5). Bei sehr niedriger Sensibilität ($\sigma = 0.1$) und bei Mittelung über ein Phonem (Abb. 6.2 und 6.3) sind die unterschiedlichen Verläufe von kleinen und großen Korpora beträchtlich. Nimmt die Anzahl der Phoneme zu, heben sich die Unterschiede zwischen den unterschiedlichen Korpusgrößen graduell auf (Abb. 6.4 und 6.5).

Nun wird das Verhalten von Korpora ab 1000 (circa 250 Wörter) bis 450.000 Zeichen (112.000 Wörter) bei einer konstanten Phonemlänge von eins näher beschrieben (Abb. 6.2). Alle 11 Korpora dieser Größenkategorie weisen bei geringer Sensibilität ein Maximum von F_1 auf unabhängig von der genauen Größe des Korpus. Lediglich der 100.000-Zeichen-Korpus (circa 23.000 Wörter) weicht stärker von dem arithmetischen Mittelwert für F_1 von 0.55 und mit einer Standardabweichung von 0.05 ab. Trotzdem folgen auch die Ergebnisse dieses Ausreißers den typischen Verläufen der übrigen Korpora. Korpora ab dieser Größe zeigen zudem noch eine Besonderheit bei mittlerer Sensibilität. Hier bleiben die Segmentierungswerte noch konstant bei F_1 -Werten um 0.1 und fallen erst bei weiterer Zunahme der Sensibilität ab.

Für die untersuchten Korpora lässt sich also festhalten, dass sehr geringe Sensibilitäten für eine hohe Segmentierungseffizienz notwendig sind und dass ab einem Schwellenwert

²Korpora größer als 20.000 Zeichen erhöhen lediglich noch ihre Robustheit bei langen Phonemketten.

6 Simulationsergebnisse und Auswertung

von circa 250 Wörtern die Korpusgröße darauf keinen Einfluss hat. Korpora ab dieser Größe haben zudem den Vorteil, dass sie selbst bei höheren Sensibilitäten noch gute Segmentierungswerte aufweisen und die Länge der Phonemketten variieren kann. Es existieren aber zwei Kombinationen aus Phonemkettenlänge und Sensibilität, die bei größeren Korpora möglichst nicht gewählt werden sollten:

- a eine hohe Sensibilität und sehr kurze Phonemketten (z.B. $\lambda = 1$ und $\sigma = 1$) und
- b eine niedrige Sensibilität und sehr lange Phonemketten (z.B. $\lambda = 10$ und $\sigma = 0.1$)

Für diese zwei Szenarien liegen die F_1 -Werte bei Null. Für die simulierten Daten konnten dafür keine Ausnahmen gefunden werden.

Während bei Korpora unter 1000 Zeichen das erste Szenario als einzig mögliches zu favorisieren wäre, würde sich das zweite als ebenso unvorteilhaft hervortun wie bei großen Korpora. Korpora kleinerer Größe (< 1000 Zeichen) sind zudem volatil (Abb. 6.3), d.h., ihre Verläufe können nicht einheitlich beschrieben werden. Dennoch lassen sich graduelle Entwicklungen, wie die Tendenz zu hohen Sensibilitäten, gut erkennen. Auffällig ist auch hier, dass Korpora unter einer Zeichenlänge von 1000 bei mittleren Sensibilitäten keine Steigung aufweisen, also indifferent sind, bevor sie dann eine bestimmte Sensibilitätsschwelle überschreiten und mit zunehmender Sensibilität auch höhere F_1 -Resultate erzielen. Mit Zunahme der Größe verschieben sich die Maxima dann deutlich zu kleineren Sensibilitätswerten. Bei einer Größe von ungefähr 50 Wörtern (250 Zeichen) ist die Verhaltensähnlichkeit zu größeren Korpora zu erkennen.

Betrachtet man die Parameterveränderungen zusammenhängend für alle Phonemlängen (in Abbildungen 6.2-6.5 für λ von 1, 5 und 9 beispielhaft verbildlicht), erkennt man die eingeschränkte Wirkung der Korpusgröße auf die Segmentierungsergebnisse. Bei Phonemketten der Größe fünf lässt sich dieser Trend bereits eindeutig erkennen (Abb. 6.4). Hier zeigen nur noch Korpora über 20.000 Wörtern ein leicht divergentes Verhalten, welches sich aber bei längeren Phonemketten sehr genau dem Trend der Mehrheit anpasst (Abb. 6.5). Wie schon aus Abbildung 6.1 zu entnehmen ist, verändert sich mit Zunahme der Länge der Phonemketten, die Wichtigkeit der Sensibilität für die Segmentierungen. Die Maxima von F_1 staffeln sich nun in direkter Proportionalität zur Korpusgröße und die Länge der Phonemketten wird bei einer Sensibilität von 0.6 unabhängig.

Dieser Zusammenhang lässt sich anschaulich darstellen, wenn die Sensibilität konstant gehalten wird (siehe Abb. 6.6-6.9 für beispielhafte Darstellungen mit konstanten Sensibilitäten von 0.1, 0.6 und 1.0). Die Kurvenverläufe von den Abbildungen 6.6 und 6.7 sind kaum unterscheidbar, obwohl sich die Sensibilität von 0.6 auf 1.0 erhöht hat. Die F_1 -Werte bleiben faktisch für die unterschiedlichen Korpusgrößen gleich. Es ist lediglich ein leichtes Absinken der Funktionenschar an beiden Extremen (bei Phonemkettenlänge=1 und Phonemkettenlänge=10) auszumachen. Aus diesen Darstellungen lassen sich aber auch noch weitere Aussagen hinsichtlich der idealen Phonemkettenlänge treffen. Abbildungen 6.8 und 6.9 zeigen, dass bei Korpora über 10.000-20.000 Zeichen und einer sehr geringen Sensibilität (hier 0.1), die Phonemkettenlänge zwischen eins und fünf liegen kann, ohne dass dies die Segmentierungsleistung entscheidend beeinflusst. Dies ist

bei den Korpora kleinerer Größe nicht der Fall, da die F_1 -Werte nur bei einer Länge von eins überhaupt Ergebnisse liefern. Mit steigender Sensibilität schwächt sich der Effekt auch bei größeren Korpora ab. In der Tat liegen dann die höchsten F_1 -Werte nur bei mittleren Phonemkettenlängen bei großen Korpora und bei kleinen Korpora weiterhin bei einer Phonemkettenlänge von eins.

6.2.2 Dateninterpretation und Erklärungen

Soll der statistische Ansatz der Wortsegmentierung aus theoretischer Perspektive psychologisch plausibel sein, so muss eine Erklärung für die unterschiedlichen Sensibilitätswerte bei Korpusgrößen unter 1000 Zeichen gefunden werden. Die Länge der Phonemkette ist dagegen nicht widersprüchlich, weil bei allen Korpusgrößen eine Länge von eins für sehr gute Segmentierungen ausreichend ist. Man könnte also annehmen, dass das Kind meist sehr kleine Lautketten bevorzugt. Das ist bei der Sensibilität nicht der Fall.

Aus Abbildung 6.1 lässt sich sehr leicht erkennen, dass die kleinen Korpusgrößen genau dort ihr höchstes Leistungsspektrum an F_1 -Werten besitzen, wo bei den großen Korpusrepräsentanten überhaupt keine Segmentierung erreicht werden konnte. Wie lässt sich dieses Ergebnis in Bezug auf die Realität des Erstspracherwerbs erklären? Es ist wohl kaum anzunehmen, dass ein Kleinkind, je nach unterschiedlicher Größe des Sprachinputs, seine Sensibilität von wiederkehrenden Lautstrukturen ändert.

Eine Möglichkeit besteht darin, eine bestimmte Quantität an Sprachinput vorauszusetzen, damit überhaupt der Wortsegmentierungsprozess beim Kind beginnen kann. Das würde dann bedeuten, dass die Wortsegmentierungsmechanismen erst nach längeren Gesprächen mit den Säuglingen anfangen einzusetzen. Diese Interpretation ist durch eine gewisse Einfachheit vorteilhaft. Man würde nach dieser Logik annehmen müssen, dass der Sprachlerner gegenüber wiederkehrenden Lautstrukturen, schon nach wenigen, vielleicht nur zwei oder drei zusammenhängenden Lautketten in unterschiedlichen Lautumgebungen, wohl aber auf Grundlage eines ausreichenden Gesamtinput, in der Lage ist, die Zusammengehörigkeit dieser Einheiten festzustellen.

Die Strategie des Sprachlerner wäre also, auf eine absolute Sicherheit zu verzichten, zu Lasten einer eventuellen Fehleranfälligkeit für falsche Segmentierungen. Genau dies kommt der Struktur des Englischen und womöglich überhaupt natürlicher Sprachen entgegen, weil eingebettete Lautstrukturen zum Einen freie Morpheme und zum Anderen auch nur Teile von Morphemen sein können, wie im folgenden Beispiel illustriert werden kann: Die Sätze */hɪz jɔː bɔːl. dɔːgi teɪks ɪt. səʊ ðɛr dɔːgi bɜːks. ɪts hɪz bɔːl naʊ. hi sɪts daʊn./* (*Here's your ball. Doggy takes it. So there doggy barks. It's his ball, now. He sits down*) müssen ohne Lücken als */hɪzjɔːrbɔːldɔːgiteɪksɪtsəʊðɛrdɔːgibɜːksɪtshɪzbɔːlnaʊhɪsɪtsdaʊn/* verarbeitet werden. */sɪts/* tritt über Wortgrenzen hinweg auf, enthält */ɪt/* als eigenständiges Morphem oder besteht aus zwei Morphemen. Eine geringere Sensibilität hätte hier den Vorteil, dass die Kombinationen */ts/* und */ks/* in */gets/* und */bɜːks/*, obwohl diese weniger häufig auftreten als */sɪ/*, doch die gleiche Berücksichtigung finden würden. Eine hohe Sensibilität erlaubt dies bei einer Phonemkettenlänge von eins nicht, weil die kriti-

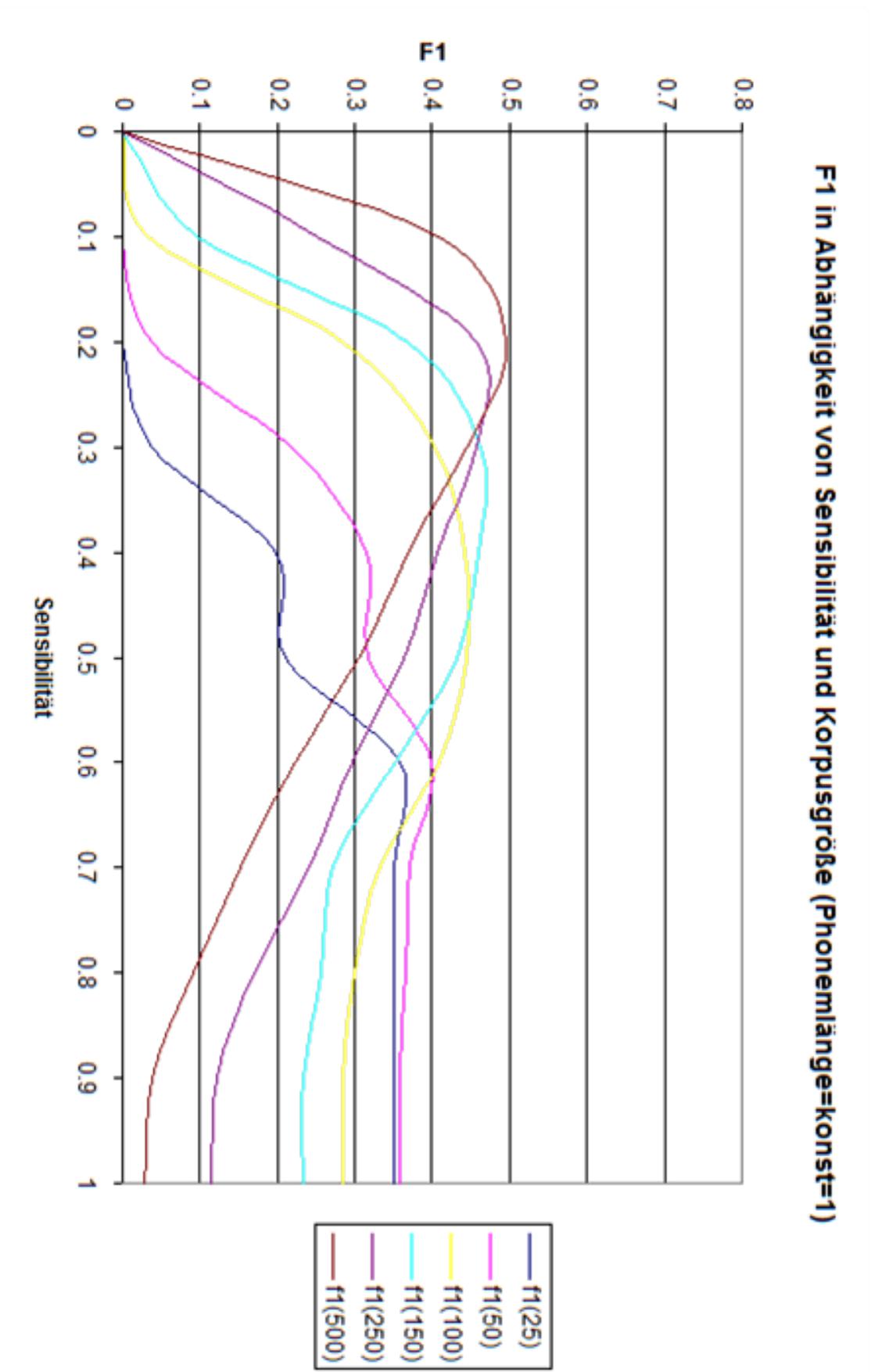


Abbildung 6.2: $F_1(\sigma)$, $\lambda = 1, \psi \leq 500$

F1 in Abhängigkeit von Sensibilität und Korpusgröße (Phonemlänge=konst=1)

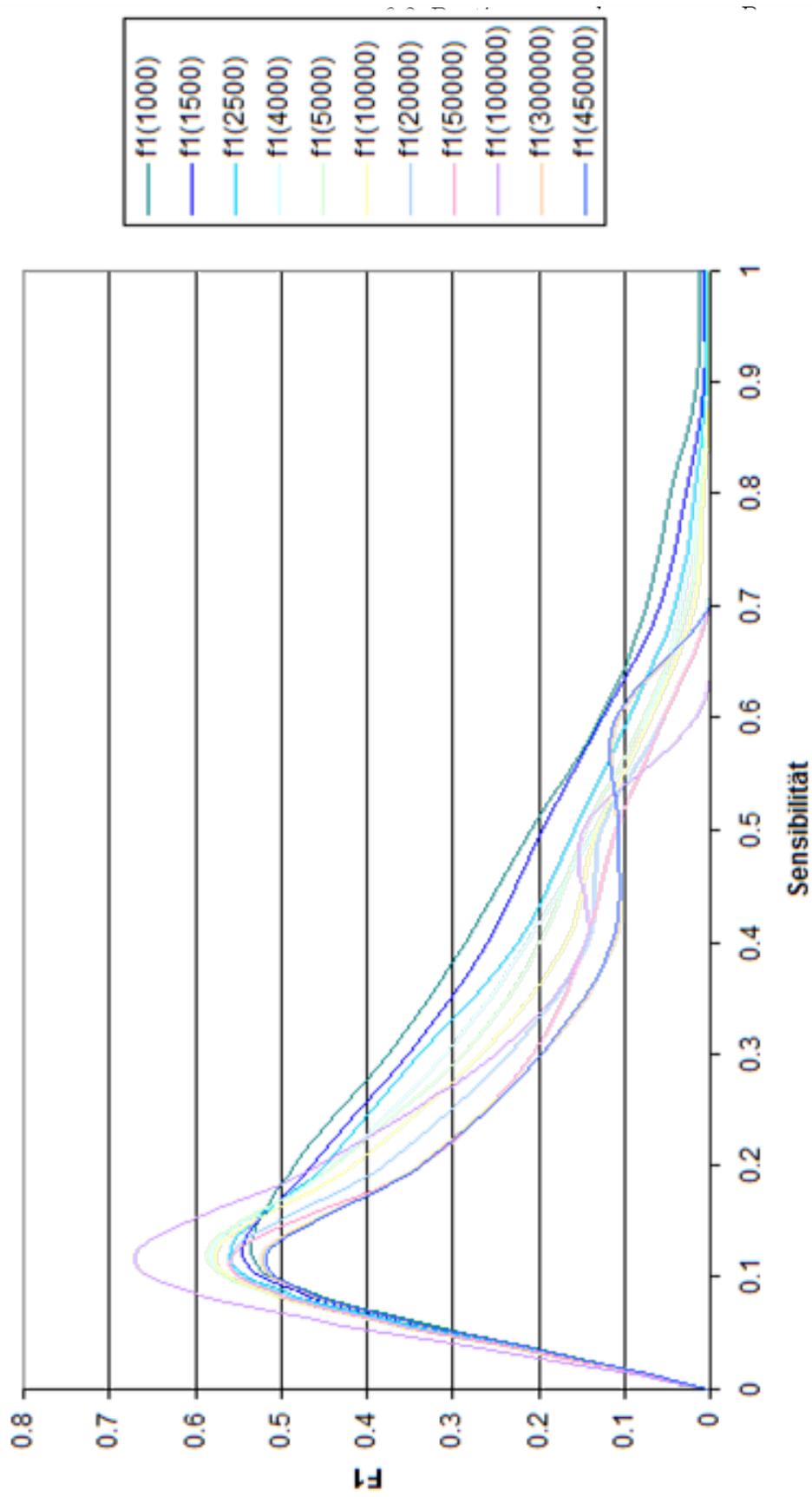


Abbildung 6.3: $F_1(\sigma), \lambda = 1, \psi \leq 1000$

erwerte

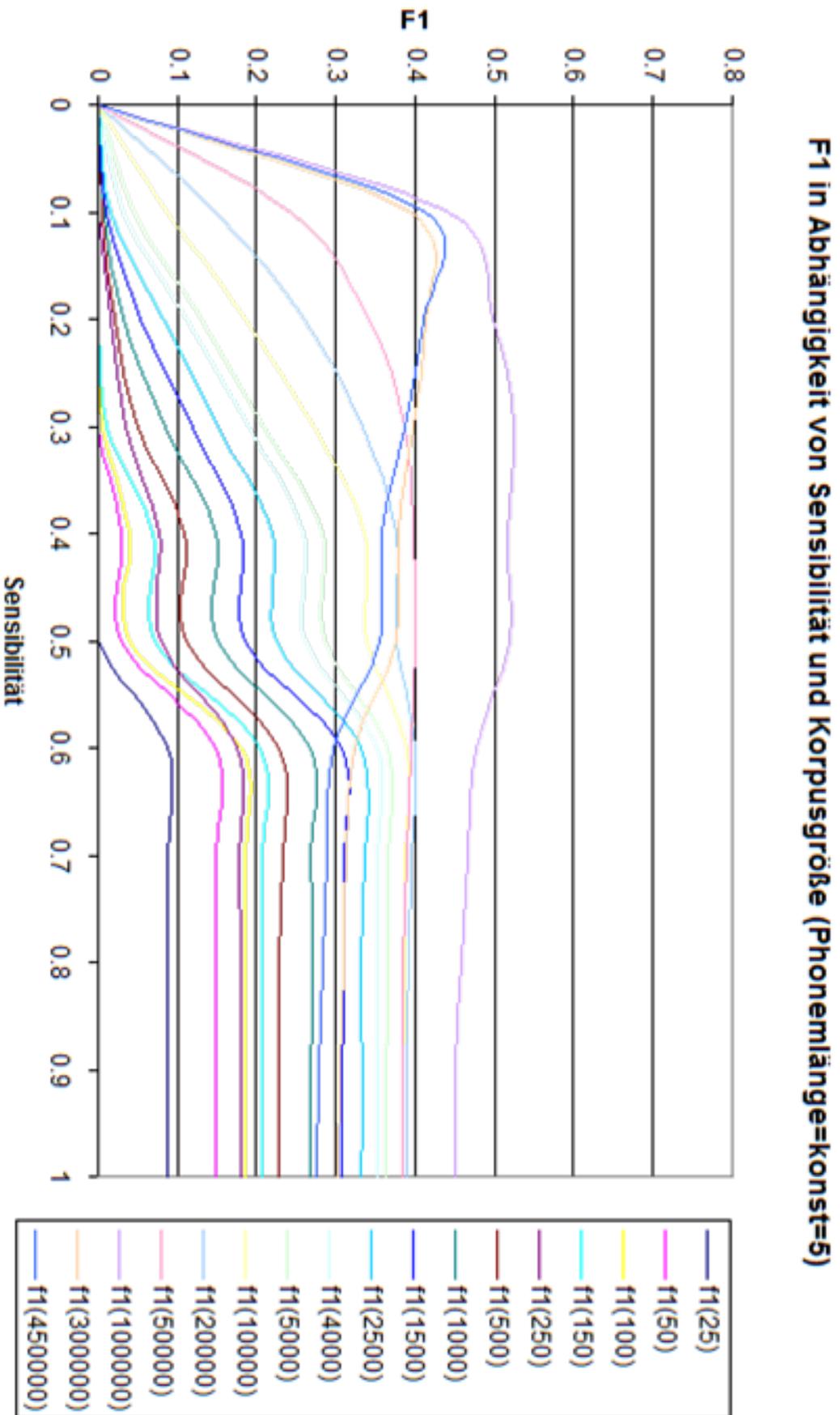
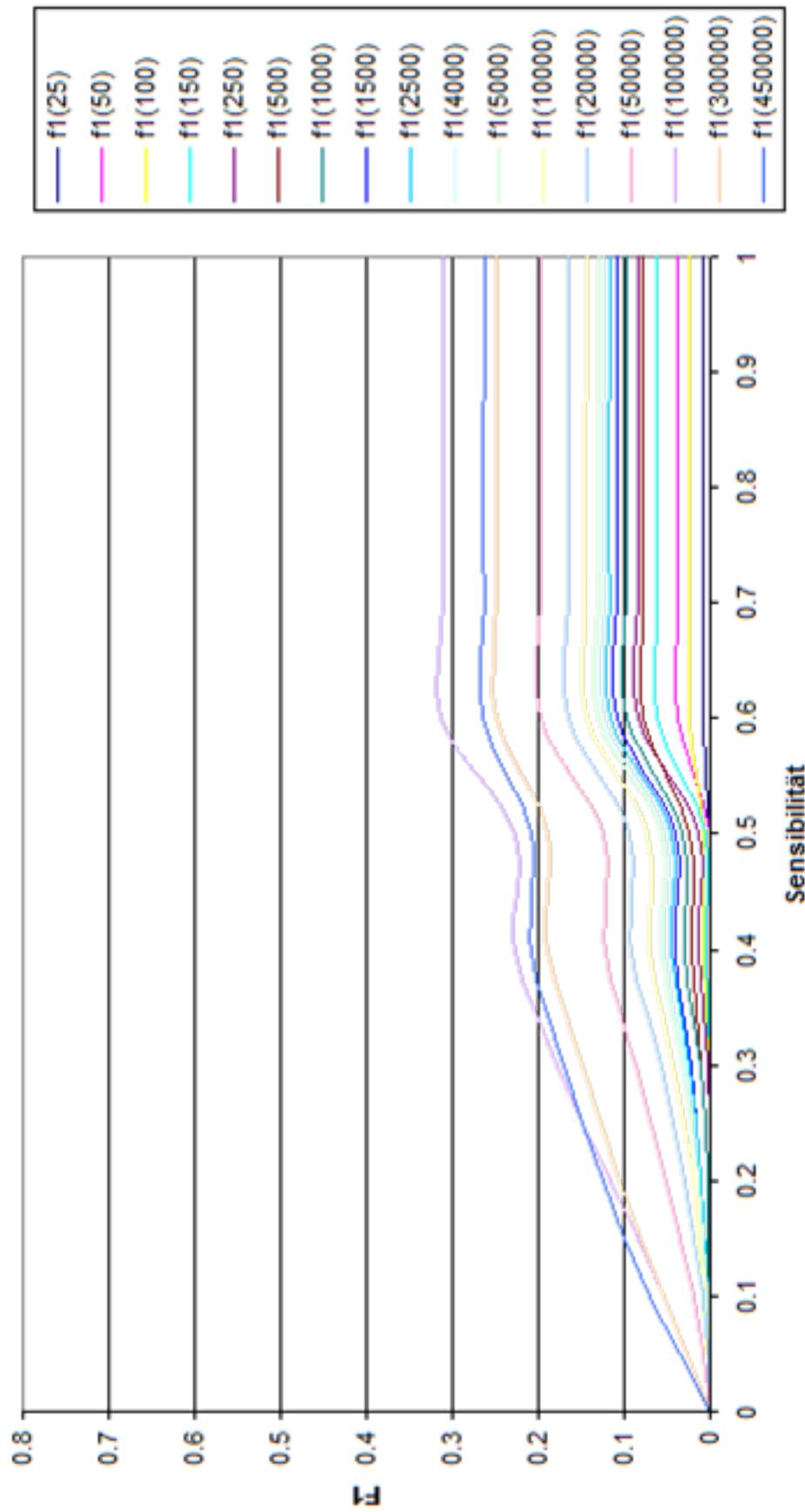


Abbildung 6.4: $F_1(\sigma)$, $\lambda = 5$

F1 in Abhängigkeit von Sensibilität und Korpusgröße (Phonemlänge=konst=9)



erwerte

Abbildung 6.5: $F_1(\sigma), \lambda = 9$

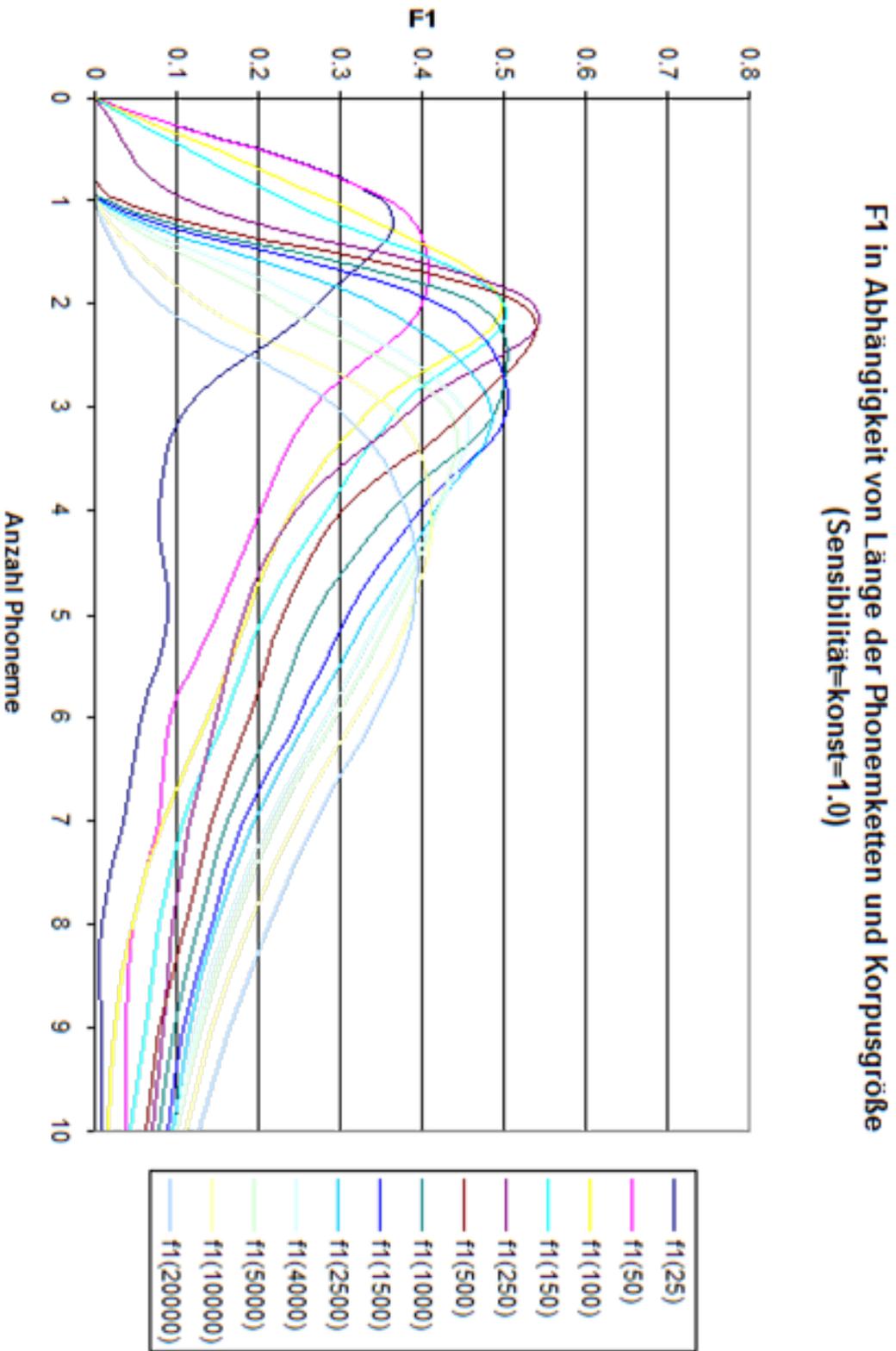


Abbildung 6.6: $F_1(\lambda)$, $\sigma = 1$, $\psi \leq 20000$

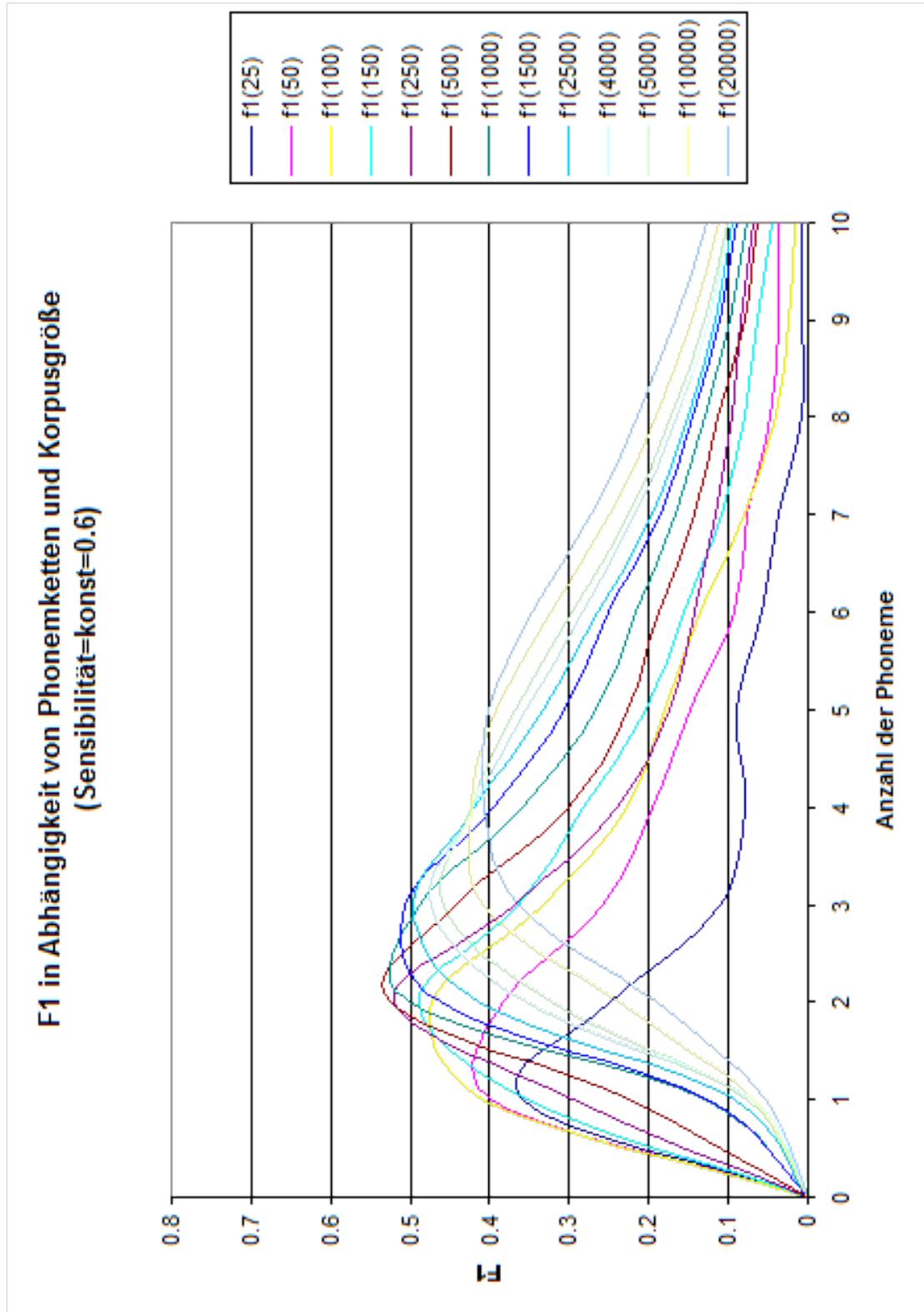


Abbildung 6.7: $F_1(\lambda), \sigma = 0.6, \psi \leq 20000$

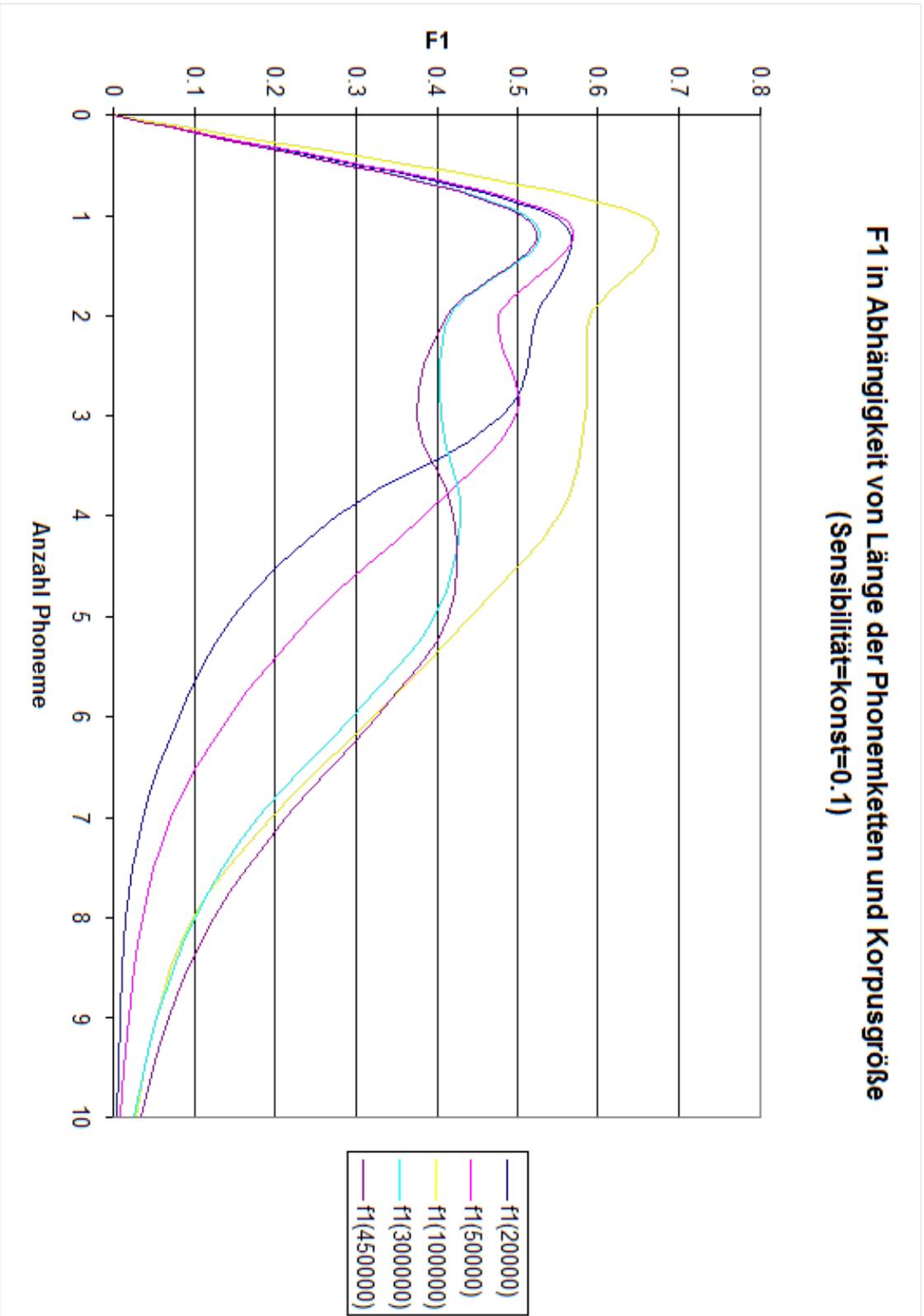


Abbildung 6.8: $F_1(\lambda)$, $\sigma = 0.1$, $\psi \geq 20000$

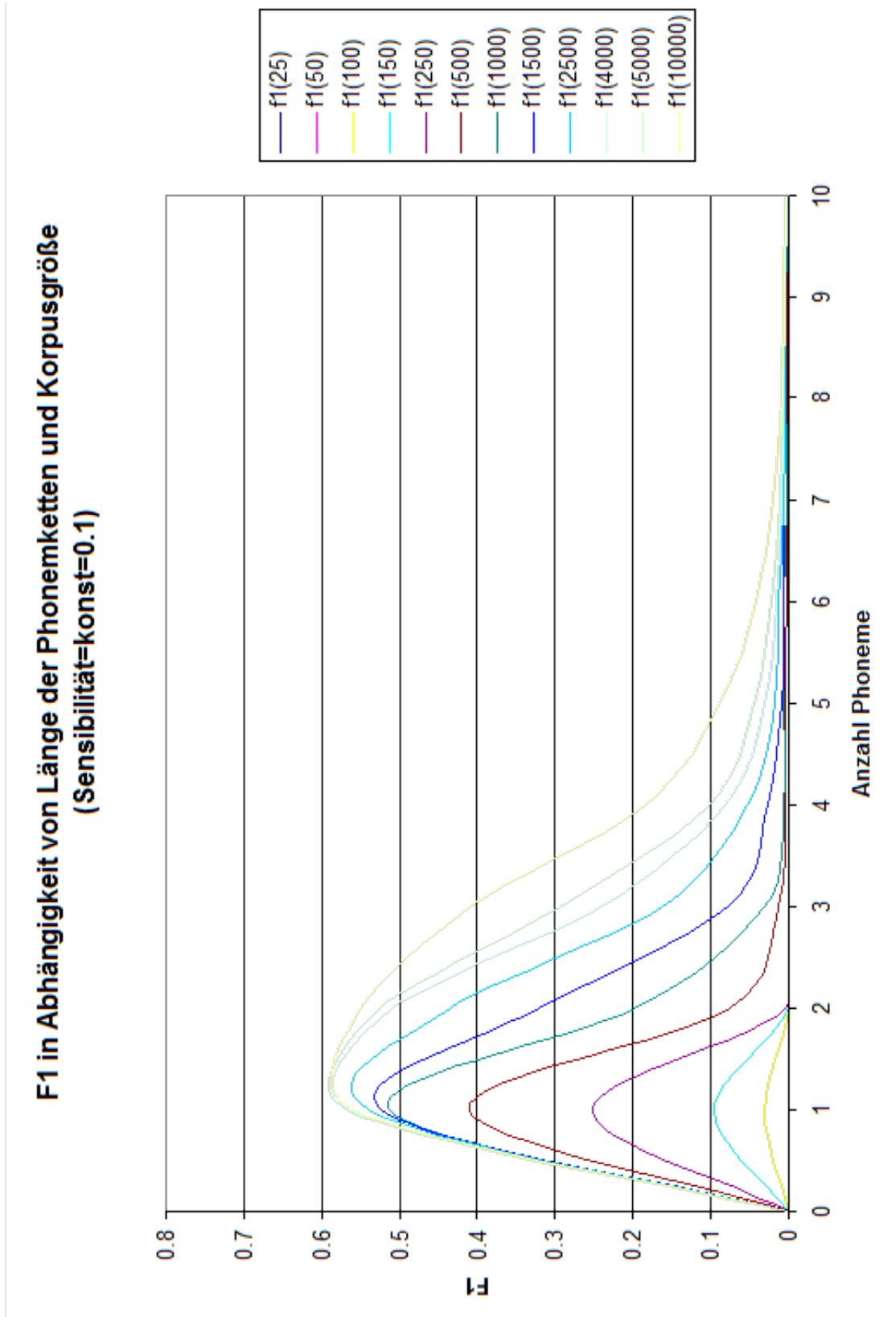


Abbildung 6.9: $F_1(\lambda), \sigma = 0.1, \psi \leq 10000$

6 Simulationsergebnisse und Auswertung

sche Grenze für die Entscheidung, eine Zusammengehörigkeit festzustellen, schlichtweg höher ist. Bei Mittelung über längere Phonemketten wird dieser Effekt abgeschwächt, denn je größer eine Phonemkette wird, desto geringer wird ihre Vorkommenshäufigkeit in einer bestimmten Umgebung. Das könnte dann auch zur Folge haben, dass */gets/* und */it/* korrekt segmentiert werden, da nun das Verhältnis von */tsit/* zu */ts/* errechnet wird.

An dem gegebenen Beispiel lässt sich auch das unterschiedliche Verhalten von kleinen Korpora erklären. Nimmt man den eben gegebenen Korpus als einzigen Input an, könnte eine niedrige Sensibilität wohl überhaupt keine Wortgrenze definieren. Dies liegt natürlich an der Eigenschaft der Sensibilität, welche die kritische Grenze der Übergangswahrscheinlichkeit definiert. Weil die Übergangswahrscheinlichkeit die Häufigkeit einer Einheit und ihres Nachfolgers ins Verhältnis zur Häufigkeit dieser Einheit ohne Nachfolger setzt, muss eine genügend große Zahl an Vorgängereinheiten überhaupt erst einmal vorhanden sein, um eine kleine Verhältniszahl (=geringe Sensibilität) zu erhalten. Im obigen Beispiel wird das */t/* in drei Umgebungen von einem */s/* gefolgt. Das */t/* kommt viermal im Korpus vor. Die Kombination */ts/* tritt demnach dreimal auf. Dadurch ergibt sich ein hoher Sensibilitätswert ($3/4=0.75$). Wenn das */t/* nun öfter auch in anderen Lautumgebungen auftreten würde (z.B. 15 Mal), wie es in größeren Korpora normalerweise der Fall ist, verringert sich der Sensibilitätswert entsprechend (hier $3/15=0.2$).

Bei kleinen Korpora handelt es sich also um nicht repräsentative Sammlungen, aus denen sich demzufolge überhaupt keine kleineren Sensibilitätswerte errechnen lassen. Deshalb können auch keine Segmentierungen vorgenommen werden; die berechneten Verteilungen liegen ja immer über dem angegebenen Grenzwert (*limit*). Wie die Simulation an verschiedenen Korpusgrößen gezeigt hat, beginnt sich mit zunehmender Größe eines Korpus ab 1000 Zeichen ein Gleichgewicht einzustellen, welches erst ab einer Größe von 5000 Zeichen den Gleichgewichtszustand erreicht. Ab dieser Größe ist ein Korpus so robust, dass genügend Phonemkombinationen vorhanden sind, um diese repräsentativ zueinander ins Verhältnis setzen zu können.

An dieser Stelle scheint eine erste Schlussfolgerung angebracht. Nimmt man an, dass es keinen Mechanismus gibt, nach dem der Sprachlerner die Sensibilität seiner Berechnungen mit der Textgröße abstimmt (also generell von einer geringen Sensibilität ausgeht), dann muss der Sprachlerner zunächst einen genügend großen Input ansammeln. Wird er nur mit kleinem Input konfrontiert, kann er keine Wortgrenzen finden. Es ist denkbar, dass kleinere Korpora dann als ganze Einheiten betrachtet werden beziehungsweise solange kleinere Inputdaten gesammelt werden bis die Größe ausreicht, um Muster erkennen zu können. An dieser Stelle könnte das Baby dann beginnen, die längeren Einheiten in kleinere Bestandteile zu zerlegen.³ Das bedeutet nicht, dass Kleinkinder generell kei-

³Dieser Befund stimmt auch mit dem Vorschlag von Tomasello (2003) und Vihman (1996) überein, der da lautet

... that young children have certain holistic phonological templates to which they assimilate their early words and even multi-word expressions, and that these are psychologically prior to any elementary so-called building blocks such as phonemes – which must be extrac-

ne kleineren Dialoge segmentieren können. Sobald sie ein erstes Paralexikon entwickelt haben, sind sie in der Lage auch andere Segmentierungsstrategien zu nutzen, welche sie sehr gut und sogar besser auf nur wenige Wörter beziehen können. Der Segmentierungsmechanismus, der hier auf die auditive Sprachwahrnehmung angewendet wurde, ist eine kognitive Fähigkeit, welche in anderen kognitiven Prozessen eine ebenso wichtige Ausgangsrolle spielt (Tomasello, 2003, 30).

Die Alternative dazu wäre ein Mechanismus, nach dem der Sprachlerner Sensibilität, Inputgröße und eventuell auch die Länge der Phonemketten aufeinander abstimmt. Dieser Mechanismus dürfte natürlich nicht den allgemeinen kognitiven Fähigkeiten von acht Monate alten Babys widersprechen, da man sich ansonsten wieder in die Nähe eines Sprachmoduls mit genauen Voreinstellungen begibt. Bislang konnte kein plausibler Mechanismus erdacht werden, sodass diese zweite Alternative verworfen werden musste. Man kann bei derzeitiger Wissenslage einfach nicht davon ausgehen, dass ein wenige Monate altes Baby einen Sprachinput bestimmter Größe erhält und aufgrund der Quantität des Inputs entscheidet, ob einige Lautkombinationen einmal mehr und einmal weniger in unterschiedlichen Umgebungen auftreten dürfen. Wahrscheinlicher ist es doch, dass der Sprachlerner grundsätzlich eine Strategie verfolgt, die da lautet: „Finde ich Lauteinheiten, die öfter zusammen auftreten, nehme ich an, dass diese zusammengehören, auch wenn ihre Bestandteile hin und wieder in anderen Umgebungen auftreten“. Die Sensibilität drückt sich in der hier formulierten Strategie in „hin und wieder“ aus. Eine hohe Sensibilität lässt ein „hin und wieder“ nicht zu, sondern lediglich ein „nur in einer einzigen“.

6.2.3 Zusammenfassung

Mit dem analysierten Datenmaterial war es möglich, Aussagen über die idealen Parameterwerte für eine effiziente Wortsegmentierung zu treffen. Hohe Segmentierungswerte richten sich nach der optimalen Kombination der Phonemkettenlänge und der Sensibilität in Bezug auf eine bestimmte Korpusgröße. Die Länge der Phonemkette ist für alle Korpusgrößen konsistent. Es zeigte sich, dass die Phonemkette möglichst klein zu wählen ist. Der Diphonbereich oder auch nur das einzelne Phonem als Grundlage der Berechnungen liefern die besten Ergebnisse für alle Korpusgrößen gleichermaßen. Andere Phonemlängen sind zwar auch möglich; gehen sie jedoch über eine Länge von fünf bis sechs Phonemen hinaus, werden auch hier keine Wortgrenzen mehr erkannt. Ein positiver Aspekt langer Lautketten ist, dass die Wahrscheinlichkeit einer zusammengehörigen Einheit sehr viel höher sein muss. Problematisch wäre es dann aber, dass man bei sehr langen Phonemketten, die Sensibilität darauf abstimmen müsste. Ein solcher Abstimmungsmechanismus entbehrt jedoch jeglicher Plausibilität zum derzeitigen Wissensstand. Somit kann die Phonemkettenlänge zwar divergieren, sollte aber nicht über fünf Einheiten hinausgehen.

ted from these larger structures ... The process is thus from whole to parts. (Tomasello, 2003, 61)

Aus dem gleichen Grund sollte man auch von einer niedrigen Sensibilität ausgehen. Auch wenn kleine Korpora dann nicht segmentiert werden können, bietet eine niedrige Sensibilität doch weit aus mehr Erklärungssicherheit, weil hier die allgemeine Fähigkeit der Mustererkennung herangezogen werden kann, die zum Beispiel auch in visuellen Verarbeitungsmechanismen bei Primaten einen sehr ähnlichen Verarbeitungsprozess nutzt (Tomasello, 2003). Abhängig von der Häufigkeit, aber eben nicht in „algebraisch-absoluter“ Analogie, werden Muster ihrem Umfeld nach in eine Struktur gezwungen, welche es dem menschlichen Gehirn ermöglicht, seine Ressourcen effizienter in den notwendigen Berechnungen einzusetzen.

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

6.3.1 Aufbereitung der Daten

Eine wichtige Beobachtung, welche für alle untersuchten Korpora Gültigkeit besitzt, ist, dass das Verhältnis der F_1 -Extrempunkte zueinander bei gegebener Sensibilität und Phonemlänge gleich bleibt (siehe Abbildung 6.10 als exemplarischen Fall eines 50.000 Zeichen Korpus für drei Lexikongrößen und konstanter Phonemkettenlänge von 10).⁴ Diese vorteilhafte Eigenschaft der Simulationsergebnisse kann genutzt werden, um neue Erkenntnisse zu gewinnen. Alle Maxima, welche durch ihre Parameterkombination von *limit* (σ) und *laenge* (λ) eindeutig bestimmt sind, können durch eine Funktion $argmax(F_1)$ beschrieben werden. Diese Funktionenschar kann dann in Abhängigkeit von der Anzahl der Rekursionen und in Abhängigkeit von der Lexikongröße berechnet werden. Dies soll hier für unterschiedliche Lexikongrößen und die Anzahl der rekursiven Durchläufe erfolgen. Es macht dabei Sinn, beide Dimensionen in Abhängigkeit voneinander darzustellen, um zu sehen, wie sich die Lexikongröße auf die Segmentierungsleistungen der Durchläufe auswirkt.

Diese Darstellung wird in Abb. 6.11 und 6.12 für verschiedene Korpusgrößen verbildlicht. Abermals ist eine Unterscheidung in große und kleine Korpusrepräsentanten angebracht. Für Korpora bis 1000 Zeichen lassen sich keine Unterschiede feststellen. F_1 ist lediglich niedriger. Bei einem kleinen Korpus kann man erkennen, dass weder die Anzahl der Rekursionen noch die Größe des Lexikons einen entscheidenden Einfluss auf die Segmentierungsleistung haben. Mit zunehmender Größe des Korpus werden die Einflüsse dieser beiden Variablen, Anzahl der Durchläufe (ω) und Lexikongröße ζ , immer deutlicher. Insbesondere fällt ein steiler Anstieg nach der ersten Rekursion bei allen Korpora über 5000 Zeichen auf. Nach dem zweiten Durchlauf bleibt der Anstieg konstant oder setzt sich nur sehr leicht fort. Interessant ist jedoch, dass die unterschiedlichen Längen der Phonemketten sich immer mehr einander angleichen bis sie bei einer Lexikongröße von 30 Wörtern nach dem fünften Durchlauf gleiche Segmentierungsleistungen errechnen können.

⁴Man bedenke, dass die Parameter für Sensibilität und Phonemlänge aus Gründen der Veranschaulichung übertrieben gewählt wurden, weil hier die Steigungen $\frac{dF_1}{d\sigma}$ sehr deutlich zu erkennen sind.

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

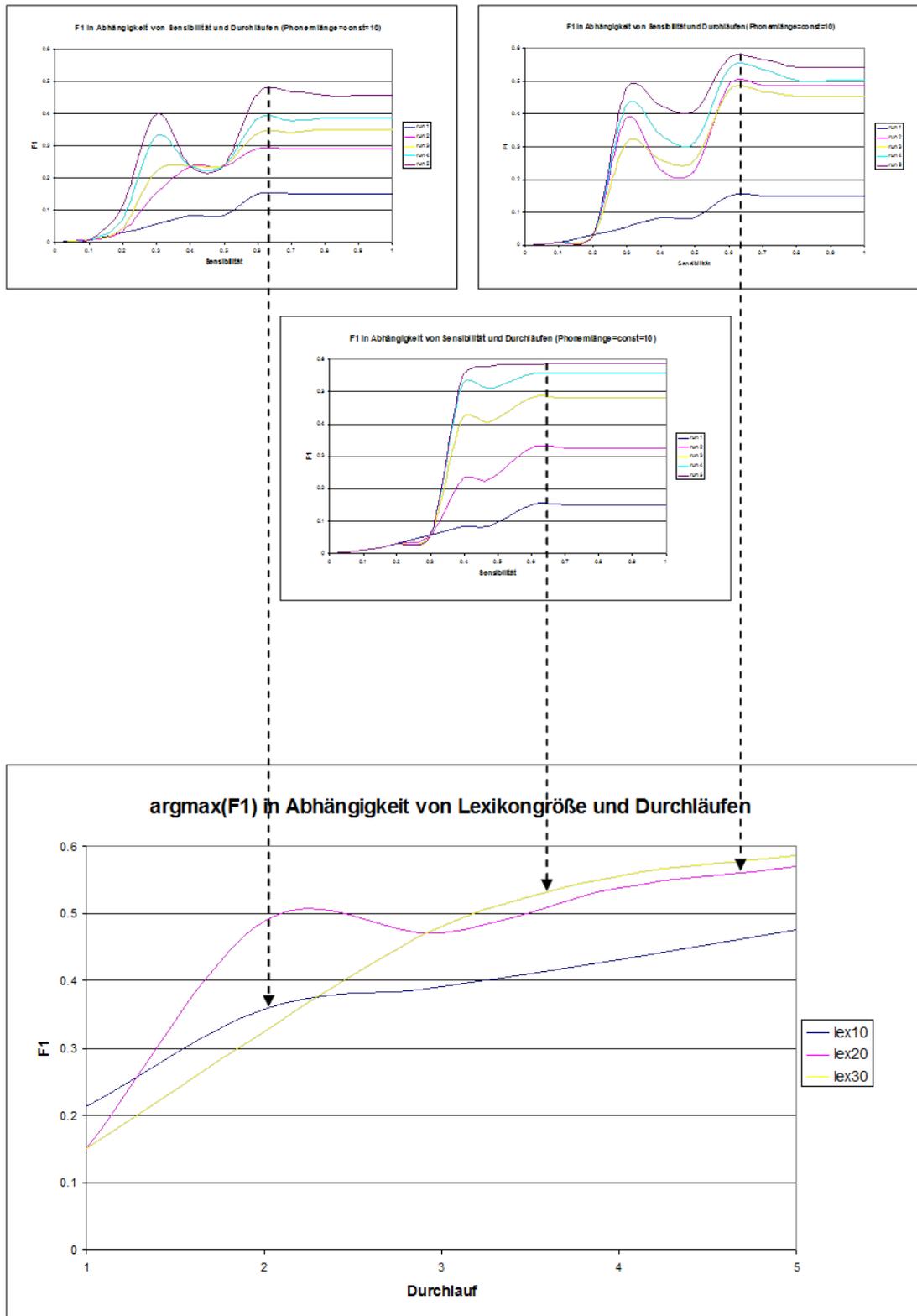


Abbildung 6.10: F_1 -Maximalwerte, $\text{argmax}((F_1)(\omega))$; $10 \leq \zeta \leq 30$, $\lambda, \sigma = \text{const}$

6 Simulationsergebnisse und Auswertung

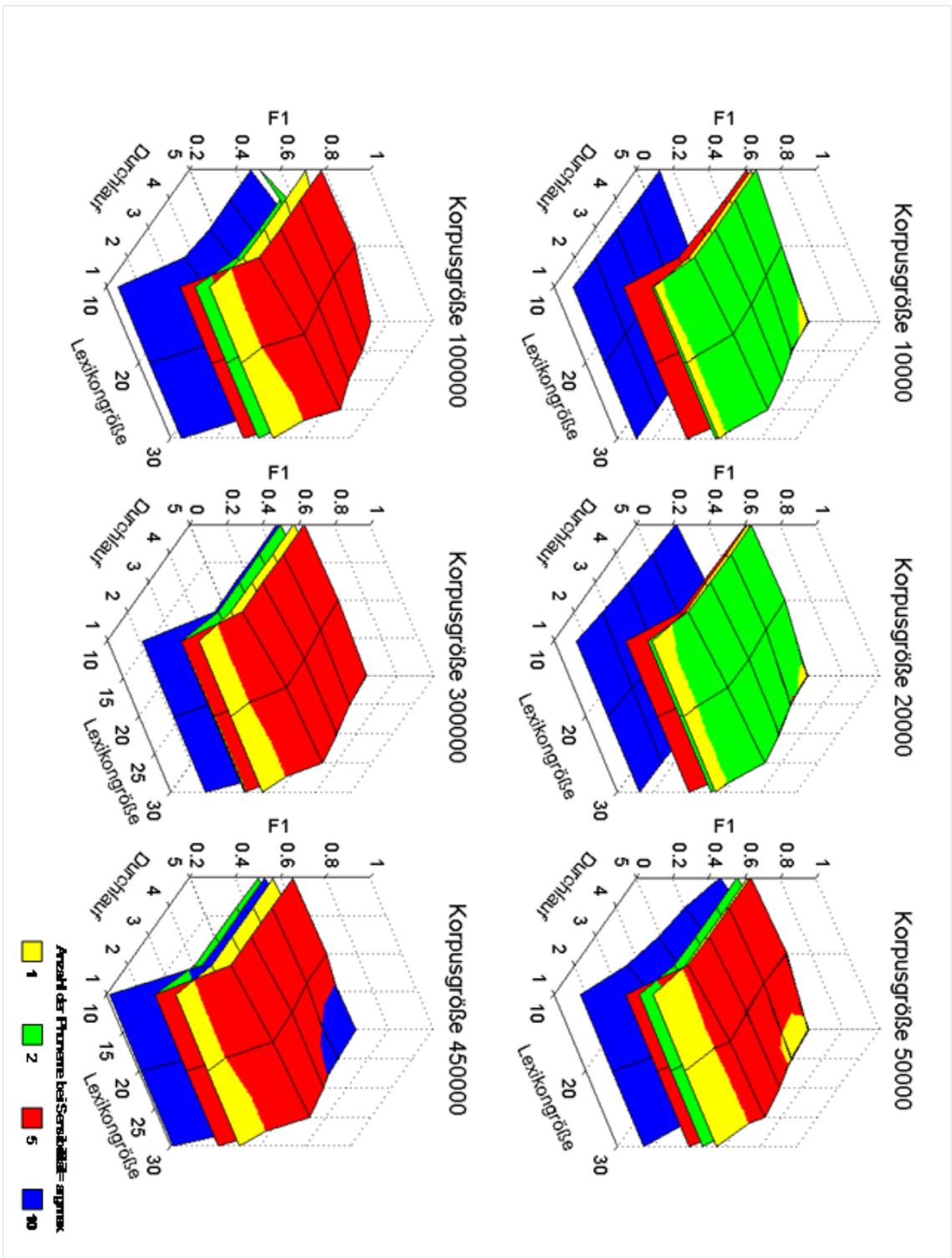


Abbildung 6.11: Lexikonentwicklung für große Korpora

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

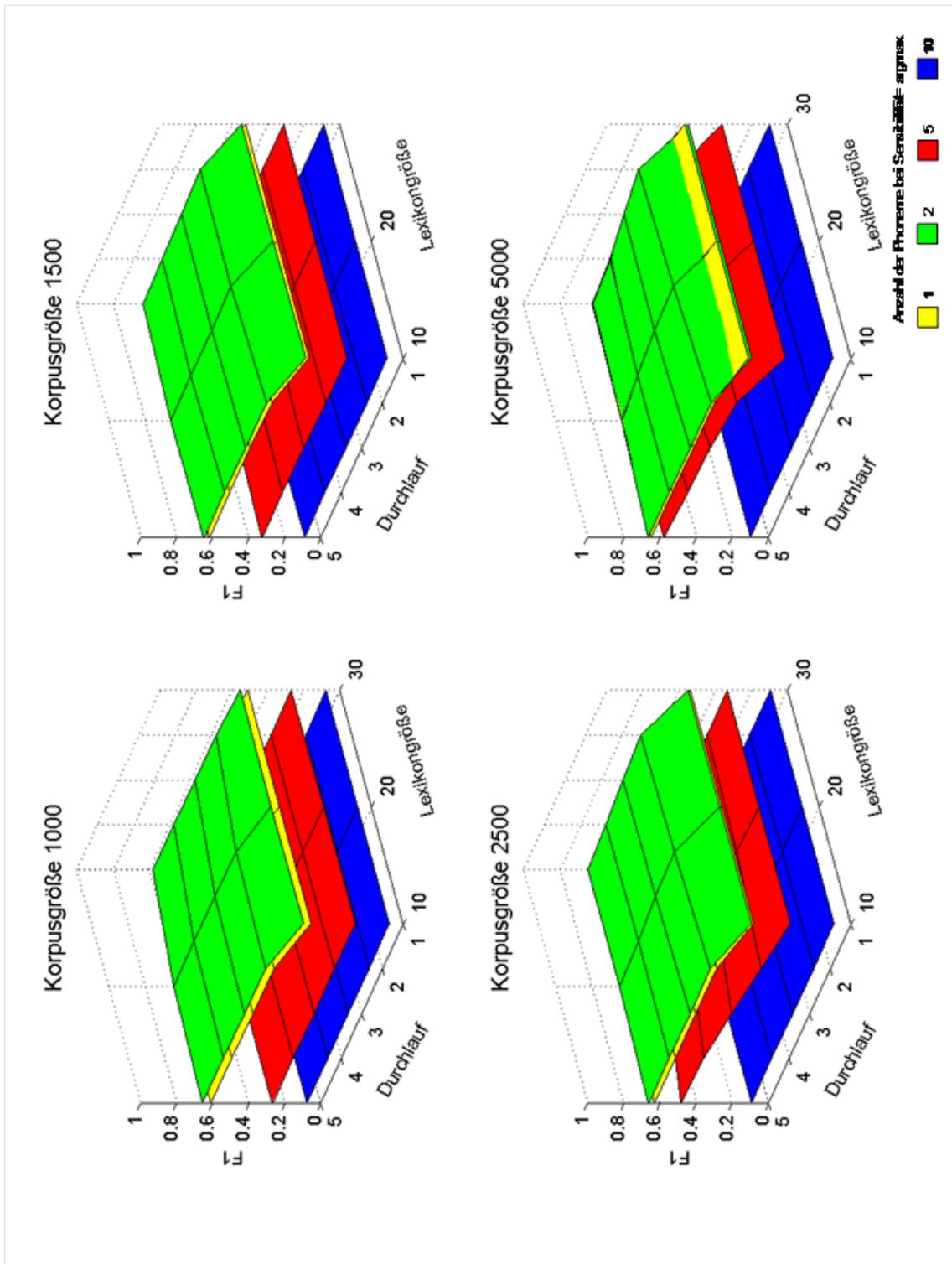


Abbildung 6.12: Lexikontwicklung für kleine Korpora

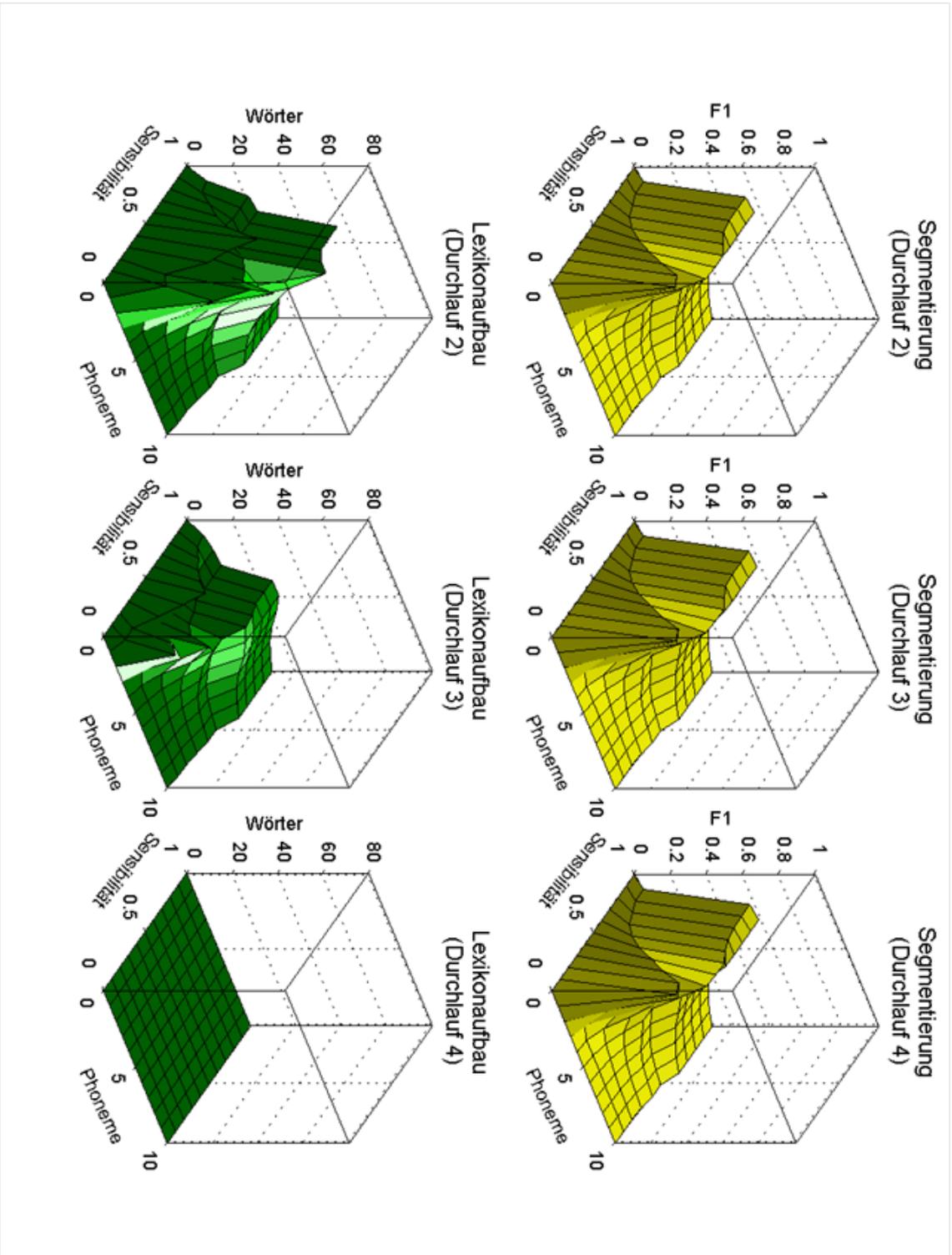


Abbildung 6.13: $\psi = 1000$, $\zeta = 10$ (Veränderung der Wortanzahl)

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

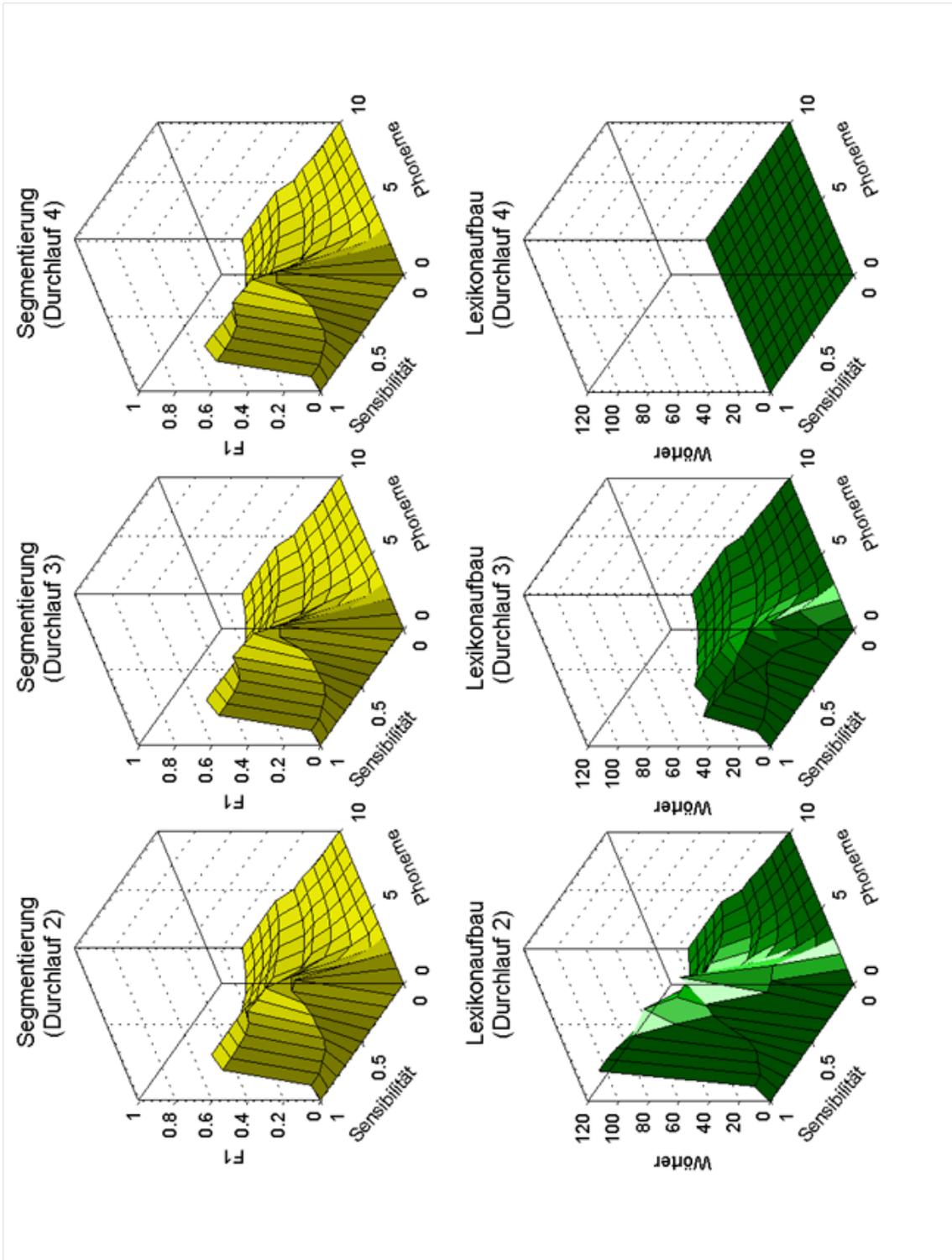


Abbildung 6.14: $\psi = 1000, \zeta = 20$ (Veränderung der Wortanzahl)

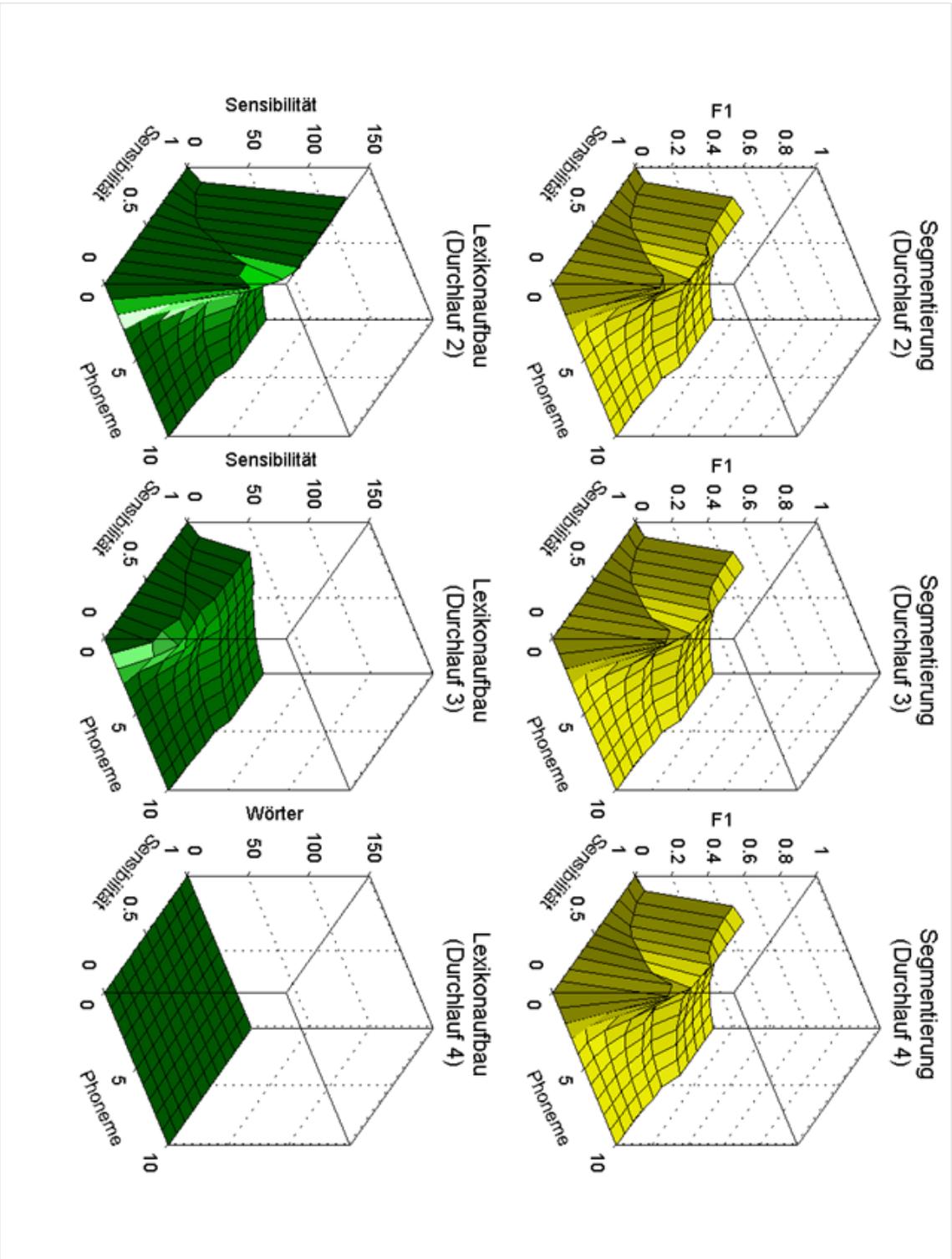


Abbildung 6.15: $\psi = 1000$, $\zeta = 30$ (Veränderung der Wortanzahl)

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

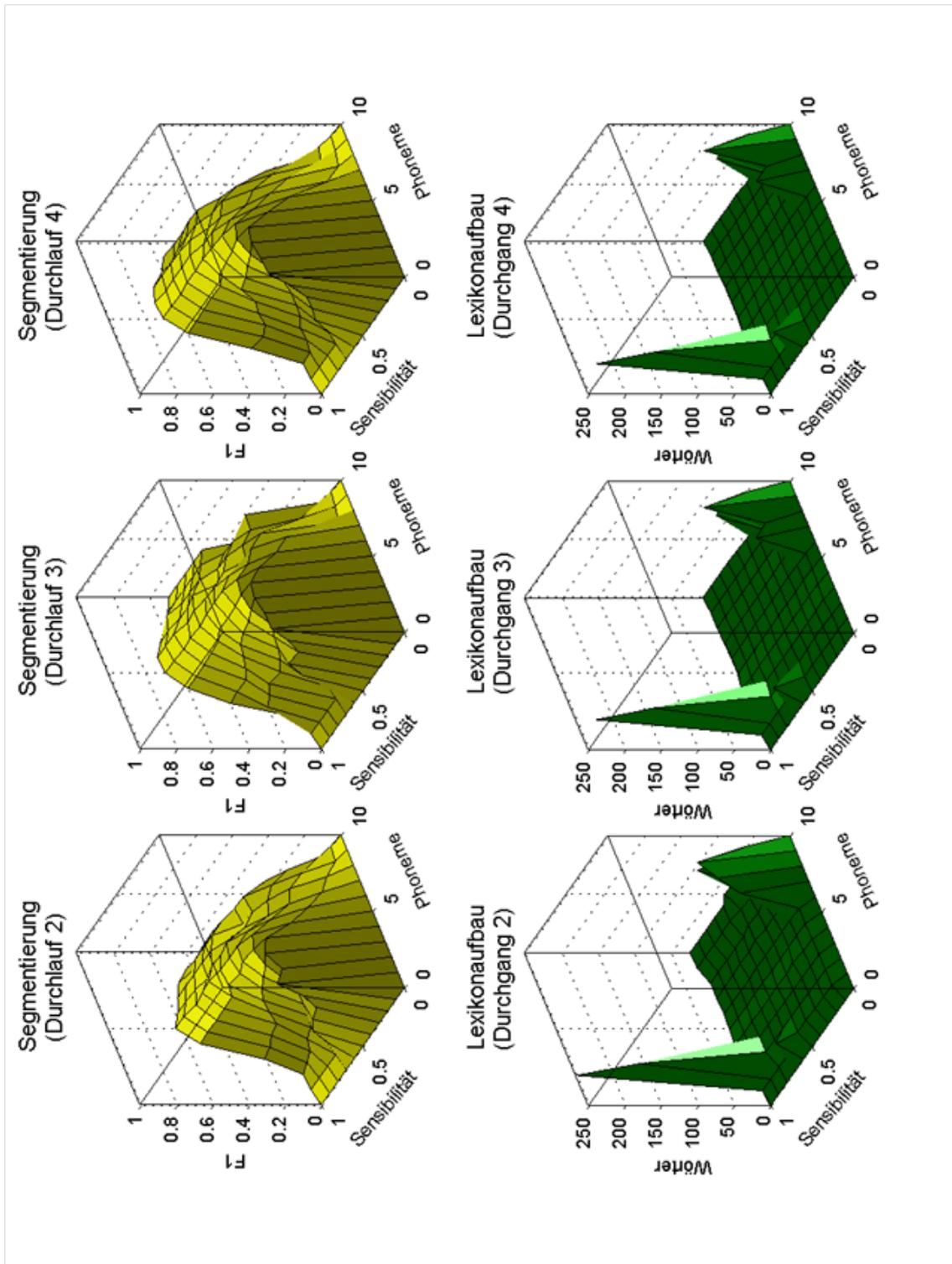


Abbildung 6.16: $\psi = 50000, \zeta = 10$

6 Simulationsergebnisse und Auswertung

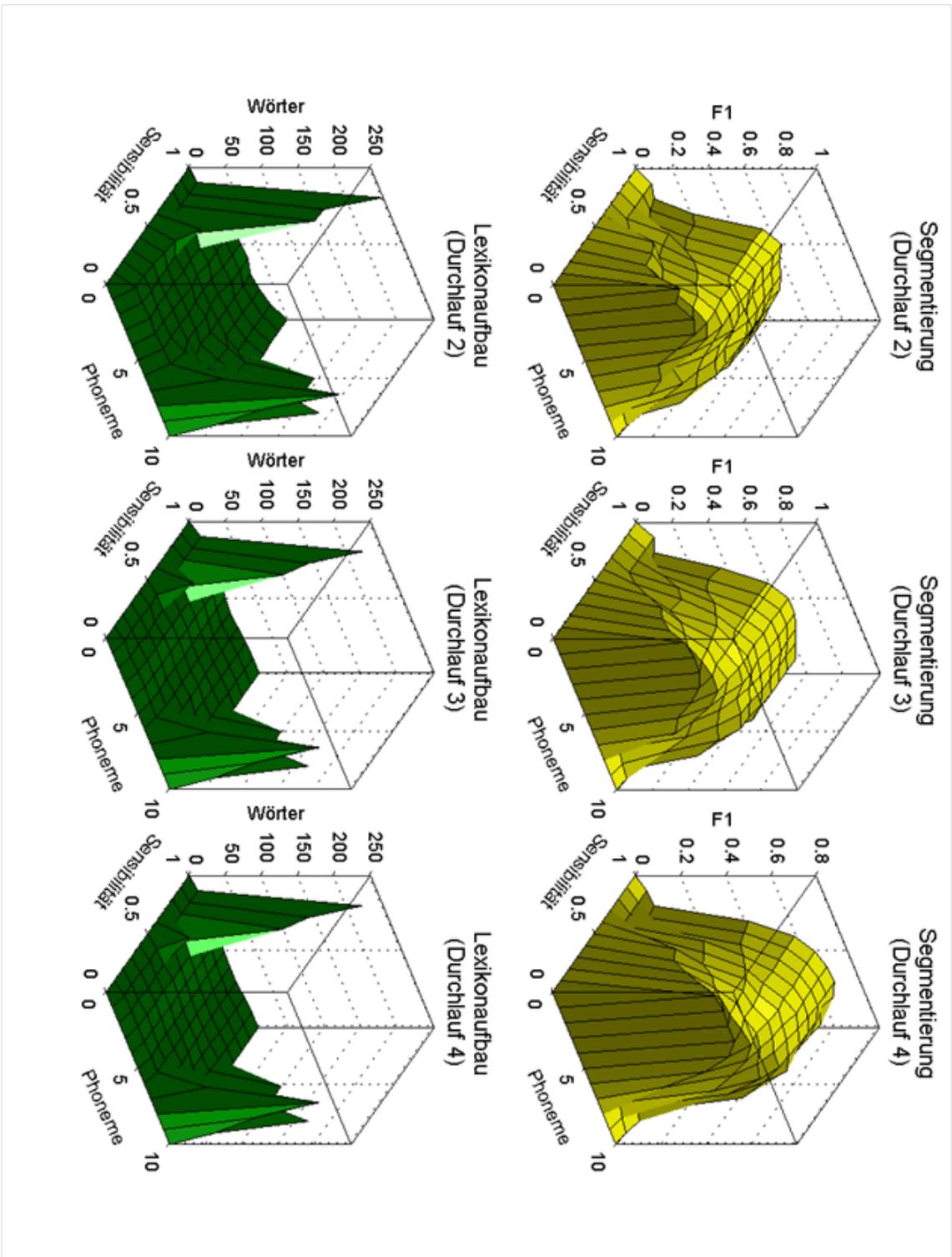


Abbildung 6.17: $\psi = 50000$, $\zeta = 20$

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

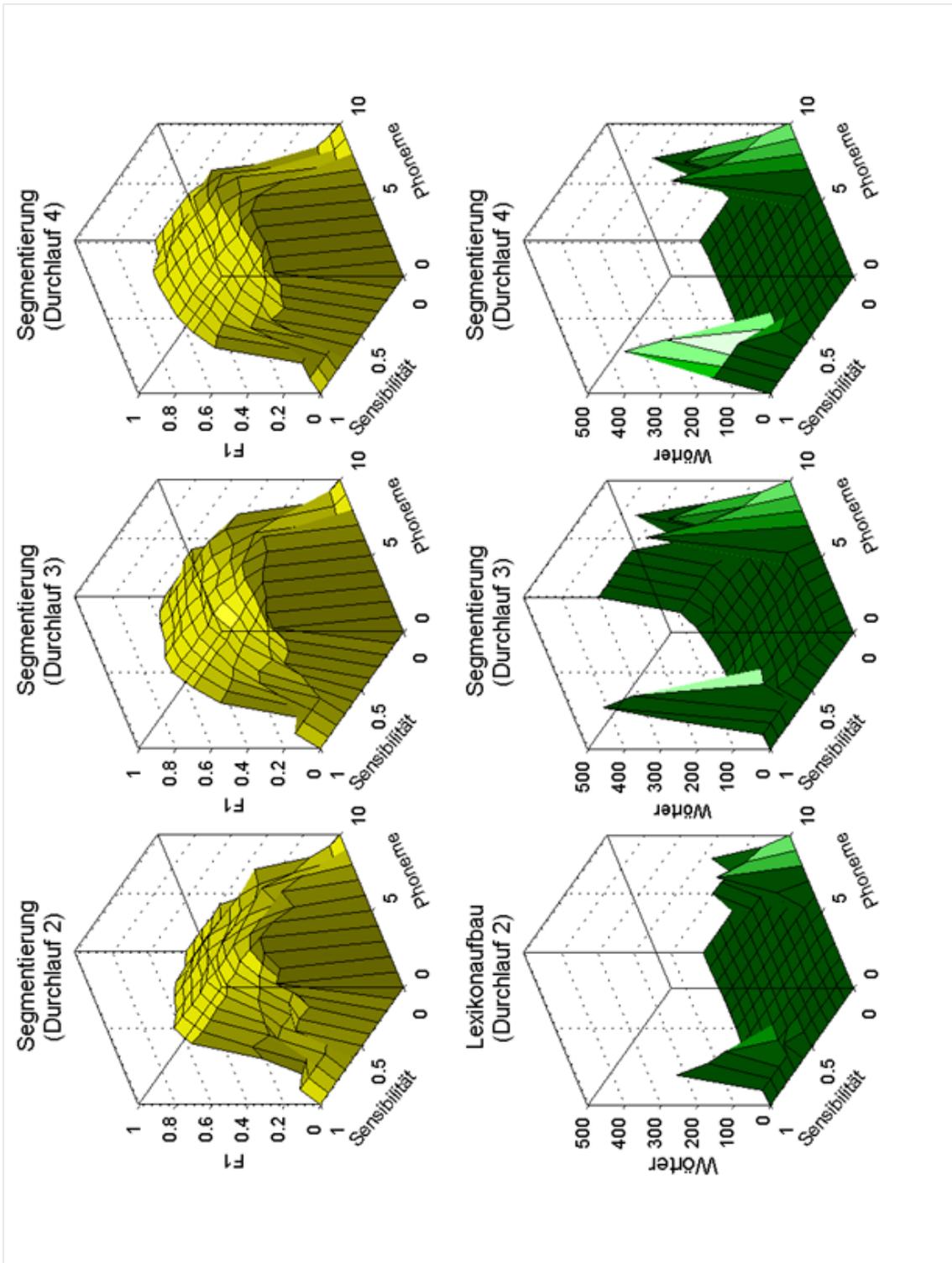


Abbildung 6.18: $\psi = 50000, \zeta = 30$

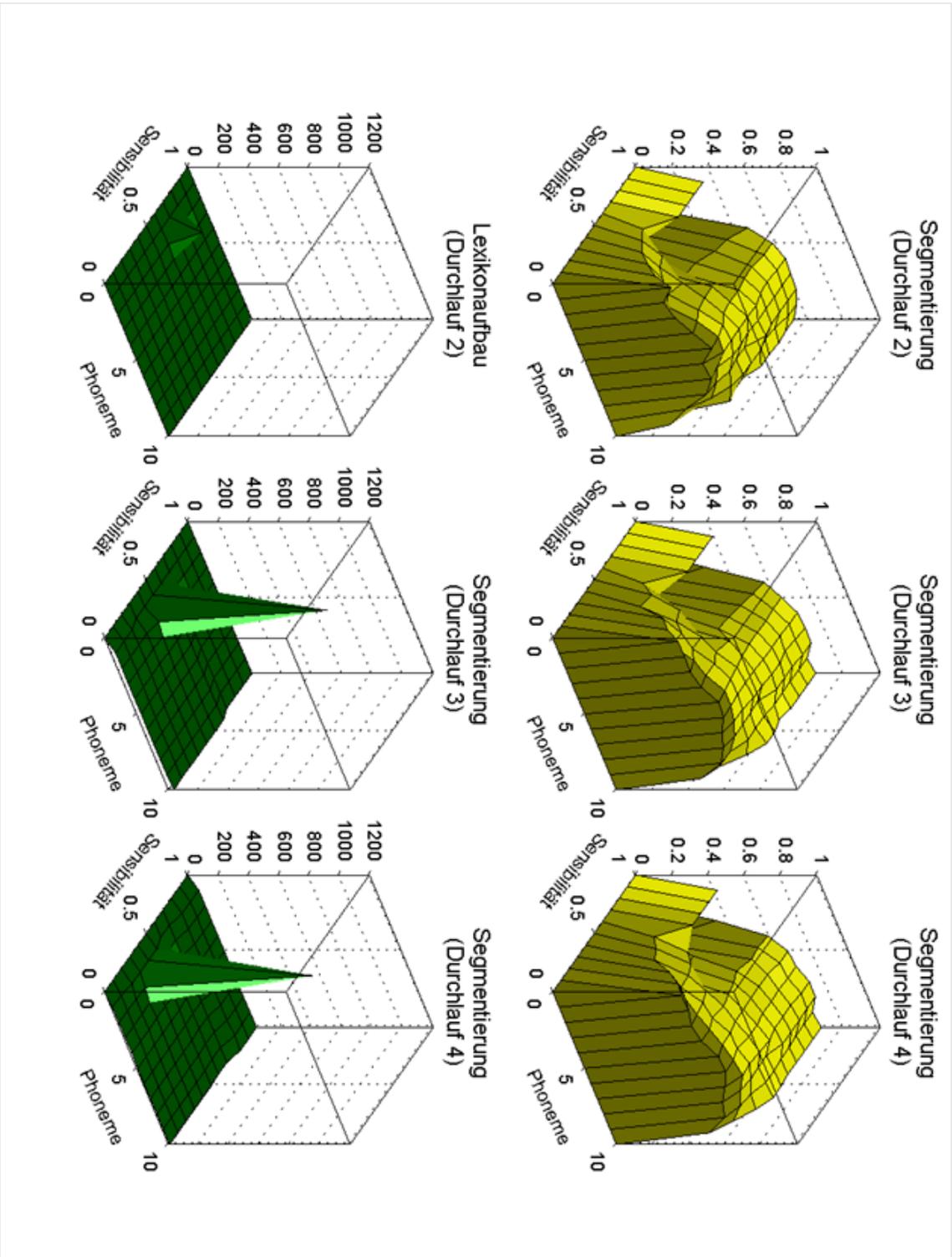


Abbildung 6.19: $\psi = 450000, \zeta = 20$

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

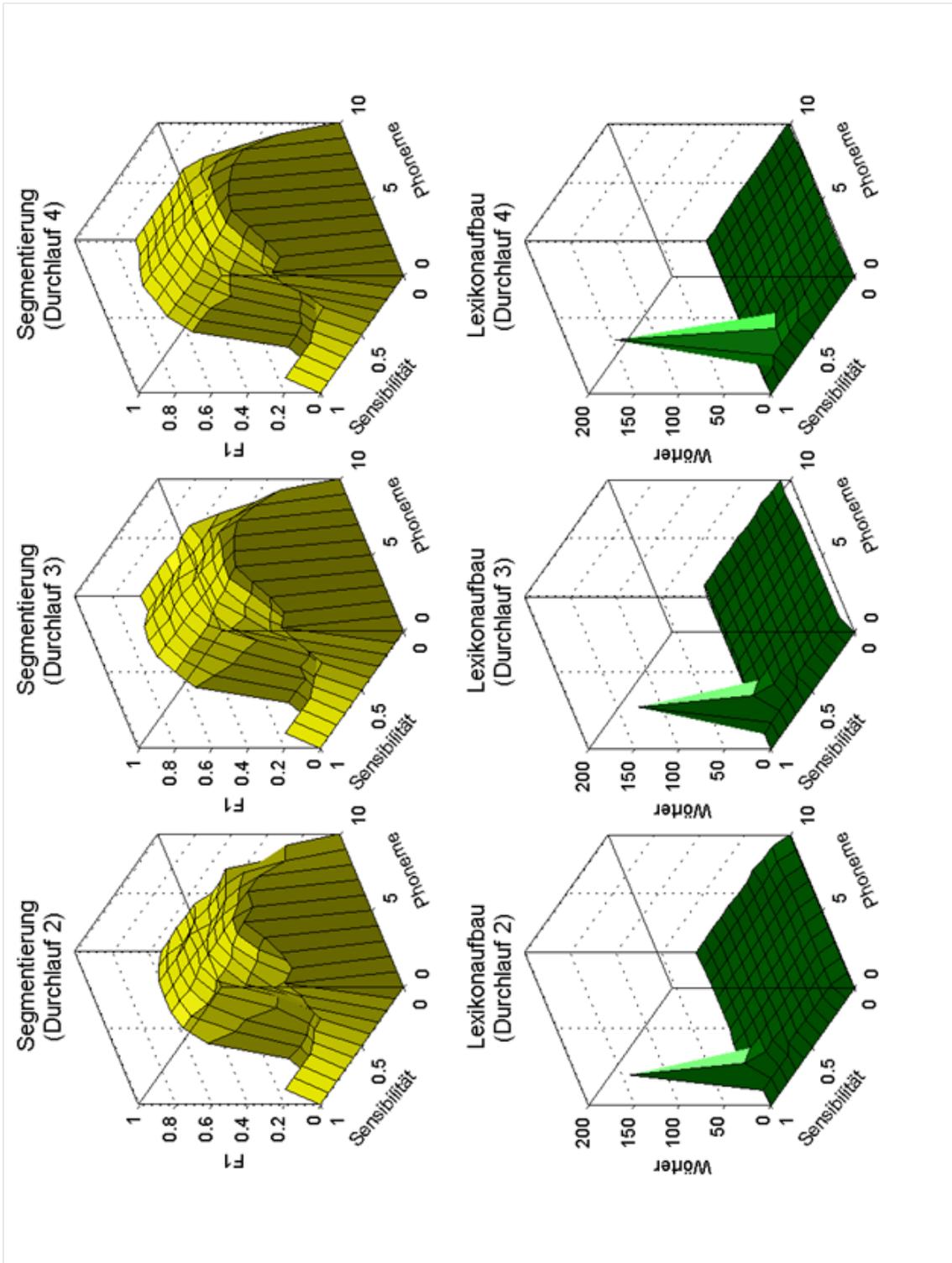


Abbildung 6.20: $\psi = 450000, \zeta = 10$

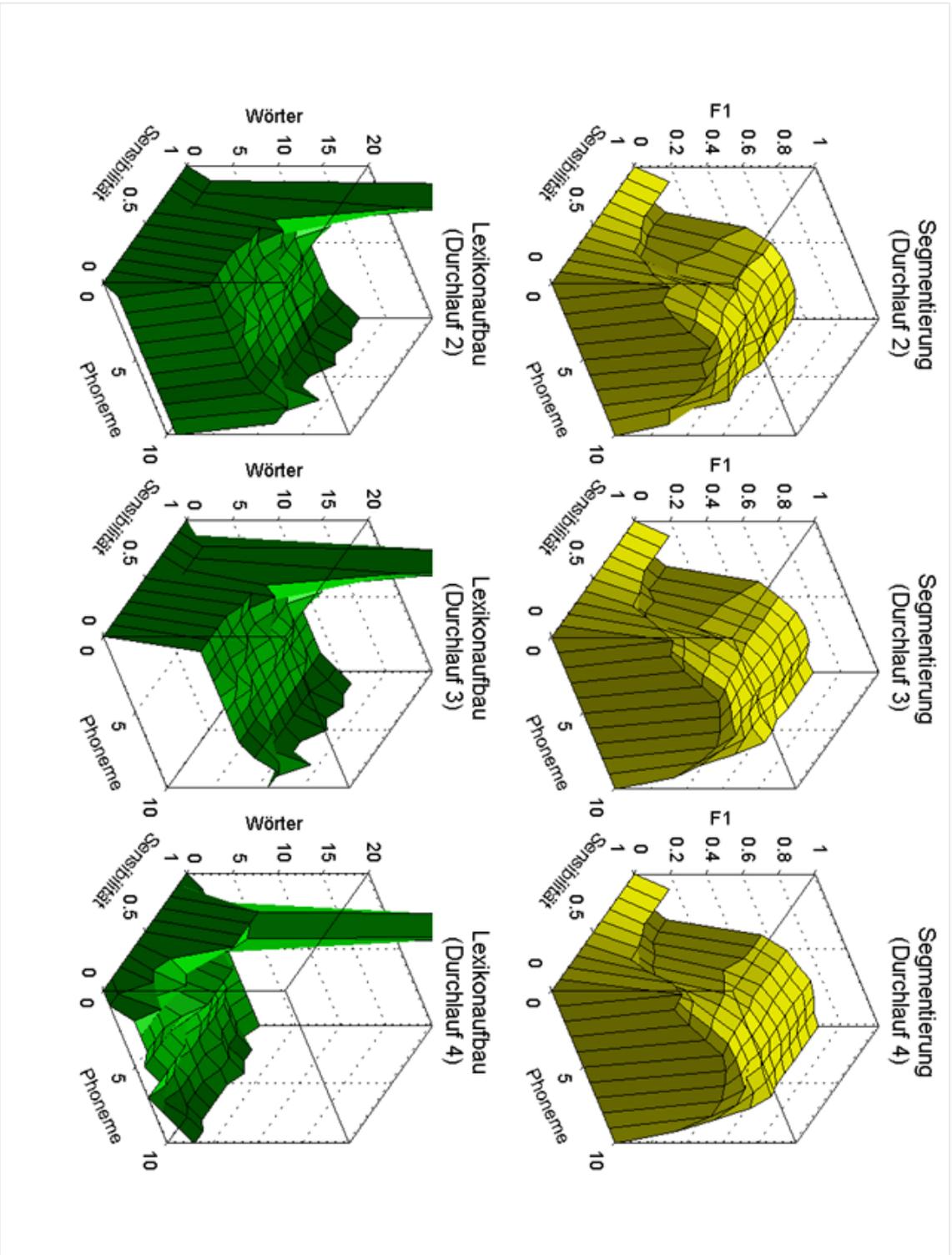


Abbildung 6.21: $\psi = 450000$, $\zeta = 10$ (ohne Ausreißer)

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

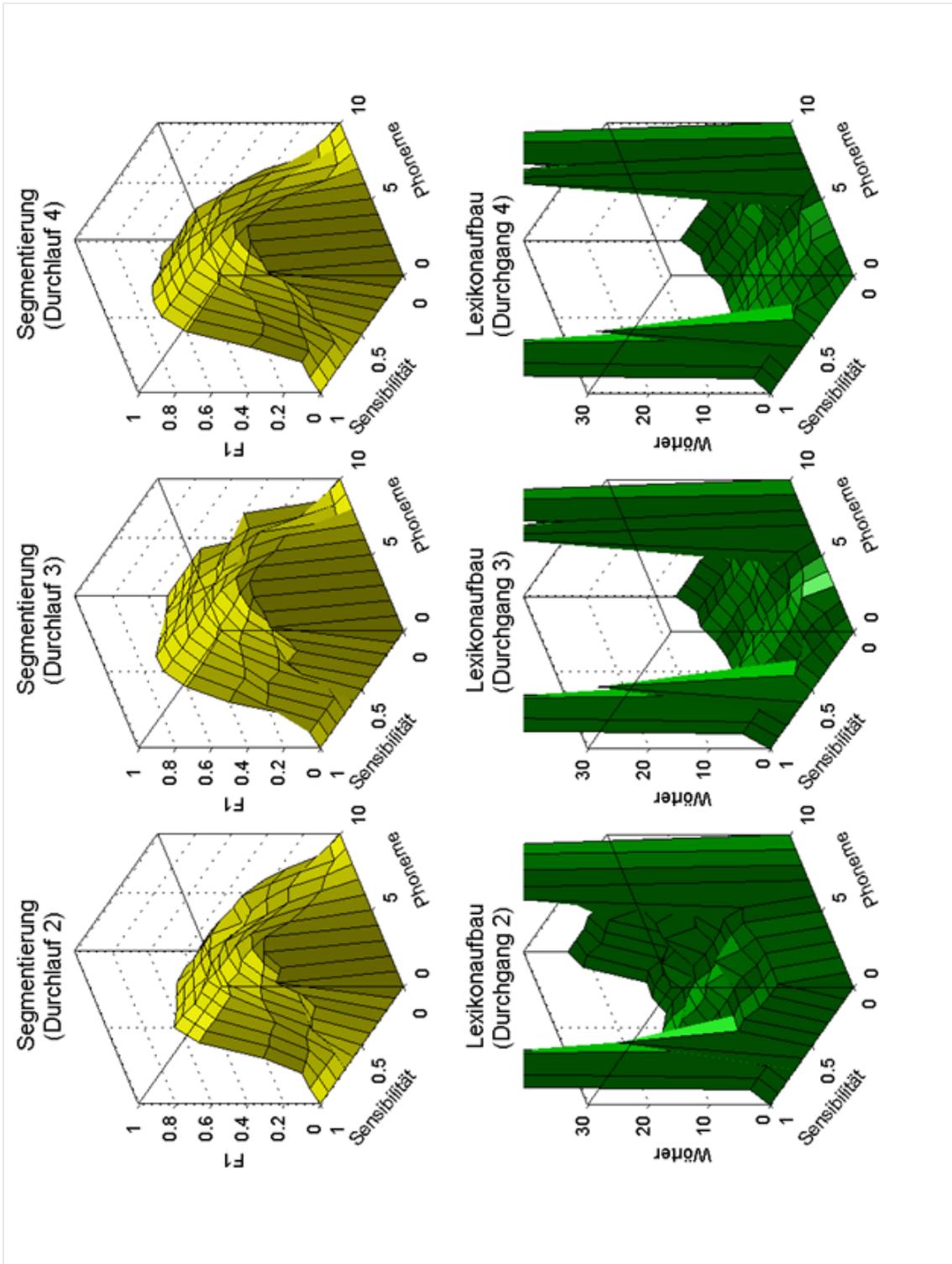


Abbildung 6.22: $\psi = 50000, \zeta = 10$ (ohne Ausreißer)

6 Simulationsergebnisse und Auswertung

Als Visualisierung sind drei Korpusrepräsentanten ausführlicher abgebildet. Sie zeigen den Verlauf des Lexikonaufbaus und die Auswirkung auf die Segmentierung in der jeweiligen Folgeperiode für drei verschiedene Lexikonfunktionen ζ_i (10, 20 und 30).⁵ In den Abbildungen 6.13-6.15 sind die Lexikonentwicklungen für einen 1000-Zeichen-Korpus zu sehen. Hierbei ist zu beachten, dass nur die neuen Wörter des Lexikons in jeder Periode berechnet wurden. Dies gilt nicht für die anderen Darstellungen. Die Abbildungen 6.16-6.18 zeigen die Lexikonentwicklung bei einem 50.000 Zeichen Korpus und letztlich stellen die Abbildungen 6.19 und 6.20 die Lexikonentwicklung für den größten Korpus mit 450.000 Zeichen dar. Für die großen Korpora sind noch die Abbildungen 6.21 und 6.22, welche die absoluten Werte der Ausreißer vernachlässigen, relevant.

6.3.2 Beschreibung der Simulationsdaten

Die bereits vorgestellte Zweiteilung in große und kleine Korpora spiegelt auch hier wieder eindeutige Verhaltensunterschiede der Segmentierungswerte wider. Zunächst seien die Abbildungen 6.13-6.15 betrachtet. Dabei fällt auf, dass die Zunahme der Wörter im Lexikon direkt proportional zur Segmentierung in der ersten Periode ist. Die Folgeperiode verzeichnet jedoch dann keinen großen Zuwachs an F_1 , sondern bleibt über alle Perioden konstant oder weist vereinzelte leichte Erhöhungen auf (zwischen 5 und 10 Prozent). Dies stimmt auch mit der Darstellung in Abbildung 6.11 überein. In der dritten Periode nimmt der Zuwachs von Wörtern in das Lexikon ab. Die äußere Gestalt des Zuwachses neuer Wörter zeigt aber weiterhin die charakteristische Form der Segmentierungsleistung. Ab der vierten Periode können keine neuen Wörter mehr segmentiert werden. Dieses Verhalten ändert sich auch nicht, wenn in jeder Periode mehr Wörter in das Lexikon gelangen könnten, d.h., die Variable ζ (Lexikonfunktion) wird von 10 auf 20 oder 30 erhöht (Abbildungen 6.14 und 6.15). Dabei werden erwartungsgemäß mehr Wörter in das Lexikon geschrieben. Die Anzahl neuer Wörter steigt in der zweiten Periode auf das doppelte. Aber dies hat nur wenig Auswirkung auf die Segmentierungsleistung in der Folgeperiode.

Große Korpora zeigen ein anderes Bild. Die Abbildungen 6.16-6.18 zeigen einen 50.000-Zeichen-Korpus. Natürlich liegen die sehr hohen Maxima der Lexikonentwicklung genau an den Extrempunkten, welche, wie in Kapitel 6.2.1 bereits herausgestellt, möglichst vermieden werden sollten, d.h. für genau die Parameterkombinationen von Phonemkettenlänge und Sensibilität, die keine richtigen Segmentierungen hervorbringen. Dieses Muster tritt in jeder Periode auf. An diesen kritischen Stellen können auch keine richtigen Segmentierungen berechnet werden. Die F_1 -Werte liegen in jeder Periode bei Null. Für die anderen Kombinationen von Phonemanzahl und Sensibilität ist eine leichte Verbesserung der F_1 -Werte zu verzeichnen. Das gilt insbesondere für höhere Sensibilitäten und längere Phonemketten. Die Segmentierung wird also robuster. Die Anzahl der zugelassenen Wörter, die im Lexikon gespeichert werden, hat auf die Form wiederum keinen Einfluss, sondern nur auf die Absolutwerte. Bei $\zeta = 30$ verdoppelt sich die Anzahl der

⁵Für den 450.000-Zeichen-Korpus wurde auf die Darstellung des Lexikons der Größe 20 verzichtet.

zum Lexikon gehörenden Einträge gegenüber $\zeta = 10$. Über die Perioden hinweg kann kein Multiplikatoreffekt aufgrund der zugelassenen Wortanzahl erkannt werden.

Wird die Größe des Korpus weiter erhöht, kehrt sich der negative Effekt der eine kritischen Stelle (hohe Phonemanzahl/niedrige Sensibilität) um (Abbildungen 6.19-6.22).⁶ Die zweite kritische Stelle bleibt erhalten und schneidet ein tiefes Tal in die Segmentierungsfunktion. Mit zunehmenden Iterationen nähert sich das Tal der mittleren Sensibilität und bleibt dann konstant. An diesen Stellen herrschen in der Wortsegmentierung quasi-chaotische Zustände, welche durch das Einfügen der Lexikoninhalte deutlich verstärkt werden. In der dritten Periode steigt so die Anzahl der Wörter, die ins Lexikon aufgenommen werden, von einigen unter Hundert auf über Tausend an. Aus der Segmentierungsfunktion (F_1 -Werte) ersieht man deutlich, dass diese Einträge in der Mehrheit keine Wörter sein können. In den Abbildungen 6.21 und 6.22 sind Lexikonentwicklungen vergrößert dargestellt, weil diese Entwicklungen in der Gesamtsicht nicht erkennbar sind.

Im Durchschnitt verbessern sich bei allen Korpusgrößen die F_1 -Werte um fünf Prozent pro Durchlauf, wobei die zweite Periode die höchste Verbesserung aufzeigt und jede weitere Periode immer weniger Verbesserungen erbringt.

6.3.3 Dateninterpretation und Erklärungen

Das Verhalten des 1000-Zeichen-Korpus könnte zwei Ursachen haben. Entweder würden sich die positiven und negativen Effekte aufheben oder die aus dem Lexikon eingefügten Wörter würden bereits durch die Übergangswahrscheinlichkeiten vollständig im ersten Durchlauf segmentiert werden. Um das herauszufinden, müssen die Komponenten des F_1 -Maßes betrachtet werden. Dies soll hier an einem Beispiel illustriert werden (siehe Tabelle 6.1).⁷ In Tabelle 6.1 werden die Daten eines 1000-Zeichen-Korpus für zwei Durchläufe und für die Phonemketten der Länge eins und zwei vergleichend gegenübergestellt. Die vier interessanten Szenarien wurden dick hervorgehoben. Aus diesen Daten ergibt sich, dass beide oben genannten Ursachen in Abhängigkeit von der Sensibilität σ und der Länge der Phonemketten λ zutreffend sind. Im ersten Durchgang werden bei einer Phonemkette und Sensibilität von eins keine Unterschiede bei richtigen und falschen Segmentierungen zu beobachten sein. Wie in der Erklärung zu Kapitel 6.2.2 bereits ausführlich dargelegt, führt die hohe Sensibilität bei nur einer Phonemkette dazu, dass Einheiten nur dann als zusammengehörig anerkannt werden, wenn sie ausnahmslos in der gleichen Umgebung vorkommen, also immer das gleiche Phonem als Nachfolger haben. Dies ist bei einem 1000-Zeichen-Korpus nur sehr selten der Fall und lässt sich aus Tabelle 6.1 direkt ablesen. Das Lexikon konnte nur sechs Einträge aufnehmen, obwohl zehn Einträge ($\zeta = 10$) zur Aufnahme erlaubt sind. Diese sechs neuen Einträge

⁶Genau genommen, wäre hier eine weitere Unterteilung der großen Korpora auch zu vertreten. Weil dies aber nur bei dem größten Korpus der Fall ist, soll darauf verzichtet werden.

⁷Dieses Beispiel steht als typischer Repräsentant für 26.400 ähnliche Berechnungen von den als klein eingestuften Korpora. Bei Bedarf einer näheren Prüfung dieser Stichprobe stehen die Daten auf einer DVD/CD-ROM-Sammlung am Fachgebiet Linguistik – Computerphilologie der Universität Kassel zur Verfügung.

6 Simulationsergebnisse und Auswertung

entsprechen genau den getroffenen Segmentierungen ($3 + 2 = 5$ Lücken = 6 Wörter). In der zweiten Periode werden nun genau auch diese Wörter wieder eingefügt und damit kommt keine neue Information hinzu. Genau diese Information wurde bereits durch die Übergangswahrscheinlichkeiten errechnet. Konsequenterweise können sich die Leistungsmaße auch nicht ändern. Erst wenn sich die Sensibilität über die Hälfte senkt, sind positive Effekte zu beobachten. Die Steigungen der Leistungsmaße liegen zwischen fünf und zehn Prozent. Dafür ist aber fast ausschließlich das Maß der Vollständigkeit verantwortlich. Die Präzision bleibt weitestgehend konstant. Damit ist klar, dass sowohl Fehler als auch richtige Segmentierungen im gleichen Verhältnis wachsen. Der positive Nettoeffekt wird also dadurch erzielt, dass nun mehr segmentiert wird, nicht aber dadurch, dass die richtigen Segmentierungen im Verhältnis zu den Fehlern wachsen.

Bei Erhöhung der Phonemkettenlänge λ auf zwei Einheiten verhalten sich die Leistungsmaße im Vergleich zu einer Phonemkette umgekehrt proportional. Das ist insbesondere für die hohen Segmentierungswerte überraschend, da man zunächst nicht annehmen würde, dass sich die Verteilung bei zwei Phonemen derart drastisch ändert. Richtige und falsche Segmentierungen wachsen im gleichen Verhältnis; heben sich also auch bei einem Phonem gegenseitig auf. Der geringe positive Effekt ist abermals auf die Vollständigkeit zurückzuführen. Allerdings ist nun die Wirkung der rekursiven Struktur schon bedeutsam. Der erste Durchlauf hat ein Lexikon mit 16 Einträgen bereitgestellt, das bei günstigster Einfügung 32 Lücken hervorbringen würde. Aufgrund der zusätzlich gegebenen Informationen werden die neu zu berechnenden Übergangswahrscheinlichkeiten auch mit dieser Information versorgt und es können noch einmal mehr Lücken entdeckt werden (drei im gegebenen Beispiel der Tabelle 6.1). Sinkt die Sensibilität nun stark ab, ist eine solche Kombination an Bigramen gegeben, bei der alle Einheiten schon richtig durch die Übergangswahrscheinlichkeiten berechnet wurden, sodass die segmentierten Bestandteile, die auch am häufigsten vorkommen, denen in Form und Zahl des Lexikons entsprechen.

Die typische Proportionalität von F_1 und neuen Wörtern im Lexikon ist bei großen Korpora auf den ersten Blick nicht mehr vorhanden (Abb. 6.19-6.22). Während bei kleinen Korpora, wie erwartet, die hohen F_1 -Werte auch in korrekten Lexikoneinträgen resultierten, zeigten die Ergebnisse, dass diese Logik bei großen Korpora nicht Stand hält. Hier stechen die Ausreißer zu deutlich hervor. Die zwei kritischen Kombinationen von Phonemkettenlänge und Sensibilität, die keine oder nur sehr wenig richtige Segmentierungen hervorbringen, verstärken diesen negativen Einfluss natürlich umso gewaltiger in den Folgeperioden. Eine Eigenschaft des F_1 -Maßes ist, nicht weiter als Null zu sinken, selbst wenn es nur falsch gesetzte Lücken gibt. Diese Situation tritt bei beiden kritischen Punkten auf. Im ersten Durchgang wurden keine richtigen Segmentierungen errechnet (siehe Kapitel 6.2.1). Folgerichtig können nun daraus auch keine korrekten Wörter in das Lexikon geschrieben werden. Hinzu kommt, dass die Varianz der Häufigkeitsverteilung der getätigten Segmentierungen äußerst gering ist, d.h. sehr viele verschiedene Bigrame kommen nur zwei- oder dreimal vor. Laut der Definition des Algorithmus werden aber alle Segmentierungen ins Lexikon aufgenommen, die genauso oft vorkommen wie der letzte reguläre Lexikoneintrag, d.h. kleiner oder gleich als die zulässige Lexikongröße

Durchlauf 1, Lexikon 10						Durchlauf 2, Lexikon 10							
Länge 1	richtig	falsch	nicht	precision	recall	F_1	Lexikon	richtig	falsch	nicht	precision	recall	F_1
1.0	3	2	249	0.60	0.01	0.02	6	3	2	249	0.60	0.01	0.02
0.9	3	2	249	0.60	0.01	0.02	6	3	2	249	0.60	0.01	0.02
0.8	3	2	249	0.60	0.01	0.02	6	3	2	249	0.60	0.01	0.02
0.7	3	2	249	0.60	0.01	0.02	6	3	2	249	0.60	0.01	0.02
0.6	20	2	232	0.91	0.08	0.15	23	20	2	232	0.91	0.08	0.15
0.5	46	27	206	0.63	0.18	0.28	63	46	27	206	0.63	0.18	0.28
0.4	62	36	190	0.63	0.25	0.35	11	71	44	181	0.62	0.28	0.39
0.3	72	62	180	0.54	0.29	0.37	22	92	61	160	0.60	0.37	0.45
0.2	99	77	153	0.56	0.39	0.46	27	135	90	117	0.60	0.54	0.57
0.1	108	62	144	0.64	0.43	0.51	26	127	78	125	0.62	0.50	0.56
0.0	0	0	252				1	0	0	252			
Länge 2													
1.0	90	39	162	0.70	0.36	0.47	16	114	46	138	0.71	0.45	0.55
0.9	90	39	162	0.70	0.36	0.47	16	114	46	138	0.71	0.45	0.55
0.8	94	41	158	0.70	0.37	0.49	17	125	51	127	0.71	0.50	0.58
0.7	97	38	155	0.72	0.38	0.50	16	129	48	123	0.73	0.51	0.60
0.6	111	45	141	0.71	0.44	0.54	21	138	69	114	0.67	0.55	0.60
0.5	124	42	128	0.75	0.49	0.59	13	158	57	94	0.73	0.63	0.68
0.4	114	41	138	0.74	0.45	0.56	25	151	73	101	0.67	0.60	0.63
0.3	112	34	140	0.77	0.44	0.56	21	146	51	106	0.74	0.58	0.65
0.2	75	20	177	0.79	0.30	0.43	86	75	20	177	0.79	0.30	0.43
0.1	24	18	228	0.57	0.10	0.16	43	24	18	228	0.57	0.10	0.16
0.0	0	0	252				1	0	0	252			

Tabelle 6.1: Vergleich der Leistungsmaße von zwei Durchläufen

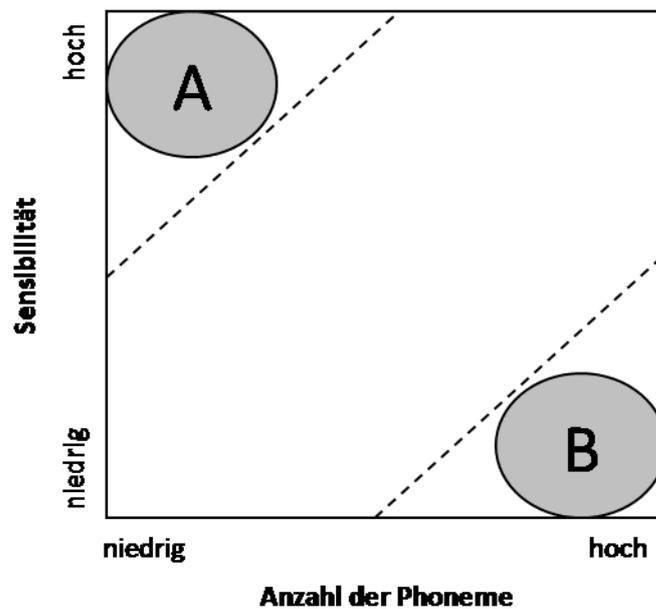


Abbildung 6.23: Widersprüchliche Parameterkombinationen

sind. Diese große Anzahl an Lexikoneinträgen, welche in der Folgeperiode zu weiteren Fehlsegmentierungen führen muss, würde für das Kind einen quasi-chaotischen Zustand bedeuten, weil es, wie sich dann erst später herausstellen würde, auf keine gewohnte Struktur zurückgreifen könnte, obwohl aus algorithmischer Sicht natürlich eine Struktur vorhanden ist. Selbst wenn man den Algorithmus so abändert, dass nur genau die spezifizierte Lexikongröße abgespeichert wird, würde es nichts an dem Ergebnis ändern: An den beiden kritischen Stellen würden keine richtigen Segmentierungen errechnet werden. Die beibehaltene Gestalt des Algorithmus hebt diesen Umstand noch einmal deutlich hervor.

Bei entsprechender Vergrößerung dieser Bereiche (Abb. 6.21 und 6.22), können die Verteilungen der Wortaufnahme ins Lexikon ohne die Ausreißer besser interpretiert werden. In beiden Korpusgrößen korrelieren die Anstiege der F_1 -Werte mit den Anstiegen der Wörter im Lexikon bei langen Phonemketten. Bei kürzeren Phonemketten sind die Anstiege ebenfalls proportional und liegen im Durchschnitt bei etwas über 10 Prozent. Hier sind also eindeutig positive Effekte der rekursiven Struktur nachweisbar. Lässt man die Ausreißer außer Acht, gibt auch hier die Form des Lexikonzuwachses die proportionale Steigung der F_1 -Werte wieder, so wie es auch bei kleineren Korpora der Fall ist. Auch hier heben sich die positiven und negativen Effekte teilweise auf, im Nettoeffekt tritt eine Besserung durch die rekursiven Strukturen ein.

Wie können nun aber die Ausreißer erklärt werden? Die betreffenden Kombinationen aus Sensibilität σ und Phonemkettenlänge λ (in Abb. 6.23 als A und B gekennzeichnet) sind aus logischer Sicht widersprüchlich und jede dieser kritischen Kombinationen

ist entweder nur für sehr kleine (A) oder nur für sehr große Korpora (B) überhaupt sinnvoll. Da die Quantität des Sprachinputs von Kleinkindern als nicht gering eingeschätzt werden kann, kommen beide Kombinationsmöglichkeiten nicht in Betracht. Die Kombination A wurde bereits in Kapitel 6.2 ausführlich erläutert und braucht hier nicht eingehender wiederholt zu werden. Hier sei nur kurz auf den logischen Widerspruch hingewiesen, welcher bei kleineren Korpora nicht existent ist. Ausgangspunkt dafür ist die Erkenntnis aus der quantitativen Linguistik, dass ein Zusammenhang zwischen Wortlänge und Vorkommenshäufigkeit besteht (Best, 2003, 25ff; 45). Einheiten, bestehend aus ein oder zwei Phonemen, kommen natürlich auch sehr häufig vor, d.h. vor allem auch in unterschiedlichen Umgebungen. Deshalb können bei hoher Sensibilität und geringer Länge der Phonemkette nur sehr wenige Segmentierungen gefunden werden. Dies gilt aber nicht bei kleinen Korpora, weil sie nicht groß genug sind, um den Effekt zwischen Silben-, Phonem- oder Wortlänge und Vorkommenshäufigkeit zu repräsentieren. Genau dies hat die Kombination A aus Abbildung 6.23 mit B gemein. Auch bei B können nur wenige Segmentierungen vorgenommen werden. Je länger eine Phonemkette wird, umso größer wird die Wahrscheinlichkeit ihrer Einzigartigkeit,⁸ d.h. ihre Vorkommenshäufigkeit wird geringer. Niedrige Sensibilitäten gehen aber genau vom Gegenteil aus, da das Verhältnis von der gesamten Phonemkette zu ihrer Teilkette, klein sein muss. Wenn die Teilkette aber schon aufgrund ihrer Länge nicht öfter vorkommt als die Gesamtkette, dann bleibt die Verhältniszahl der Sensibilität groß und es werden nur wenige Lücken eingefügt. Bei sehr großen Korpora steigt nun aber auch wieder die Wahrscheinlichkeit, dass auch lange Phonemketten öfter vorkommen können. So hebt sich dieser Effekt bei sehr großen Korpora auf.

Nun bleibt die Frage offen, warum dann die Lexikoneinträge zahlenmäßig so hoch sind. Wie bereits am Anfang der Diskussion angemerkt wurde, treten fast alle Segmentierungen dieser kritischen Regionen nur ein- oder zweimal auf. Die Architektur des Algorithmus erlaubt, alle Segmentierungen mit gleicher Häufigkeit in das Lexikon aufzunehmen. Folglich werden für diese beiden kritischen Bereiche auch hier alle segmentierten Einheiten aufgenommen, da sie alle nur einmal vorkommen. Teilweise umfassen diese Ketten einige Hundert Phoneme.

Schließlich bleiben die in den Abbildungen 6.11 und 6.12 gezeigten Zusammenhänge eine Erklärung schuldig. Bei diesen Darstellungen sind im Wesentlichen zwei Beobachtungen interessant. Erstens: Warum schwächt sich der steile Anstieg von F_1 nach der zweiten Iteration stark ab? Eine Erklärung dafür ist auch für das nächste Kapitel 6.4 unmittelbar relevant. Zweitens: Warum profitieren längere Phonemketten von der rekursiven Struktur so sehr viel mehr als kürzere?

Die erstgenannte Beobachtung hat ihre Ursache in der *Embeddedness* der Morpheme. Diese Problematik wurde schon in Kapitel 6.2.1 am Beispiel der Fehlsegmentierungen hoher Sensibilität erläutert. Der hier auftretende Effekt kann aber genauer benannt werden. Im Englischen entsprechen einige der grammatischen Morpheme mit hoher Frequenz einfachen Phonemen oder einer Zweierphonemkette. Zum Beispiel treten das Pluralmor-

⁸Dies geht auch aus den Statistiken der quantitativen Linguistik in Best (2003) hervor.

6 Simulationsergebnisse und Auswertung

phem (mit seinen Allomorphen [s], [z] und [ɪz]) sowie das Vergangenheitsmorphem für regelmäßige Verben (mit seinen Ausprägungen [d], [t] und [ɪd]) in einem repräsentativen Korpus im Vergleich zu anderen Phonemketten überproportional häufig auf. Auch die dritte Person Einzahl und das Possessivum, beide als /s/ markiert, sind hochfrequent. Gleiches gilt für die Verlaufsform und das Gerundium, welche mit /ɪŋ/ transkribiert werden. Insbesondere stehen so /s/, /t/ und /ɪŋ/ als sehr häufig vorkommende grammatische Marker zur Diskussion, denn sie treten einige Male in Kombination mit sehr häufigen Wörtern auf.

Ein Beispiel soll diesen Sachverhalt illustrieren. Das Paralexikon umfasst anfangs beispielsweise Wörter wie /ðæt/, /wʌt/, /mɒmi/, /beɪbi/, /bɔl/, /du/, /gəʊ/ oder /lʊk/. Sie treten sehr oft in einer günstigen Umgebung auf, sodass sie als Einheiten mittels der Übergangswahrscheinlichkeiten erkannt werden können. Es reichen nun nur wenige Einzelfälle aus, die sie mit den entsprechenden gebundenen Morphemen gemeinsam haben und so werden die Einheiten im Zuge der rekursiven Einfügung voneinander getrennt. Das /s/ als verkürzte Form in /ðəts/, /wʌts/ und als dritte Person in /lʊks/; das /z/ als Possessivsuffix in /mɒmiz/ und Pluralsuffix in /beɪbiz/ oder /bɔlz/; das /ɪŋ/⁹ in /dʌɪŋ/ oder /lʊkɪŋ/ sind sehr wahrscheinliche Kandidaten für das Lexikon im zweiten Durchgang. Im gewählten Beispiel würde das /s/ schon einmal pro Wort als gebundenes Morphem ausreichen und es stände dreimal einzeln als Kandidat für das Paralexikon bereit. Ähnliches gilt natürlich für das /z/, aber auch für das /t/ (siehe auch Tabelle 6.2). Beginnt nun der zweite Durchlauf, werden gemäß der Einsetzungsprozedur zuerst die längsten Phonemketten (einer Länge ξ) aus dem Lexikon mit dem Input abgeglichen, gegebenenfalls markiert und Lücken eingefügt. Wenn keine Phonemkette ξ übrig ist, setzt die Prozedur mit $\xi - 1$ fort bis $\xi = 0$ ist. Ist kein Eintrag im Lexikon enthalten, bleibt der Input an diesen Stellen ohne Lexikonabgleich und wird gemäß den Übergangswahrscheinlichkeiten segmentiert. Das letzte Szenario wird ab dem zweiten Durchlauf immer unwahrscheinlicher, weil ja nun einzelne Phoneme im Lexikon enthalten sind, die sich sehr oft in längeren Phonemketten wieder finden. Das bedeutet natürlich, dass diese Ketten an der falschen Stelle auseinander gerissen werden. So könnte es sein, dass der sequentielle Abgleich schließlich bei der Phonemkette /sɪzɜz/ (scissors) keine Einträge für /sɪzɜz/, /sɪzɜ/, /sɪz/ oder /sɪ/ findet. Das /s/ ist aber im Lexikon enthalten, sodass zunächst eine Lücke nach dem /s/ eingefügt wird. Sollte nun auch /ɪz/ noch nicht im Lexikon stehen, wird als nächster Kandidat /z/ eingefügt, der zum Einen das /ɪ/, zum Anderen das /z/ voneinander trennt. Letztlich ist die gesamte Kette in einzelne Phoneme zerlegt. Wiederholt sich dieser Vorgang nur wenige Male, werden im Lexikon nun bald auch /ɪ/ oder /ɜ/ als eigenständige Einträge stehen.

Wäre das Paralexikon größer und würde es auch die Einträge mit grammatischen Morphemen umfassen, sollte dieses Problem nicht mehr auftreten. Das bedeutet, wenn das anfängliche Lexikon nicht nur /ðæt/, sondern auch /ðəts/ enthält, könnte das /s/

⁹Das /ɪŋ/-Morphem wird allerdings auch unabhängig von der rekursiven Struktur als eigenständige Form segmentiert, weil es in genau den passenden Lautumgebungen vorkommt. Das ist für die anderen gebundenen Morpheme des Englischen nicht der Fall. /ɪŋ/ nimmt also eine Sonderrolle ein. Siehe dazu auch die Anmerkungen in Kapitel 6.5.

6.3 Einfluss rekursiver Strukturen auf die Lexikongröße und F_1

auch nicht abgetrennt werden. Es gilt also, je größer das Paralexikon ist, desto geringer ist das Problem der *Embeddedness*.

Diese Einflüsse können anfangs noch von den positiven Effekten überlagert werden, sodass sie für die ersten Durchläufe nicht mit dem F_1 -Maß erfasst werden. So treten die positiven Auswirkungen dann auf, wenn das Wort */mɔmi/* zwar ins Lexikon aufgenommen wurde, weil es einige Male einer korrekten Segmentierung unterlief, in vielen anderen Umgebungen aber auseinander gerissen wurde oder Teil einer längeren Phonemkette blieb und erst durch die Rekursion richtig segmentiert wurde. Wenn */mɔmi/* keine Possessiv- oder Mehrzahlmarker aufweist, wird es durch das Einsetzen im Folgedurchgang in jeder der Umgebungen korrekt zusammengefügt beziehungsweise an der rechten Stelle aus der Phonemkette gelöst. Das Wort kann in einem Korpus gut zwanzig Mal vorkommen und vielleicht nur die Hälfte korrekt segmentiert sein, so würde sich die F_1 -Leistung im zweiten Durchlauf immer noch verdoppeln.

Die letztgenannte Beobachtung ist auf die Verteilungsdichte von langen Phonemketten zurückzuführen. Wie bereits ausgeführt wurde, gibt die quantitative Linguistik die wesentlichen Anhaltspunkte (vgl. Best, 2003). Die Verteilung einzelner Phoneme ist schon bei kleinen Korpora repräsentativ, d.h., sie stehen in einem Mengenverhältnis zueinander, das für die jeweilige Sprache typisch ist. Bei 40 bis 50 Phonemen erscheint dies auch plausibel. Die nächste Kombinationsstufe von Phonemen zu Diphonen oder Bigramen steigt exponentiell an. Für diesen Fall kann nur ein größerer Korpus eine repräsentative Verteilung aufweisen. Genauso verhält es sich mit längeren Phonemketten. Erst ein sehr großer Korpus kann den Gebrauch längerer Phonemketten in einer Sprache repräsentativ wiedergeben. Diese quantitativen Gesetzmäßigkeiten finden sich auch in den Daten der Abbildungen 6.11 und 6.12. Schon bei kleinen Textsammlungen liegen die F_1 -Werte für Phonemketten der Länge eins und zwei bei 60 Prozent. Unabhängig von der Korpusgröße werden sich diese Werte nicht mehr ändern. Das liegt daran, dass ihre Verteilung schon ab 5000 Zeichen repräsentativ verteilt ist und deshalb konstant bleibt. Für größere Phonemketten ist die repräsentative Verteilung noch nicht erreicht. Man kann in Abbildung 6.12 sehr genau sehen, wie in Abhängigkeit von der Phonemanzahl und Korpusgröße, die F_1 -Werte zunächst asymptotisch gegen die 60-Prozent-Marke streben und dann diese schließlich auch erreichen. Dies legt andererseits auch den Schluss nahe, dass die Übergangswahrscheinlichkeiten bis zu diesem Wert, einen Korpus segmentieren und dass darüber hinaus die Übergangswahrscheinlichkeiten keine zusätzlichen Informationen hergeben können, weil die Verteilung der Phonemketten, egal welcher Länge, nicht genauer im Korpus kodiert ist. Für die restlichen 40 Prozent divergieren die Verteilungsgesetze von dem tatsächlichen Sprachgebrauch. Somit müssen andere Mechanismen für die vollständige Segmentierung ergänzend hinzugezogen werden.

6.3.4 Zusammenfassung

Die weitere Fortentwicklung des Segmentierungsmechanismus geht von den bereits vorgestellten Ergebnissen aus. Bislang wurde durch die Simulation gezeigt, dass mit einem

6 Simulationsergebnisse und Auswertung

repräsentativen Korpus ein F_1 -Maß von etwa 0.5 und etwas darüber erzielt werden kann, wenn die Parameter Sensibilität und Länge günstig gewählt werden. Die Verwendung einer rekursiven Struktur, bei der die häufigsten Segmentierungen wieder in dem Korpus eingefügt werden, steigert das F_1 -Maß auf 0.6 nach der zweiten Iteration und steigt danach nur langsam oder nicht weiter an. Handelt es sich um einen sehr großen Korpus, werden sich die F_1 -Werte auch nach der zweiten Iteration verbessern bis sie das F_1 -Maß von 0.6 erreicht haben. Dies geschieht unabhängig von der Phonemkettenlänge. Die Parameter Sensibilität und Länge der Phonemkette sind durch die rekursive Struktur bis auf zwei logisch widersprüchliche Kombinationen frei wählbar, d.h. das F_1 -Maß erreicht unabhängig von der genauen Kombination dieser Parameter immer den Wert von 0.6 oder darüber.

Die wichtige Erkenntnis aus dem vorliegenden Datenmaterial ist, dass sich die im Kapitel 4.1 diskutierte Frage der kleinsten Repräsentationseinheiten für die Berechnung von Übergangswahrscheinlichkeiten als irrelevant darstellt. Da F_1 bei allen Phonemkombinationen gleicher Länge annähernd konstant bleibt, können wir zu Recht schließen, dass dies auch für Kombinationen unterschiedlicher Längen der Fall sein muss.¹⁰

6.4 Entwicklung und Eigenschaften der Lexikonfunktion

Ausgestattet mit den „Wortsegmentierungsheuristika“ aus den Kapiteln 6.2 und 6.3 können nun spezifische Fälle genauer modelliert werden, deren Ergebnisse wieder auf Korpora rück geschlossen werden können, die das gleiche Verhalten zeigen. Die Überlegung ist denkbar einfach. Zuerst wird gezeigt, dass eine bestimmte Anzahl an Korpora das gleiche Verhalten für verschiedene Parameterausprägungen aufweist (siehe Kap. 6.2.1 und 6.3.2). Nun wird ein Korpus aus dieser Gruppe herausgenommen und einer genaueren Analyse unterzogen. Die so erbrachten Ergebnisse an dem Einzelkorpus können dann wieder auf die gesamte Gruppe übertragen werden, da die Wahrscheinlichkeit hoch ist, dass sich die anderen Mitglieder der Gruppe gleich verhalten. Korpora der Größe von ungefähr 2500 Wörtern, die einem Mutter-Kind-Dialog von ein bis zwei Stunden entsprechen, repräsentieren nach den Ergebnissen der oben erfolgten Analyse alle Korpora, die zwischen 5000 und 50.000 Zeichen liegen.¹¹ Größere Korpora können leider nicht verwendet werden, weil kein zuverlässiges Datenmaterial in diesem Umfang zur Verfügung steht.

Wie in der Zusammenfassung von Kapitel 6.3.4 angedeutet, besteht eine Möglichkeit, den Algorithmus zu verbessern darin, nicht den gleichen Korpus bei den Durchläufen zu benutzen, sondern verschiedene. Zugleich ist dies auch näher an der Realität des Kindes. Man muss wohl davon ausgehen, dass die Sprachlerner ebenfalls mit unterschiedlichem

¹⁰Dieser Schluss kann derzeit nur logisch erfolgen, weil die Komplexität von allen möglichen Kombinationen verschiedener Lautmengen zu groß ist, um sie in einer repräsentativen Simulation nachzuweisen.

Es wäre jedoch möglich, den Zufall über die genauen Parameterausprägungen entscheiden zu lassen.

¹¹Korpora ab 20.000 Zeichen zeigen für sehr lange Phonemketten ein leicht divergentes Verhalten (Abb. 6.8), was aber nicht das Gesamtbild beeinflussen kann.

Material gefüttert werden. Tabelle 6.2 zeigt die Entwicklung des Paralexikons während der ersten fünf Durchläufe, wobei in jedem Durchlauf ein anderer 10.000-Zeichen-Korpus verwendet wurde.

Die Resultate verdeutlichen zweierlei. Erstens: Es macht auf den ersten Blick im Ergebnis für die F_1 -Werte keinen Unterschied, ob verschiedene Korpora benutzt werden oder ob nur ein Korpus rekursiv hinzu gegeben wird. Dieses Ergebnis darf genau genommen auch nicht stark davon abweichen, bedenkt man den Anspruch der Repräsentativität der Korpora, welche als solche in den Kapiteln 5.3 und 6.1 deklariert wurde. Das zweite Ergebnis stellt allerdings einen Unterschied zu der Ein-Korpus-Variante heraus. Es ist auffällig, dass die Steigung von F_1 nicht im zweiten Durchlauf (siehe Abb. 6.12), sondern erst im vierten Durchlauf geschieht; dann allerdings auch mit dem gleichen Ergebnis. Den Grund für diese Verzögerung muss man schon in der Verwendung von unterschiedlichen Korpora suchen. Jeder Korpus bringt natürlich neue Wörter ein, welche nicht in dem bereits vorhandenen Lexikon vorkommen. Mit dem dritten Korpus wiederholen sich aber auch die Wörter. Dies erkennt man an dem abnehmenden Verlauf des Wörterzuwachses (dritte Zeile in Tabelle 6.2). Mit dem Zuwachs von F_1 in der vierten Periode steigt auch der Zuwachs neuer (aber falscher) Wörter wieder an. Allerdings auch hier mit einer sich verdoppelnden Fehlerquote pro Durchlauf, sodass sich hier positive und negative Effekte aufheben und F_1 für einige weitere Durchläufe konstant bleibt bevor es dann auf das Ursprungsniveau abnimmt. Im Ergebnis bleibt festzuhalten, dass die Verwendung unterschiedlicher Korpora der psychologischen Realität zwar näher kommt, dass aber davon die Segmentierungsleistung zunächst unberührt bleibt. Dies gilt auch für konstante Lexikonfunktionen.

Der zuvor beobachtete Verzögerungseffekt (siehe Tabelle 6.2) bei einer konstanten Lexikonfunktion (ζ), wie $LexFunc = 10$, gibt jedoch einen entscheidenden Wissensgewinn. Aus den Überlegungen im Kapitel 6.3.3 ist das Hauptproblem der rekursiven Struktur hervorgegangen: Das „Eingebettetsein“ von Morphemen ineinander. An dieser Stelle wurde die Lösung dazu in einem größeren Lexikon gesehen, denn wenn das Lexikon sowohl die eingebetteten als auch die längeren Morpheme enthält, kann unter den Morphemen und deren morphemischen Einbettungen ein Wettbewerb wie im Kohortenmodell initiiert werden (siehe Kap. 5.4.5 und 3.1). Dieser Wettbewerb fügt dann genau den Eintrag der eingeteten Morpheme ein, welcher am Besten an die betreffende Stelle passt. Der entscheidende Parameter ist hier die Länge des Morphems, also die Anzahl der Phoneme, aus dem sich das Morphem zusammensetzt. Die Länge wurde als Entscheidungskriterium gewählt, weil der Zusammenhang zwischen Vorkommenshäufigkeit und Wortlänge statistisch signifikant ist (Best, 2003). Je länger das Wort, umso weniger kommt das Wort in einem Korpus vor. Dieser Zusammenhang wird hier genutzt, denn im Umkehrschluss bedeutet es ja, dass die Wahrscheinlichkeit, dass es sich um ein langes Morphem handelt, sehr viel größer ist, als die, dass es sich um nur einen Teil dieses Morphems handelt oder um zwei verschiedene Morpheme, die zufällig nebeneinander auftreten. Das längere Morphem würde also besser passen und den Wettbewerb gewinnen.

6 Simulationsergebnisse und Auswertung

ω	1.	2.	3.	4.	5.					
Lexikon	ðæt	20	ðæt	114	s	405	s	694	s	975
	wæt	19	ju	111	ju	182	t	487	t	961
	jə	19	ðer	68	ðæt	164	ju	293	ɪ	400
	lɪtɹɹ	14	wæt	56	ðer	112	ðæt	229	ju	380
	ðer	13	ənd	47	wæt	92	wæt	129	ðæt	307
	ɹɹju	13	s	44	ənd	75	ənd	117	i	292
	ju	13	jə	42	get	69	ðer	112	ənd	201
	ðerjuɡəʊ	12	get	30	kən	56	ɹm	91	ɹm	187
	ənd	12	mərgɹɹ	26	ɡəʊ	52	wi	85	wæt	129
	ɹm	12	kən	20	jə	42	get	69	u	112
	bək	11	ɡəʊ	15	t	35	let	66	ðer	112
	get	11	lɪtɹɹ	14	ɹm	34	kən	56	wi	85
			wer	14	let	29	ɡəʊ	55	d	71
			ɹɹju	13	wi	28	jə	42	get	69
			ɹm	12	mərgɹɹ	26	i	40	z	68
			ðerjuɡəʊ	12	wer	14	ɪ	35	let	66
			bək	11	lɪtɹɹ	14	u	34	r	58
					ɹɹju	13	jɹ	29	kən	56
					ðerjuɡəʊ	12	mərgɹɹ	26	ɡəʊ	55
					ɹm	12	lɪtɹɹ	14	k	50
				bək	11	wer	14	ɹ	50	
						ɹɹju	13	ɹ	50	
						ðerjuɡəʊ	12	jə	42	
						ɹm	12	jɹ	29	
						bək	11	mərgɹɹ	26	
								lɪtɹɹ	14	
								wer	14	
								ɹɹju	13	
								ðerjuɡəʊ	12	
								ɹm	12	
								bək	11	
Wörter	12	17	21	25	31					
Zuwachs	12	5	4	4	6					
Fehler	0	1	2	4	9					
F_1	0.52	0.51	0.52	0.64	0.64					

Tabelle 6.2: Lexikonentwicklung für $\zeta = 10$

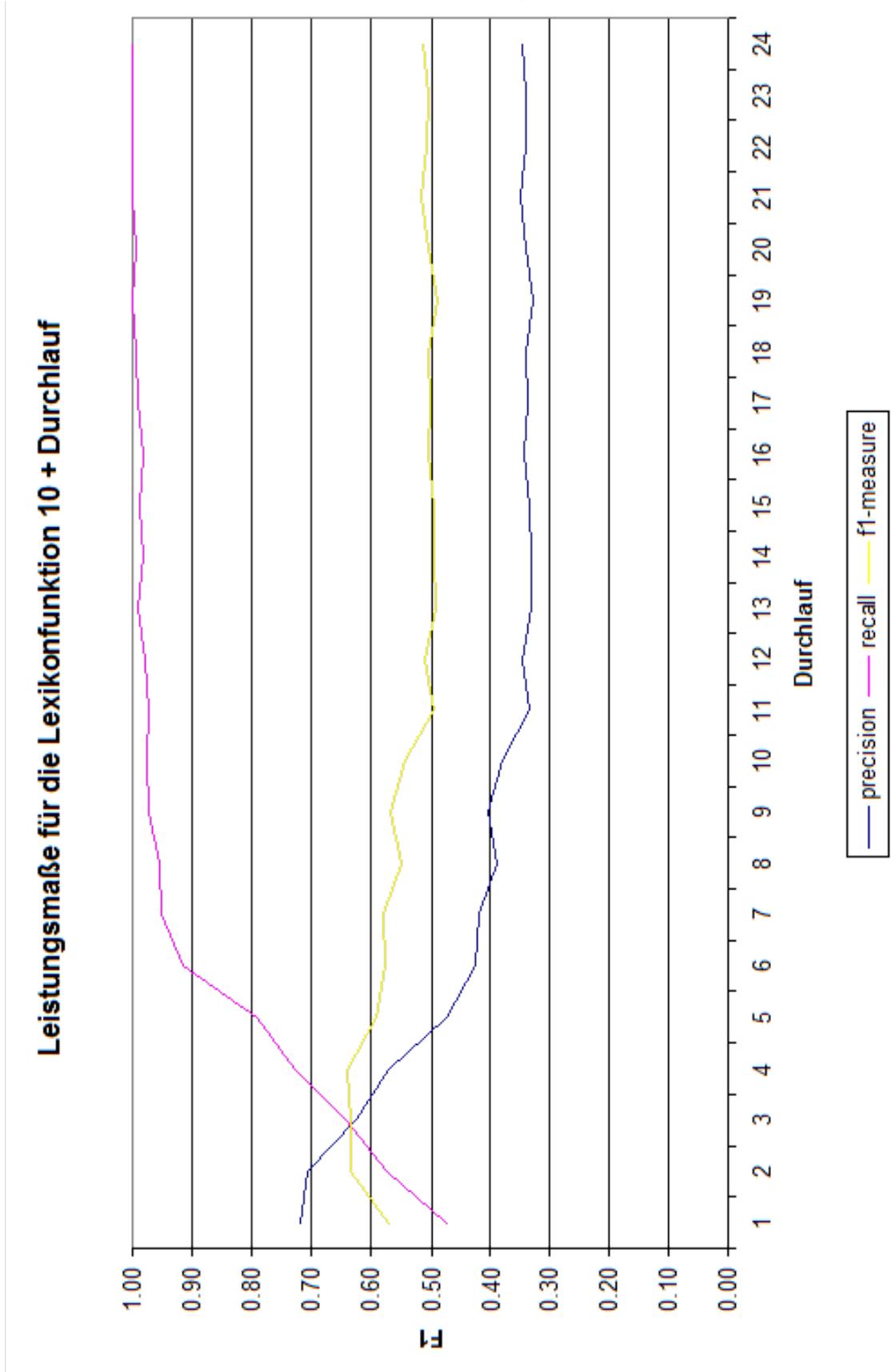


Abbildung 6.24: $F_1(\omega)$ für $\zeta = 10 + \omega$

6 Simulationsergebnisse und Auswertung

ω	richtig	falsch	nicht
1	1149	452	1280
2	1420	587	1049
3	1558	937	862
4	1798	1343	671
5	1881	2090	495
6	2248	3085	207
7	2323	3256	117
8	2276	3600	107
9	2376	3524	70
10	2404	3974	60
11	2354	4725	68
12	2496	4735	51
13	2442	4993	26
14	2394	4844	45
15	2434	4910	27
16	2476	4808	42
17	2494	4920	19
18	2546	4959	17
19	2464	5112	2
20	2534	4962	13
21	2585	4836	1
22	2545	4954	2
23	2542	4975	0
24	2615	4971	1

Tabelle 6.3: Segmentierungswerte für $F_1(\omega) = 10 + \omega$

Bei kleineren Lexikonlisten wird dieser Wettbewerb nicht entstehen können, weil es nicht genügend konkurrierende Einträge gibt, d.h. Morpheme, die ineinander enthalten sind. Ein großes Lexikon muss demnach zu weniger Fehlsegmentierungen führen.

Der entdeckte Verzögerungseffekt kann für die Entwicklung eines solchen Lexikons bedeutsam sein. Über eine gewisse Anzahl von Perioden könnte sich eine lange Liste an Wörtern ansammeln, die man erst nach dem Erreichen einer bestimmten Größe in einem *Top-down*-Verfahren als zusätzliche Segmentierungsinformation hinzugibt. Demnach wird das *Embeddedness*-Problem stark abgemildert und die Segmentierungsleistung muss konsequenterweise ansteigen. Dafür müssen zwei Größen durch weitere Simulationen geschätzt werden:

- a wie groß muss eine solche Liste von Wörtern sein (ζ) und
- b wie viele Perioden (μ) müssen vergehen,

damit sich die Segmentierungsleistung deutlich steigert beziehungsweise der Mix aus den eingebetteten Morphemen optimal wird. Beide Fragen hängen am Ende von der Größe des Paralexikons ab. ζ legt fest, wie viele Einträge in jeder Periode in das Paralexikon aufgenommen werden und die Zahl der Perioden μ besagt, wie oft ζ zum Paralexikon addiert werden soll bevor es zur Segmentierung genutzt wird.

Zur Beantwortung der ersten Frage können wir über eine konstante Anzahl von Perioden ζ unterschiedliche Lexikongrößen zuweisen, d.h., in jeder Periode werden die ζ -häufigsten *Tokens* eines mit den Übergangswahrscheinlichkeiten segmentierten Korpus akkumuliert. In Abbildung 6.25 wurden die Ergebnisse für $\zeta = 10$ und $\zeta = 30$ abgebildet. Beide Lexikonfunktionen zeigen einen konkaven Verlauf, wobei *LexFunc''*(30) erst ab 30 Perioden Null erreicht und *LexFunc''*(10) bereits ab etwa 25 Perioden. Weitere Werte für ζ zwischen 10 und 30 staffeln sich dazwischen. Erwartungsgemäß steigt die Lexikonfunktion von $\zeta = 30$ steiler an und erreicht nach 30 Perioden auch 260 korrekte Wörter. Ein Wert von $\zeta = 10$ wird dagegen die 100-Wörtergrenze wohl nicht wesentlich überschreiten können. Aus dieser Perspektive heraus würde eine Entscheidung für eine möglichst hohe Lexikongröße getroffen werden. Die falsch segmentierten Wörter betragen bei $\zeta = 30$ allerdings ungefähr 20 Prozent und dieser Betrag ist statistisch relevant, da man nicht davon ausgehen darf, dass der Sprachlerner weiß, welcher der Einträge falsch oder richtig ist. Wählt man ζ nur geringfügig höher als 30, beginnt die Fehlerquote exponentiell anzusteigen. So muss man davon ausgehen, dass ζ unter 30 Einträgen liegen sollte. Für eine Lexikonfunktion von $\zeta = 10$ spricht eine deutlich niedrige Fehlerquote im statistisch unsignifikanten Bereich.¹² Unterstellt man dem Lexikon auch rekursive Organisationsalgorithmen, wie sie weiter unten zur Diskussion gestellt werden, ist eine geringe Fehlerquote eine Grundvoraussetzung, weil sonst keine Muster erkannt werden können.

Die Werte der Simulation lassen sich näherungsweise mit Trendlinien beschreiben (Abb. 6.26). Einige Werte weichen von diesem Durchschnitt minimal ab, was aber

¹²Für die magische Zahl sieben (Miller, 1956) ist sie sogar Null.

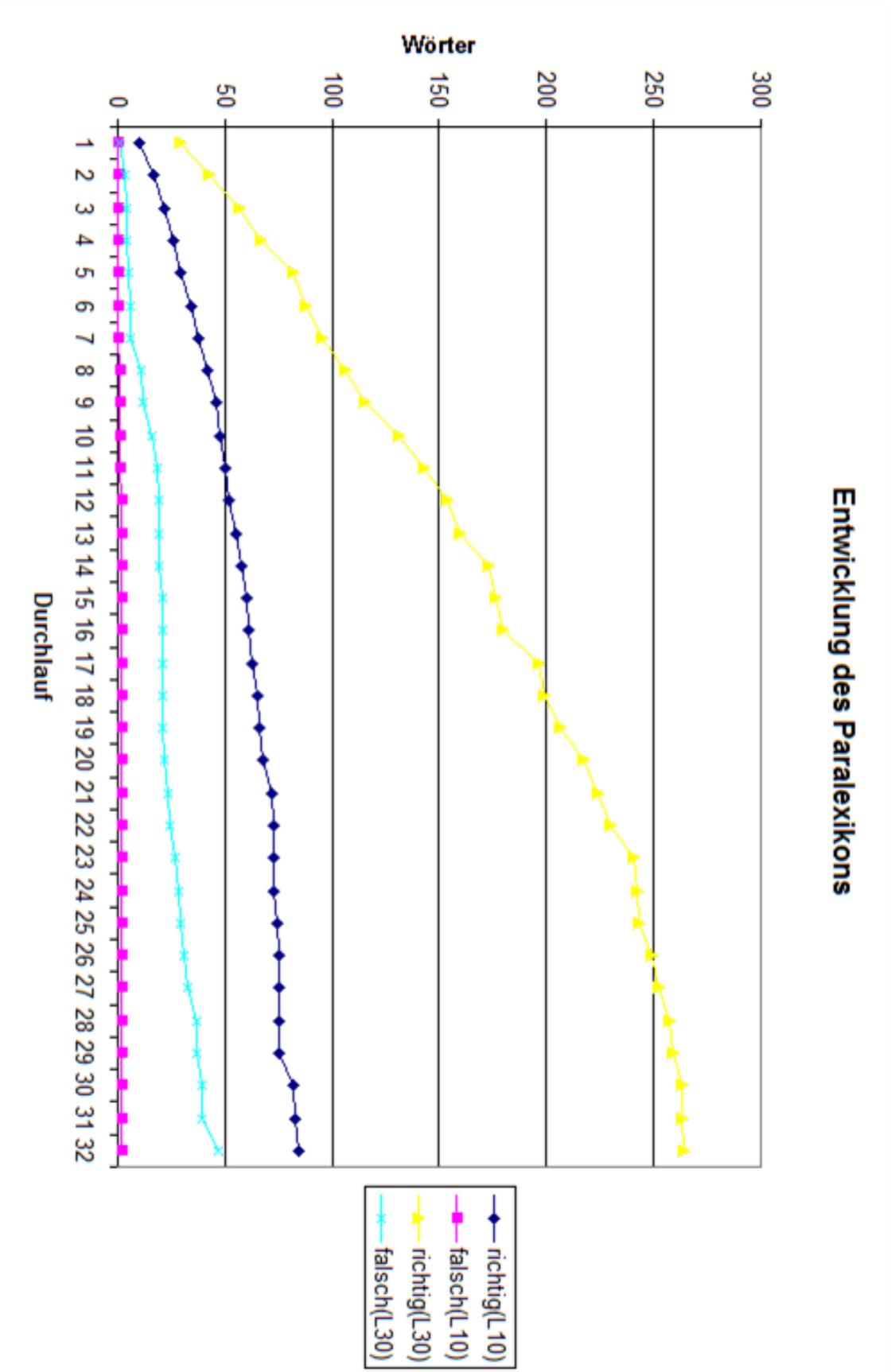


Abbildung 6.25: Entwicklung des Paralexikons

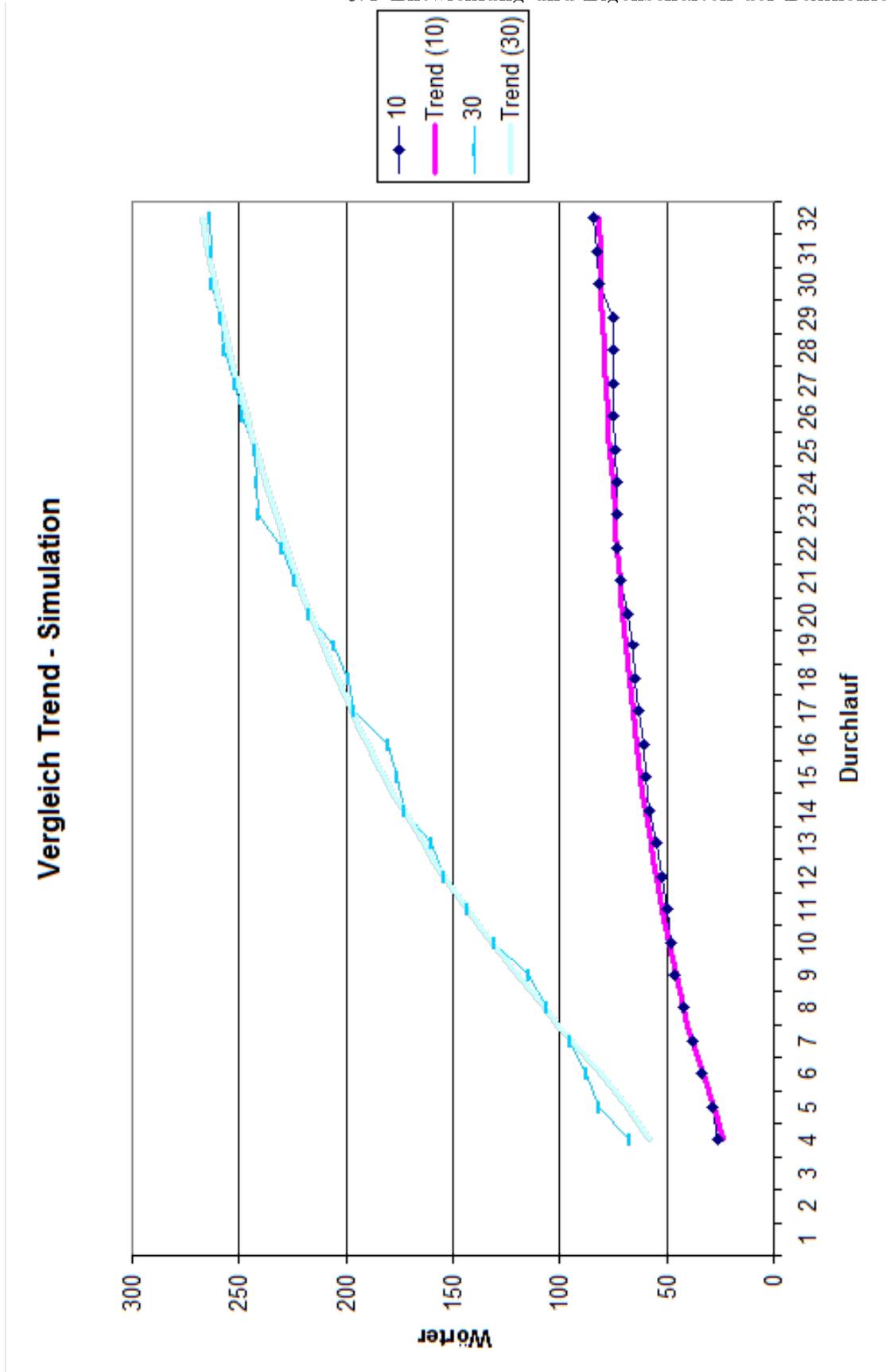


Abbildung 6.26: Trendlinien für $\zeta = 10$ und $\zeta = 30$

6 Simulationsergebnisse und Auswertung

nicht entscheidend für die Entwicklungsrichtung ist. Der funktionale Zusammenhang der Trendlinien entspricht dabei exakt einer allgemeinen Funktion der Form:

$$f(x) = \frac{a}{e^{\frac{b}{x}}} + c \text{ beziehungsweise } f(x) = ae^{\frac{b}{x}} + c \text{ für } 2 \leq x \leq \infty^{13} \quad (6.1)$$

Diese Funktion ermöglicht Vorhersagen im Unendlichen. Lässt man die Zahl der Perioden beziehungsweise der Durchläufe ω unendlich groß werden, erhält man die maximale Größe des Paralexikons.

$$\lim_{\omega \rightarrow \infty} \frac{a}{e^{\frac{b}{\omega}}} + c = a \frac{a}{e^0} = a + c \quad (6.2)$$

Für $\zeta = 10$ wurden folgende Koeffizienten näherungsweise numerisch ermittelt:

- $a = 90$,
- $b = 10$ beziehungsweise -10 und
- $c = 17$

und damit ein Grenzwert von 107 Wörtern errechnet. Analog dazu erfüllen

- $a = 340$,
- $b = 14$ beziehungsweise -14 und
- $c = 48$

für $\zeta = 30$ die Funktion. Das ergibt einen Grenzwert von 398 Wörtern. Somit können die Obergrenzen der Paralexika auch quantitativ spezifiziert werden.

Eine Lexikonfunktion von 30 Einträgen pro Durchlauf ist die maximale Größe, welche man bei 10.000 Zeichen sinnvollerweise wählen sollte, weil bei höheren Funktionswerten die falschen Segmentierungen exponentiell wachsen. Als untere Grenze erscheint es angemessen 10 Wörter zu unterstellen. Kleinere Lexikonfunktionen sind zwar auch möglich, aber dann weichen die F_1 -Werte doch deutlich ab. Sie liegen mindestens fünf Prozent unter den Ergebnissen von Lexikonfunktionen zwischen 10 und 30 Wörtern. Drei weitere Stichproben für $\zeta = 15$, $\zeta = 20$ und $\zeta = 25$ (Abb. 6.28) haben gezeigt, dass sich die Parameterwerte für Gleichung (6.1)

$$f(x) = \frac{a}{e^{\frac{b}{x}}} + c$$

nahezu linear zu einander verhalten und deshalb auch linear interpoliert werden können, sodass schließlich auch eine lineare Funktion die Grenzwerte für jedes gegebene Sigma eindeutig beschreibt. Man betrachte also, wie sich für jedes x (=Durchlauf ω) die Punkte der einzelnen Lexikonfunktionen zueinander verhalten. Da sie für jedes x im Verhältnis

¹³Der Definitionsbereich wurde aufgrund des anfänglich unterschiedlichen Verlaufs der Funktion so gewählt. Man kann an der Konstante bereits erkennen, dass der Anfang des Lexikons in Periode 1 nicht dem Ergebnis der Simulation entspricht.

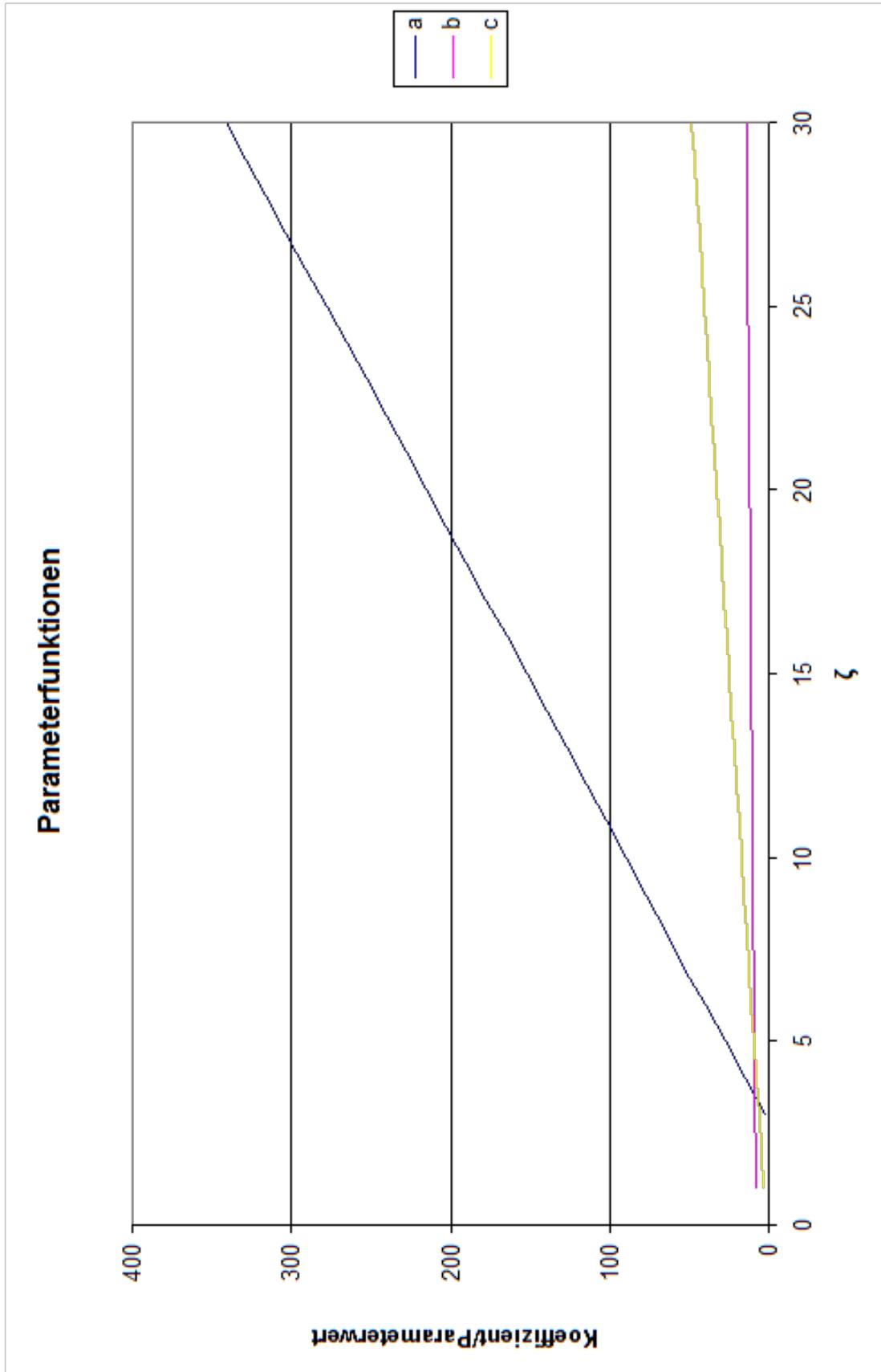


Abbildung 6.27: Lineare Interpolation der Koeffizienten

6 Simulationsergebnisse und Auswertung

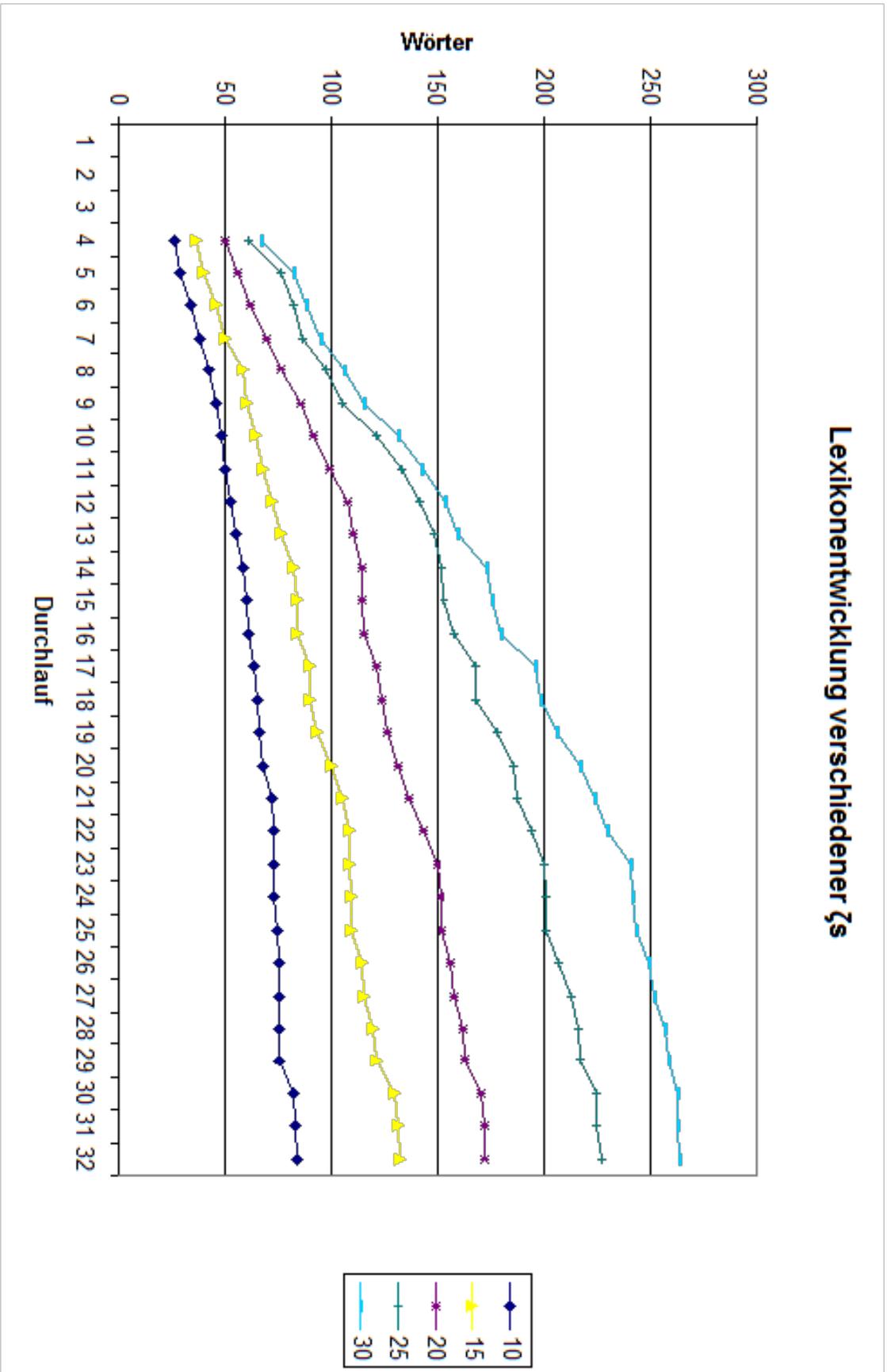


Abbildung 6.28: Entwicklung der Paralexika für $10 \leq \zeta \leq 30$

6.4 Entwicklung und Eigenschaften der Lexikonfunktion

gleich bleiben liegt Linearität vor. Eine einfache Verhältnisgleichung zwischen den y-Werten der Funktionen, welche die Lexikonentwicklung beschreiben, bei gegebenen x, liefert die Funktionsgleichungen (grafisch in Abbildung 6.27).

$$f(\zeta_a) = 12,51\zeta_a - 35,1 \quad (6.3)$$

$$f(\zeta_b) = 0,2\zeta_b + 8 \quad (6.4)$$

$$f(\zeta_c) = \frac{31}{20}\zeta_c + 1,5 \quad (6.5)$$

Für die Berechnung des Grenzwertes für ein beliebiges $\zeta_i > 2$, addiert man analog zur oben erfolgten Rechnung die betreffenden Parameterfunktionen 6.3 und 6.5

$$f(\zeta_a) + f(\zeta_c) \quad (6.6)$$

und man erhält:

$$f(\zeta_i) = \frac{1}{5}(70,3\zeta_i - 168) \quad (6.7)$$

Die Parametergleichungen 6.3-6.5 können auch in die Ausgangsgleichung 6.1 eingesetzt werden, sodass man für jedes ζ_i die entsprechende Wortanzahl in einem bestimmten Durchlauf ω erhält:

$$f(\omega) = \frac{12,51\zeta - 35,1}{e^{\frac{\zeta}{5\omega} + \frac{8}{\omega}}} + \frac{31}{20}\zeta + 1,5 \text{ für } 2 \leq \omega \leq \infty, 10 \leq \zeta \leq 30 \quad (6.8)$$

Gleichung (6.8) beschreibt die Funktionenschar für die Werte $\zeta = 10$, $\zeta = 15$, $\zeta = 20$, $\zeta = 25$ und $\zeta = 30$). Diese Funktionen entsprechen natürlich genau den Trendvorhersagen von Abbildung 6.29. Man siehe dazu Abbildung 6.28 für einen Vergleich mit den tatsächlichen Messwerten der Simulation.

In welchem Zusammenhang muss man die falschen Segmentierungen für ein jedes ζ betrachten? Die Aufnahme von falsch segmentierten Einheiten ins Lexikon lassen sich in keiner linearen Abhängigkeit von ζ beschreiben wie Abbildung 6.30 zeigt. Es lässt sich aber ein kubischer oder gar ein Polynom des vierten Grades als Trend für alle Verläufe von $\zeta > 15$ erkennen. In Abbildung 6.31 ist dieser Verlauf als gleitender Durchschnitt beispielhaft für $\zeta = 25$ eingezeichnet und kann mit folgenden Koeffizienten spezifiziert werden:

$$f(\omega) = 1,21 \cdot 10^{-4}\omega^4 - 5,85 \cdot 10^{-3}\omega^3 + 4,97 \cdot 10^{-2}\omega^2 + 1,44\omega - 1,24 \cdot 10^{-1} \quad (6.9)$$

Für ζ müssen die Koeffizienten mittels der Messwerte bestimmt werden. Ein offensichtlicher funktionaler Zusammenhang der Fehlerfunktionen (Abb. 6.30) ist nicht gegeben. Erkennbar ist jedoch, dass mit abnehmenden ζ , der Einfluss des dritten Polynoms immer größer wird, d.h., der Anstieg zwischen den beiden Wendepunkten ist für eine zunehmende Zahl an Durchläufen sehr klein oder sogar Null.

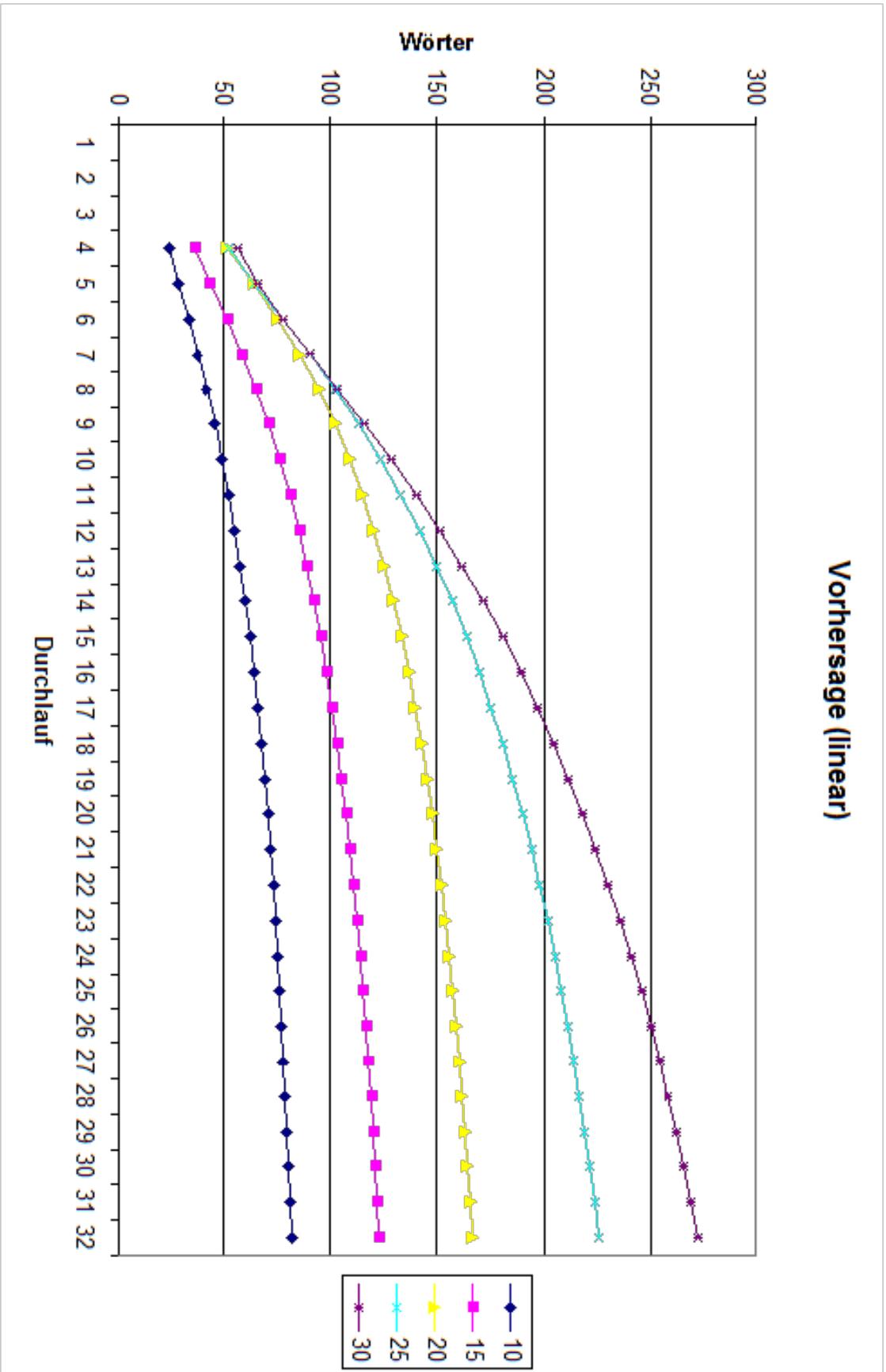


Abbildung 6.29: Vorhersage der Entwicklung des Paralexikons

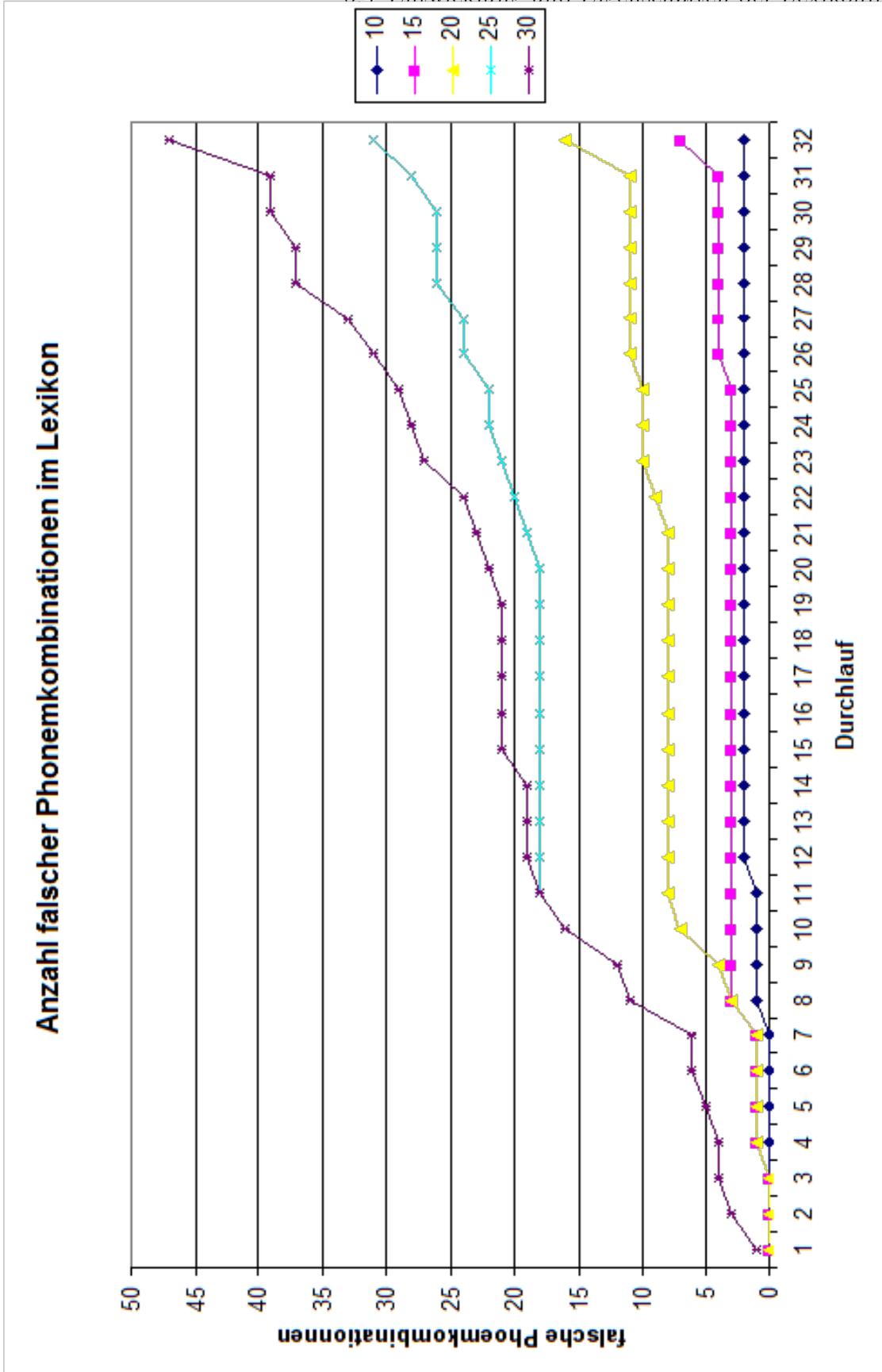


Abbildung 6.30: Fehlerhafte Einträge im Paralexikon

6 Simulationsergebnisse und Auswertung

Verfolgt man den Verlauf des Polynoms weiter (Abb. 6.32), wird sehr schnell klar, dass nach nur wenigen Perioden über 30 die Aufnahme falscher Sequenzen ins Paralexikon stark überwiegt. Nach 52 Perioden ist die Zahl zwischen richtigen und falschen Sequenzen gleich und nach nur zwei oder drei weiteren Perioden überwiegen fehlerhafte Segmentierungen. Der Verlauf der Fehlerkurve in dieser Art ist in einer weiteren Fortführung nicht möglich, da die Phonemsequenzen endlich sind. Die Skalenerträge werden also bei zunehmenden Durchläufen abnehmen und die Gesamtheit der Kurve zu einer S-Gestalt formen. Interessant ist dennoch, dass für $\zeta > 20$ die Schnittpunkte zwischen Fehlerfunktion und Lexikonentwicklung sehr nahe bei einander liegen. Sie streuen um den Wert 52 Durchläufe.

Unabhängig davon, wie die Gestalt der Fehlerfunktion von den Durchläufen größer als 30 aussieht, lässt sich mit Gewissheit sagen, dass eine Mustererkennung nach weiteren Segmentierungsgesetzmäßigkeiten im Paralexikon nach $f''(\omega) = 0$ nicht mehr möglich ist. Das Verhältnis zwischen Fehlern und richtigen Segmentierungen ist schlichtweg zu groß, um sie bei der Suche nach Gesetzmäßigkeiten vernachlässigen zu können. Da sich bei allen Funktionen mit $\zeta > 15$ ein derartiger Verlauf der Fehlerfunktion andeutet, musste die Entscheidung für eine Lexikonfunktion von 10 Wörtern pro Durchlauf fallen. Dies wird noch durch einen glücklichen Zufall erleichtert.

Überschreitet ζ eine bestimmte Höhe (um 10 Wörter) wirken sich die unterschiedlichen ζ -Werte nicht mehr signifikant auf die F_1 -Werte aus (siehe Tabelle 6.4), wie man nach der oben angestellten Überlegung intuitiv annehmen würde. Es geht also nicht nur um die Größe des Paralexikons, sondern auch um die Qualität der Einträge. Bei großem ζ ist die Wahrscheinlichkeit zwar höher, dass mehr Wörter aufgenommen werden, weil sie in dem jeweiligen Korpus auch entsprechend häufig vertreten sind; sie kommen aber auf Dauer und im Durchschnitt eher selten bei anderen Korpora vor (siehe Anhang B). Das hat zur Folge, dass sie bei Einsatz der Rekursion keinen Beitrag leisten, weil sie nicht im Korpus stehen. Ein möglichst großes Paralexikon ist zwar weiterhin zu präferieren, insbesondere wenn es die grammatischen Morpheme am Eintrag spezifiziert, aber dies muss immer auch in Abwägung zu der Qualität der Einträge und zu den Fehleinträgen geschehen.

Tabelle 6.4 zeigt warum. Die Perioden 29 und 30 geben für beide ζ die gleichen Messwerte, weil es sich bei beiden Simulationen um die gleiche Anordnung von Korpora handelt. Nach 30 Perioden wird das akkumulierte Paralexikon eingefügt und die F_1 -Werte verbessern sich in beiden Simulationen um fast 30 Prozent. In Periode 31 sehen die Werte für das größere Paralexikon etwas besser aus, weil die Präzision aufgrund geringerer falscher Segmentierungen, besser ausfällt. Aber schon in der Folgeperiode verdoppeln sich die falschen Segmentierungen bei $\zeta = 30$. Dieses Ergebnis unterstreicht die vorher angestellte Überlegung, dass die Größe des Paralexikons ausschlaggebend für eine deutliche Erhöhung der Segmentierungsleistung ist. Die Fehlerquote kann man davon jedoch nicht unabhängig betrachten. Die Größe des Lexikons hängt von ζ und von der Anzahl der Perioden vor Einsatz der Rekursion (μ) ab. Während μ möglichst groß zu wählen ist, darf ζ 30 Wörter nicht wesentlich überschreiten, weil die Fehlerquote sonst expo-

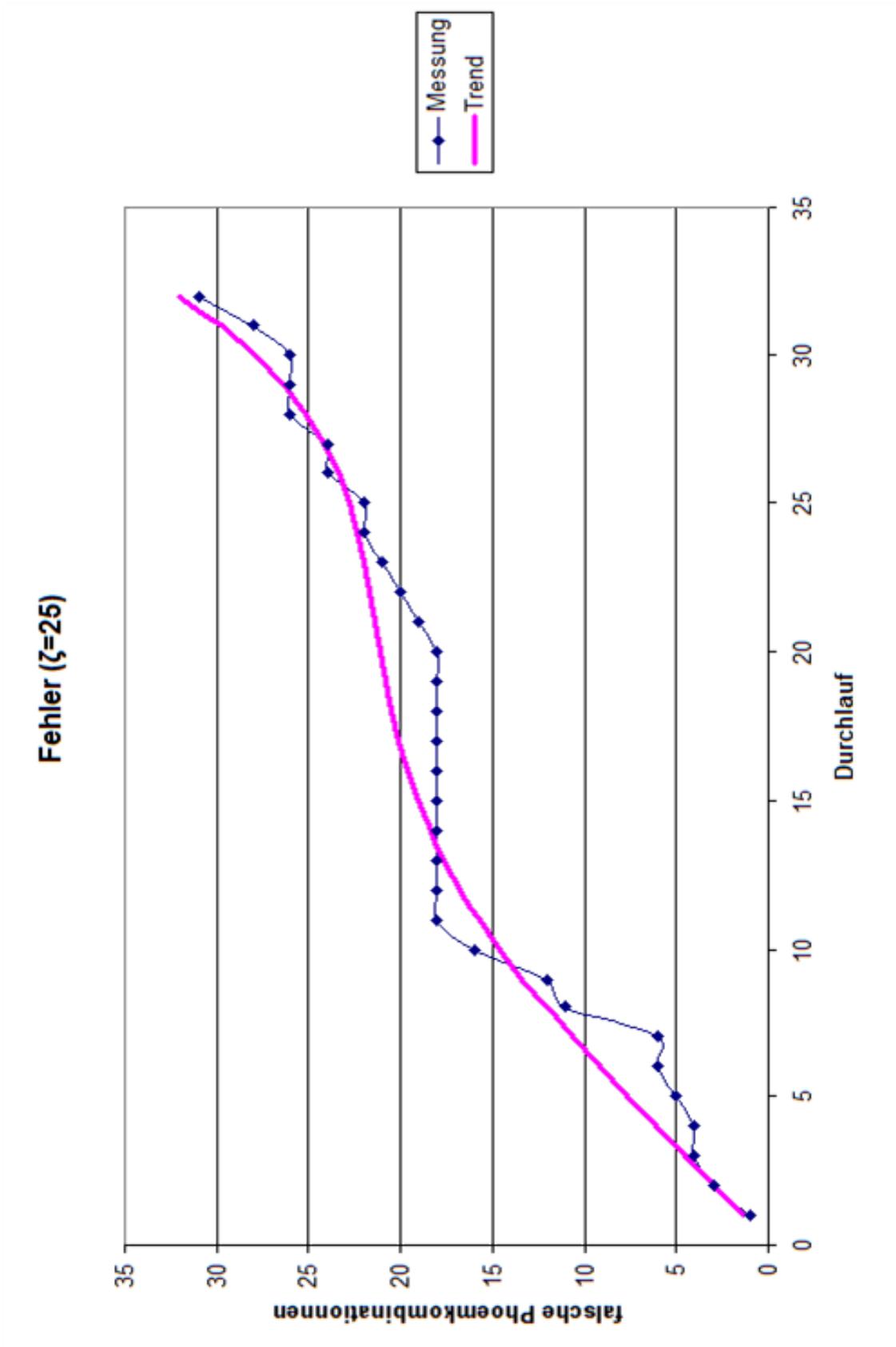


Abbildung 6.31: Vergleich zwischen tatsächlichen Fehlern und Vorhersage

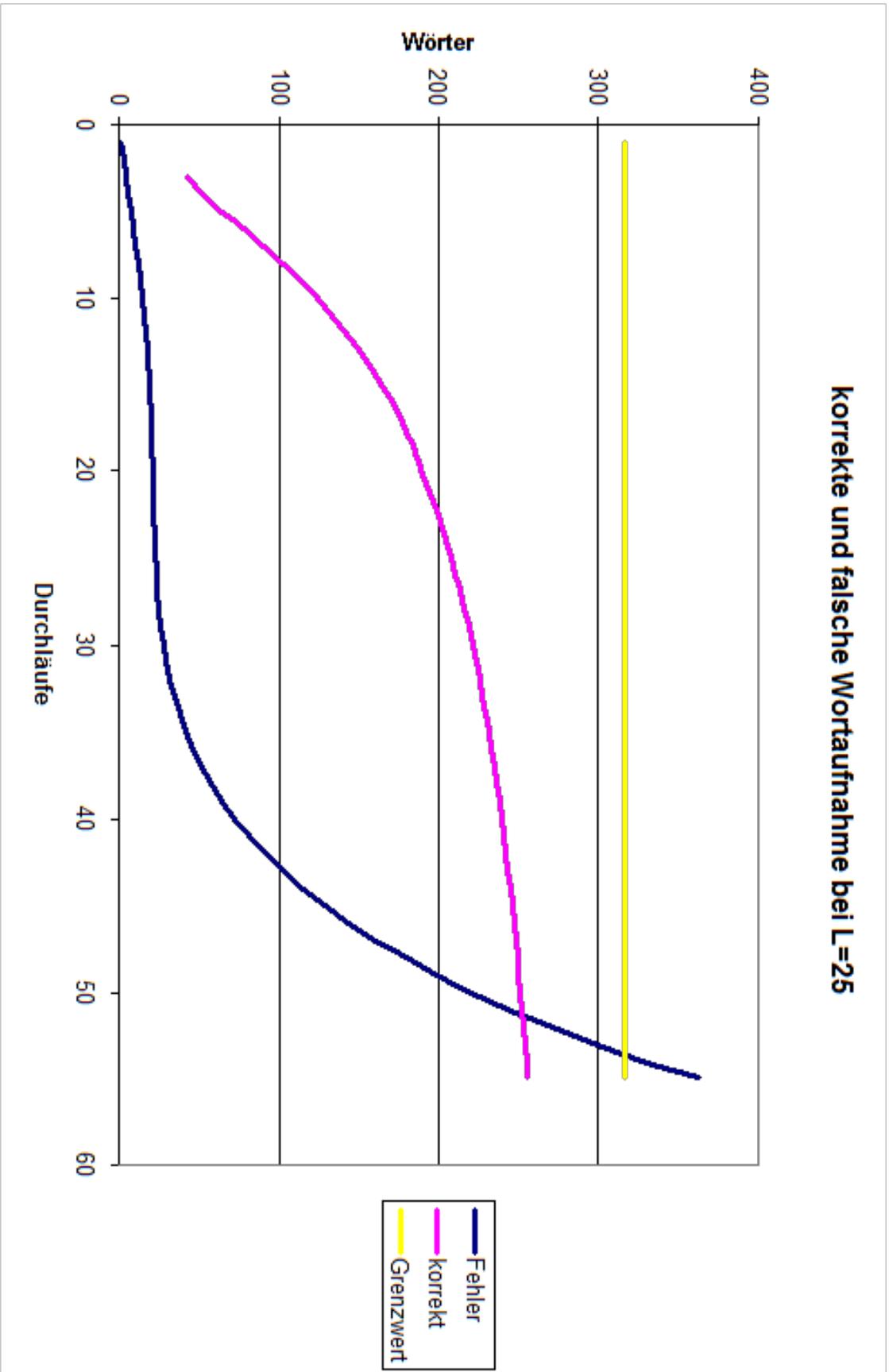


Abbildung 6.32: Trends für falsche und richtige Wörter im Paralexikon

	Periode	richtig	falsch	nicht	precision	recall	F_1
$\zeta = 10$	29	1132	493	1427	0.70	0.44	0.54
	30	1146	416	1467	0.73	0.44	0.55
	31	2127	602	322	0.78	0.87	0.82
	32	2164	759	432	0.74	0.83	0.78
$\zeta = 30$	29	1132	493	1427	0.70	0.44	0.54
	30	1146	416	1467	0.73	0.44	0.55
	31	2133	548	316	0.80	0.87	0.83
	32	2283	1000	313	0.70	0.88	0.78

Tabelle 6.4: Vergleich der Leistungsmaße für $\zeta = 10$ und $\zeta = 30$

nentiell ansteigt. Abbildung 6.33 zeigt schematisch und Abbildung 59 faktisch wie sich F_1 in Abhängigkeit vom Rekursionsanfang entwickelt. Vergehen nur wenige Perioden, enthält das Paralexikon erst wenige Einträge (siehe Abb. 6.25), aber auch keine Fehler. Mit steigender Periodenzahl wachsen die Einträge logarithmisch. Auf gleiche Weise verhalten sich die F_1 -Werte. Beobachtbar ist auch, dass sich die Lexikonfunktionen mit wachsender Periodenzahl weniger stark absenken und dies lässt den Schluss zu, dass sie sich bei entsprechend hoher Periodenzahl ω dem Verlauf der Funktion, welche die F_1 -Maximalwerte beschreibt $\text{argmax}(\text{LexFunc})$, angleicht – der Verzögerungseffekt im weiteren Fortgang also gänzlich aufgehoben wird. Dieser Gedanke wurde in Abbildung 6.33 schematisch dargestellt. Die allmähliche Hebung der Funktionenschar LexFunc tritt besonders deutlich bei einem Rekursionsanfang nach 25 Perioden auf.

Die anfangs gestellten Fragen nach der Größe des Paralexikons und der notwendigen Periodenzahl μ , um die Lexikongröße zu erreichen, kann nun zusammenfassend beantwortet werden. Die Größe des Paralexikons sollte 30 Einträge nicht überschreiten, da ansonsten auch schon Einträge ins Paralexikon aufgenommen werden, die nur zwei- oder dreimal segmentiert wurden. Darunter würden sich sehr viele zufällige und korpuspezifische (idiosynkratische) Segmentierungen wiederfinden, die aber keinen Worteinheiten entsprechen und normalerweise unter der Annahme der häufigkeitsabhängigen Verarbeitung des Gehirns nicht länger gespeichert werden würden. Nimmt ζ einen geringeren Wert an, geht dies natürlich nur mit den Kosten eines kleineren Paralexikons einher. Im Allgemeinen ist dies nicht wünschenswert, weil die zuvor gemachten Beobachtungen eindeutig auf den positiven Effekt eines großen Paralexikons hinweisen. Dieser Umstand trifft natürlich auch hier zu. Dagegen aber gilt es, die hohe Fehlerquote abzuwägen.

Die Entscheidung für einen spezifischen ζ -Wert unterliegt also einer doch recht weitläufigen Interpretation. Diese ist hier zugunsten eines kleinen ζ ausgefallen. Dafür können zwei Gründe angegeben werden. Es konnte gezeigt werden, dass die F_1 -Werte nicht wesentlich voneinander abweichen. Das widerspricht zunächst einmal dem Befund, ein möglichst großes Paralexikon anzulegen. Man fragt: Wenn sich die Größe des Paralexikons nicht auf die Segmentierungsleistung auswirkt, dann wären die vorangestellten Überle-

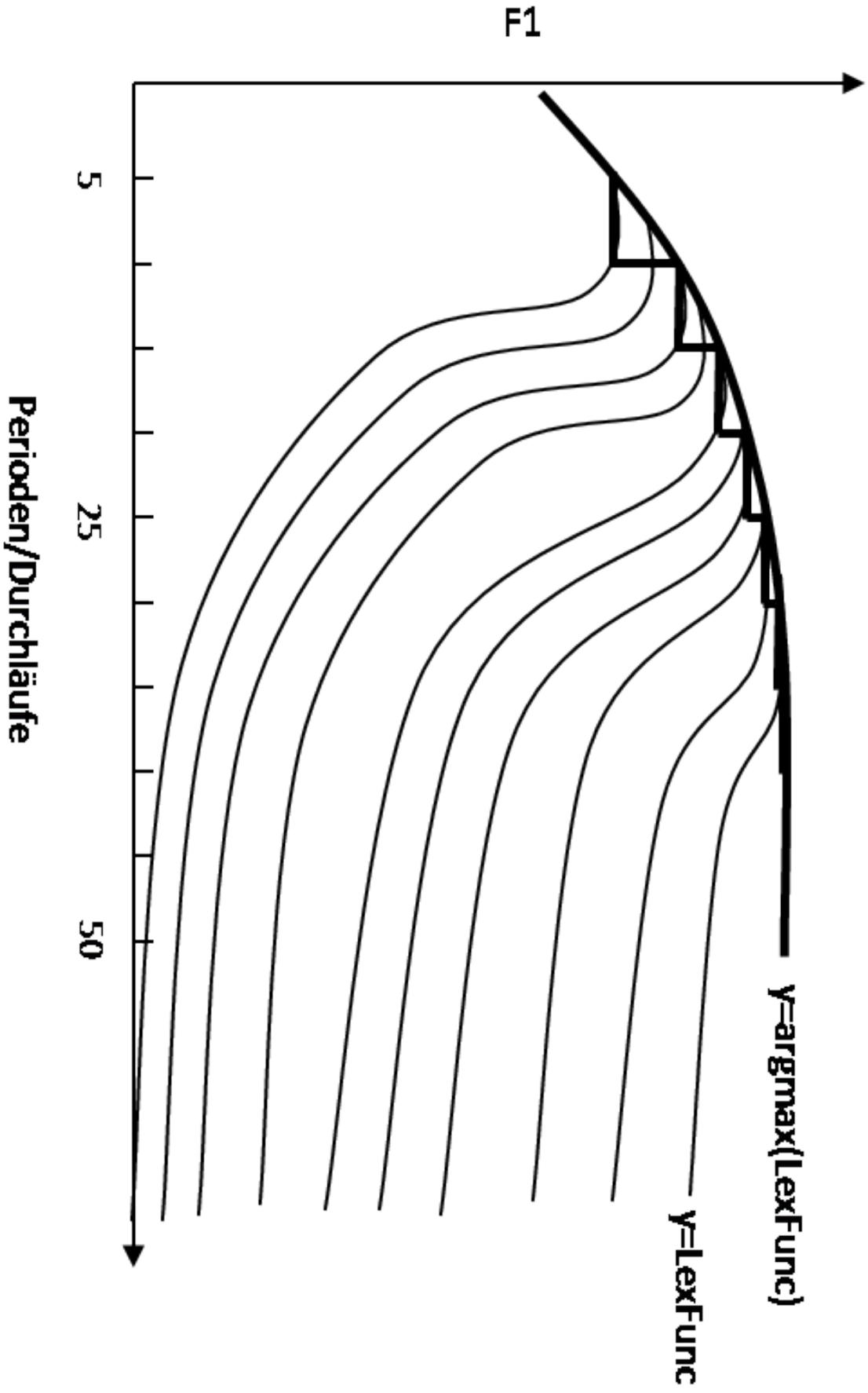


Abbildung 6.33: Zusammenhang zwischen F_1 und μ ; $\zeta = 10$ (schematisch)

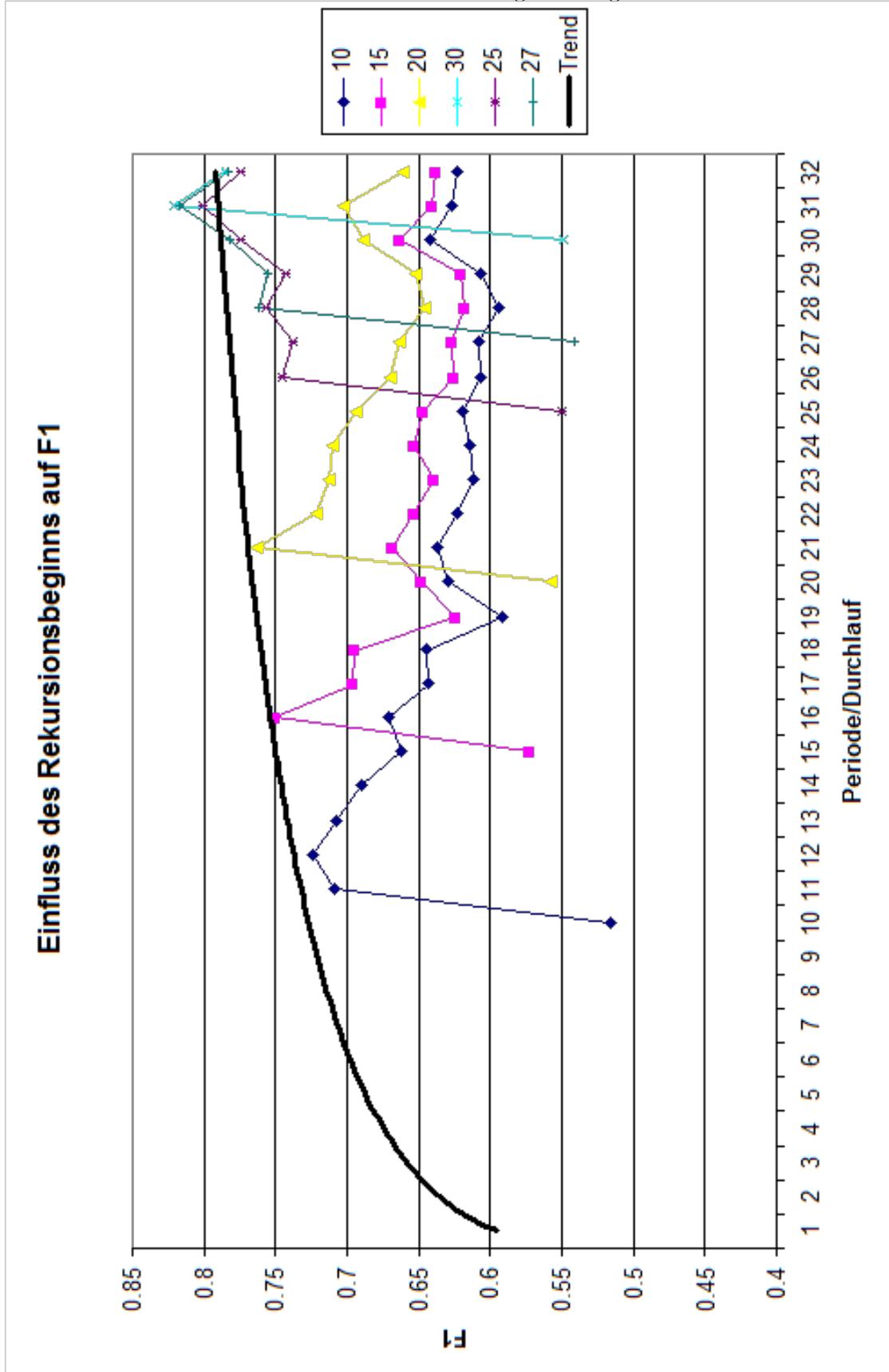


Abbildung 6.34: Zusammenhang zwischen F_1 und μ ; $\zeta = 10$ (faktisch)

gungen und Beobachtungen falsch beziehungsweise fehlinterpretiert. Das ist aber nicht der Fall. Eine nähere Analyse des Datenmaterials zeigte, dass es sich um ein Kritisches-Masse-Phänomen handelt. Dabei fallen die entsprechenden F_1 -Werte zwischen $\zeta = 10$ und $\zeta = 30$ in das gleiche Intervall. Da nun $\zeta = 10$ keine statistisch relevanten Fehler aufweist, fiel die Entscheidung auf ein kleineres Paralexikon.

Der zweite Grund für diese Entscheidung ergibt sich aus der Beantwortung folgender Fragen: Reicht die Größe des Paralexikons aus, um weitere Gesetzmäßigkeiten der Sprache eindeutig abzuleiten? Kann man mittels eines internen Rekursionsprozesses feinere Segmentierungen vornehmen? Dies ist mit Wortlisten von 80 Einträgen sicherlich gut möglich, wenn man auch einer Liste von 250 Einträgen eine höhere Zuverlässigkeit (Reliabilität) zugestehen müsste. Wiederum wirkt sich auch dabei die hohe Fehlerquote des größeren Paralexikons aus, da man nicht davon ausgehen darf, dass der Sprachlerner falsche und richtige Segmentierungen bereits unterscheiden kann. So würden die Falscheinträge einmal die Mustererkennung weiterer Gesetzmäßigkeiten erschweren (z.B. $/zj\partial r/$, $/t\partial\partial t/$ oder $/z\partial\partial t/$ in der Phonotaktik), zum anderen würden längere Phrasen nicht in ihre Bestandteile oder nur in falsche zerlegt werden können (bei $/werzj\partial r/$, $/izj\partial r/$ oder $/iz\partial\partial t/$). Dies ist bei einer geringen Fehlerquote nicht der Fall, sodass die Entscheidung wiederholt auf ein kleines ζ fallen sollte (siehe Anhang B für das Datenmaterial beider Paralexika).

Mit der zweiten Frage sollte eine Beantwortung auf die Anzahl der Perioden für die Erstellung des Paralexikons gegeben werden. Aus den Daten in Abbildung 6.25 geht hervor, dass sich für $\zeta = 10$ nach 20 Perioden nur noch geringe Wortzuwächse verzeichnen lassen, für $\zeta = 30$ wird die dritte Ableitung erst in den 20er Perioden negativ und erhält dann nach 30 Perioden keinen signifikanten Zuwachs mehr. Favorisiert man also ein niedriges ζ , dann hat man bereits nach 25 Perioden eine dem Maximum schon recht nahe Größe des Paralexikons erreicht.

Letztlich seien die Eigenschaften des Paralexikons detaillierter erklärt. Mit einem kritischen Blick auf das Datenmaterial, das die bislang ermittelten Idealwerte benutzt ($\zeta = 10$, $\mu = 30$, $\sigma = 0.1$, $\lambda = 2$), können Gesetzmäßigkeiten hergeleitet werden. Dies geschieht unter Berücksichtigung experimenteller Befunde aus der Psycholinguistik und Entwicklungspsychologie.

6.5 Eigenschaften des Paralexikons

Die Tabellen 6.5 und 6.6 zeigen das Analysematerial, welches für die Erkennung weiterer Muster verwendet werden kann. Erst wenn ein solche Wörterliste mit den Übergangswahrscheinlichkeiten und einer rekursiven Einsetzungsprozedur erstellt ist, können weitere Informationen aus diesem Material gewonnen werden, so wie es in Jusczyk et al. (1993a), Jusczyk et al. (1993b), Jusczyk et al. (1994), Jusczyk et al. (1999a), Jusczyk und Aslin (1995), Jusczyk (1998) und Jusczyk (1999) beschrieben wird. Dieser wichtige erste Schritt, die Erstellung des Paralexikons, wird in der Literatur sehr elegant umgan-

gen, indem es immer implizit vorausgesetzt oder als Annahme, die noch näher zu klären sei, vermerkt wird. In der Tat ist dieses Material aber die Grundlage für weitergehende Segmentierungsmechanismen, die hier beispielhaft an der Wörterliste gezeigt werden.

Das in Tabelle 6.6 erstellte Paralexikon umfasst in etwa 80 Einträge. Jeder dieser Einträge kann als Einheit erkannt werden. Dies ist der erste wichtige Schritt. Jeder Eintrag hat ein bestimmtes Betonungsmuster, welches beispielsweise in Jusczyk et al. (1999c), Johnson und Jusczyk (2001) oder Norris' MSS (Norris et al., 2001) eine entscheidende Rolle spielt. Worteinheit und Betonung können aber erst anhand einer solchen Wortliste in einen Zusammenhang gebracht werden, nicht schon in einem unsegmentierten Text. Die Information, wo die Einheit aufhört und endet, ist in einem unsegmentierten Text nicht vorhanden. Die hier gelisteten Einträge erlauben diesen Zusammenhang herzustellen. Alle Wörter sind mit nur einer Ausnahme (*/əvker/*) anfangsbetont. Da man nun die Wortgrenzen kennt, kann man beide Informationen – Wortgrenze und Stelle der Betonung – zusammenfügen und die Regel ableiten, dass der Wortanfang durch die betonte Silbe definiert sei.

/əvker/ ist auf der zweiten Silbe betont und hat auf */əv/* nur eine sekundäre Betonung. Mit fortlaufender Entwicklung des Paralexikons werden sich auch diese Fälle häufen und die Regel wird eine Anpassung erfahren müssen. Dieser Zustand konnte auch experimentell belegt werden (siehe Kap. 3.2.1). Erst mit neun beziehungsweise zehneinhalb Monaten tritt die Regeländerung ein (Jusczyk et al., 1999c). Nach den hier angestellten Überlegungen würde das Kind genau diese Zeit brauchen, um genügend Material im Paralexikon zu sammeln. Bis dahin würde das Kind jedoch mit der *Betonung-gleich-Wortanfang-Regel* verfahren. Wenn es */əvju/* hört, bei welchem das */əv/* betont ist, hat es sogleich einen Fall, welcher die Regel wieder bestätigt.

/ɪj/ kann als einziger Eintrag kein dominantes Betonungsmuster aufweisen, weil es am Wortende immer unbetont auftritt. Auf diese Weise könnte es als Sonderfall auch schon von einem Säugling erkannt werden.

Interessant ist doch nun, wie das Kind mit Einheiten umgeht, welche wirklich mehrere Wörter enthalten und welche so zunächst nicht mit der Regel in Einklang zu bringen sind. Der Sprachlerner könnte sich einerseits denken, dass nicht alle primären Betonungsmuster mit einer Wortgrenze zusammenfallen. Andererseits könnte es die betreffenden Einträge der Regel anpassen. Passt das Kind die Regel an, bleibt das Paralexikon wie es ist und es mag sich im späteren Verlauf des Spracherwerbs als positiv herausstellen, wenn mehr Ausnahmen in das Paralexikon gelangen. Die andere Möglichkeit wäre es, das Lexikon zu ändern. Diese Variante würde sich auf die Segmentierung sicherlich auch kurzfristig positiv auswirken. Der Sprachlerner müsste jedoch dann sicherlich später die Regel anpassen, um die Ausnahmen der zweitbetonten Silben zu bewältigen. Beide Strategien würden zum Ziel führen. Für die zweitgenannte Strategie hätte der Sprachlerner bereits nach der ersten Periode genügend Material zur Verfügung, um die Musterunterschiede im Paralexikon zu beseitigen. Von den 10 Einträgen des ersten Durchgangs fallen die Wortanfänge bei neun Einträgen mit der starken Betonung zusammen. Beim letzten Eintrag */ðerjugəv/* stimmt dies zwar auch, aber innerhalb der Einheit gibt es

6 Simulationsergebnisse und Auswertung

Rang	$\omega = 1$	$\omega = 2$	$\omega = 3$	$\omega = 4$	$\omega = 5$	6	
1	øt	øt	ju	øt	øt	øt	93
2	wat	øer	øt	ju	ju	ju	89
3	je	mrgan	øer	øer	jor	and	56
4	ltal	ju	lets	drju	wat	øts	48
5	ju	wat	and	wat	øer	jor	47
6	øer	je	drju	lets	and	wat	46
7	drju	wer	hr	jor	hr	øer	45
8	and	ltal	mrgan	and	drju	mrgan	44
9	dun	øts	wat	wer	lets	hr	37
10	øerju	drju	je	hr	øts	drju	37
11		and	øker	mrgan	wer	lets	31
12		dun	wer	hr	øts	drju	25
13		øerju	ltal	hr	wer	drju	24
14		momi	kan	øker	hr	hr	23
15		lets	wi	red	je	hr	19
16		jor	øts	ltal	øker	je	17
17		hr	ru	kan	red	øker	16
18			hav	wi	ltal	red	14
19			dun	tu	kan	ltal	14
20			øerju	øts	kan	kan	13
21			momi	ru	lok	wi	13
22			jor	dun	kadlitedi	tu	13
23				øerju	kan	prt	13
24				momi	ru	kadlitedi	13
25				ða	hand	wot3	12
26					dun	ap	12
27					øerju	lok	12
28					momi	kan	12
29					ða	ts	12
30					hand	hand	12
31					dun	øerju	12
32					dr	øerju	12
33					momi	tr	11
34					ða	momi	11

Tabulle 6.5: Lexikonentwicklung für $\omega = 6, \zeta = 10$

Rang	Eintrag	Anzahl	Rang	Eintrag	Anzahl	Rang	Eintrag	Anzahl
1	ju	543	29	its	34	57	draijør	12
2	jør	482	30	ør	28	58	si	12
3	ænd	359	31	jø	28	59	ap	12
4	ðæt	238	32	wer	25	60	ðabøl	12
5	hir	189	33	dzus	24	61	krti	12
6	mømi	185	34	kam	23	62	hø	12
7	wat	174	35	lital	23	63	wønt	12
8	ðer	139	36	før	22	64	ðerjugøu	12
9	ðλ	123	37	dumj	22	65	hænd	12
10	sam	88	38	løk	21	66	gøu	11
11	mørgan	87	39	tjikλn	18	67	beibi	11
12	ol	80	40	it	18	68	wø	11
13	ðøts	79	41	red	16	69	watørjudumj	11
14	ij	68	42	døunt	16	70	nøiz	10
15	køn	67	43	λn	15	71	mømiz	10
16	dilλn	66	44	trai	15	72	ðis	10
17	tu	65	45	iz	14	73	dzλst	10
18	θij	65	46	kλmøn	14	74	ðøtsrart	10
19	mændi	64	47	øuju	13	75	kønju	10
20	ørju	60	48	øukei	13	76	fu	9
21	lets	59	49	her	13	77	tøn	9
22	mør	59	50	kλdlitedi	13	78	zænd	9
23	get	47	51	bøk	13	79	hiz	9
24	høv	47	52	priti	13	80	bλbal	8
25	øn	43	53	λm	13	81	fat	8
26	wi	40	54	nøuju	12	82	nøu	8
27	wøtz	37	55	kønjuser	12	83	of	8
28	øukei	35	56	wλn	12	84	raund	8

Tabelle 6.6: Lexikon nach 30 Durchläufen bei $\zeta = 10$, $\mu = 30$

6 Simulationsergebnisse und Auswertung

noch zwei weitere Betonungen gleicher Intensität.¹⁴ Da nun /ju/ bereits als Einheit im Lexikon steht, könnte das ‚analytische‘ Kind, diese Teilmenge mit genau dem Mechanismus substituieren, welchen es auch bei der rekursiven Struktur anwenden würde und von Marcus et al (1999) experimentell überprüft wurde (siehe Kap. 3.2.4). Im zweiten Durchgang würde das Kind auch so mit /vrju/ verfahren. Dieser Vorgang könnte sich auch mit /əvkei/ wiederholen.¹⁵ Ab der dritten Periode steht aber das /ɪŋ/ im Lexikon. Nach der vorgestellten Logik müsste es auch aus /dʊŋ/ herausgelöst werden. Eine mögliche Erklärung könnte das nicht vorhandene Betonungsmuster von /ɪŋ/ sein, die das Kind dazu verleitet, keine Trennung vorzunehmen. Die Regel lautet schließlich „starke Silbe = Wortanfang“. Da dies nicht zutrifft, findet auch keine Segmentierung statt. Ähnlich könnte man auch bei /əvkei/ argumentieren. Dieser Eintrag hat eine abweichende Struktur von allen anderen. Weil das /əv/ immerhin noch eine sekundäre Betonung trägt, könnte dies als ein Hinweis gewertet werden, dass es nicht wie andere Einheiten behandelt wird.

Gegen eine Rekursion innerhalb des Paralexikons, wie gerade diskutiert, spricht aber wieder das Problem des Eingebettetseins, denn man kann nicht einfach voraussetzen, dass das Kind nur genau die richtigen Einheiten auswählt, welche tatsächlich echte Zusammensetzungen sind. Zwar ist dieses Problem nun im Vergleich zum unverarbeiteten Input minimal – im gegebenen Paralexikon gibt es nur wenige Fälle (/ʌn/ in /dɪʌn/, /mɔrgʌn/, /tʃɪkʌn/ oder /wʌn/) –, aber es ist existent und bei größeren Paralexika könnte die *Embeddedness* durchaus zu chaotischen Segmentierungen führen.¹⁶ Besser wäre es, wenn der Sprachlerner nun parallel zu der prosodischen Segmentierung, erste Bedeutungen zuordnen könnte und beginnt das morphologische Regelsystem zu entwickeln. Eine abschließende Beantwortung der Frage kann aber erst mit dem Wissen über die genaue Struktur des mentalen Lexikons erfolgen, weil man erst dann weiss, ob beispielsweise das /ɪŋ/ überhaupt einzeln repräsentiert werden sollte.¹⁷

Andererseits gibt es auch Hinweise in der Literatur, dass Kleinkinder längere Phrasen als Einheiten zumindest passiv als atomare Bedeutungsstrukturen verwenden (Tomasello (2003); Peters (1983)). Dies spricht für eine großzügige Regelauslegung und mag kurz-

¹⁴Wie Houston et al. (2004, 129) aufzeigen, ist die Speicherung als ein paralexikalischer Eintrag in diesem Fall tatsächlich beobachtbar, wenn keine prosodischen Informationen einfließen. Bei Zunahme der Wortakzente würde die Phrase jedoch separiert werden.

¹⁵Das „guitaris-Experiment“ (Jusczyk et al., 1999c) gibt den Beweis dafür, dass das anglophone Kind im Alter von siebeneinhalb Monaten nur in der Lage ist, das stark/schwach-Silbenmuster zu nutzen.

¹⁶Algorithmisch ließe sich dieses Problem durch eine zusätzliche Annahme fast vollständig lösen: Eine rekursive Einsetzung ist innerhalb des Paralexikons immer dann möglich, wenn alle Bestandteile eines segmentierten Eintrags bereits im Paralexikon vorhanden sind. Diese Lösung funktioniert für das Beispielllexikon in Tabelle 6.6 nur für das [ɪŋ]-Morphem nicht.

¹⁷Auch hierfür gibt es bereits implementierte Simulationen (Baayen und Schreuder, 2000) und seit längeren entsprechende experimentelle Befunde (Taft und Forster (1975); Taft und Forster (1976); Taft (1979)) zu den genauen Prozessabläufen. Siehe auch Clahsen und Rothweiler (1992) und Clahsen (1999) zur Frage, ob das morphologische Regelsystem ganz im Sinne des Nativismus als ein eigenständiges Modul mit einfacher oder dualer Verarbeitung fungiert. Passend zu dieser Arbeit stellt die Studie von Baayen (1992) einen Bezug zwischen der Produktivität eines Morphems und dessen Häufigkeit her.

fristig zu einer Leistungsminderung führen, da einige Lücken ja nicht erkannt werden, wie zum Beispiel in /*ðerjugəv*/.¹⁸

Problematisch sind /*hə*/ und /*zənd*/ als fehlerhafte Segmentierungen. Hier muss hinzugefügt werden, dass sie aus lediglich einem Korpus stammen und ihre Zahl nicht weiter erhöht haben. Man könnte hier nach dem Grundsatz der Häufigkeit davon ausgehen, dass sie bei der Fortentwicklung des Lexikons mit einigen anderen seltenen Kandidaten nicht weiter beachtet werden, weil sie eben nicht mehr als atomare, eigenständige Einheiten vorkommen und im Verhältnis zu den anderen Einträgen zu klein sind, sonst vergessen werden.

Schließlich lassen sich aus dem Paralexikon noch einige der phonotaktischen Merkmale der Sprache ableiten. Alle Kombinationen, welche innerhalb einer Einheit vorkommen, gelten danach als erlaubte Kombinationen. Auf gleiche Weise wären alle Kombinationen, die überhaupt nicht vorkommen, phonotaktische Marker, welche definitiv eine Wortgrenze bestimmen. Auch diese Information kann erst auf Basis einer vorhandenen Wortliste erfolgen, da man davor nicht weiß, ob man sich nicht doch innerhalb eines Wortes befindet. Und auch hier ist der Sprachlerner erst mit neun bis zehn Monaten in der Lage, diese phonotaktischen Informationen zu nutzen (Jusczyk et al., 1994). Das könnte ein weiterer Hinweis darauf sein, dass das Kind erst eine gewisse Zeit braucht, um das Datenmaterial in der hier beschriebenen Weise anzulegen.

6.6 Zusammenfassung

Nach der Analyse des umfangreichen Datenmaterials¹⁹ konnte eine optimale Kombination von Phonemkettenlänge λ und Sensibilität σ gefunden werden. Diese Kombination erlaubt einen Text mit einer Segmentierungseffizienz von $F_1=0.55$ zu bearbeiten. Dabei hat sich gezeigt, dass die genaue Phonemkettenlänge λ kein entscheidendes Kriterium ist, sondern recht weitläufig divergieren kann. Ab 5000 Zeichen lässt sich für jede Korpusgröße eine Schnittmenge aller F_1 -Maxima ermitteln (Phonemkettenlänge $\sigma = 1$ bis 4; Sensibilität $\lambda = 0.1$). Die Sensibilität σ hingegen sollte minimale Werte annehmen. Allerdings ist für Korpora unter 1000 Zeichen diese Kombination nicht zielführend. Daraus ergibt sich, dass der Sprachlerner einen Input von mindestens 1000 Wörtern als zusammenhängende Berechnungsgrundlage nehmen müsste. Geht man von weniger als 250 Wörtern aus, muss die Änderung der Sensibilität in Abhängigkeit von der Korpus-

¹⁸Brent (1997, 26) betrachtet in seinem DR-Algorithmus derartige Zusammensetzungen immer als ein Wort, wenn keine Teilmenge im bereits segmentierten Material zur Verfügung steht (*autonomous-bootstrapping-Ansatz*).

¹⁹Jeweils 32 Korpora einer Korpusgröße bis zu einer Größe von einschließlich 10.000 Zeichen. (ab 20.000 Zeichen acht, ab 50.000 Zeichen vier, ab 100.000 Zeichen zwei und darüber ein Korpusrepräsentant, siehe Tabelle 5.2, Kap. 5.2.4.1) wurden untersucht. Jeder der Korpusrepräsentanten einer Größe enthielt fünf Durchläufe von jeweils drei unterschiedlichen Lexikongrößen (10, 20, 30). Jeder Durchlauf listete für zehn Sensibilitätsstufen zwischen Null und eins und Phonemketten zwischen eins und zehn alle Messwerte auf. Daraus ergibt sich eine Anzahl von 600.000 F_1 -Werten und 480.000 Angaben zur Wortanzahl, die in bestimmter Ordnung mit einander verglichen werden mussten.

größe erklärt werden. Ein Mechanismus dafür ist aber nicht plausibel, sodass von einem größeren Informationsinput ausgegangen werden darf.

Die Einführung einer rekursiven Struktur bedeutet, dass die Übergangswahrscheinlichkeiten des gleichen oder eines anderen Korpus noch einmal berechnet werden, aber unter Berücksichtigung der Information aus den vorhergehenden Berechnungen. Die Lexikonfunktion beschreibt, wie viel von dieser Information verwendet werden soll. Sie definiert die Anzahl der am häufigsten segmentierten Einheiten für jede nun folgende Periode. Es hat sich gezeigt, dass eine konstante Funktion der Form $\zeta = konst.$ (10, 20 oder 30) anderen funktionalen Verläufen aus Gründen der Schlichtheit vorzuziehen ist. Ein Unterschied in den F_1 -Werten war auch bei sehr komplexen logarithmischen, Exponential- oder Winkelfunktionen höherer Ordnung nicht erkennbar und so sollte nach dem *lex parsimoniae* die einfachste Struktur gewählt werden. Der glückliche Umstand, dass die Sensibilitäten σ und Phonemkettenlängen λ für alle Maximalwerte F_1 für einen Korpus über alle Perioden hinweg gleich bleiben, ermöglicht eine übersichtliche Darstellung. F_1 kann in Abhängigkeit von der Anzahl der Perioden und der Lexikonfunktion²⁰ dargestellt werden. So kann man anhand der gegebenen Daten beobachten, dass die Segmentierungsleistung einen signifikanten Zuwachs bis zu 10 Prozent mit der zweiten Periode verzeichnet, danach jedoch nur noch sehr leicht wächst oder gar konstant bleibt. Dabei gleichen sich Fehlsegmentierungen und richtige Segmentierungsleistungen aus. Der Zuwachs erklärt sich durch ein starkes Wachsen der Vollständigkeit, also dadurch, dass mehr segmentiert wird. Die Lexikonfunktion nimmt darauf erst bei großen Korpora entscheidend Einfluss und dann auch nur hinsichtlich ihrer Robustheit, d.h., ein weiteres Spektrum an Sensibilitäten und Phonemkettenlängen ist für hohe F_1 -Werte möglich. Bei einem Vergleich zwischen dem Wortzuwachs im Paralexikon und F_1 konnte der Einfluss des Paralexikons auf die Segmentierungsleistung begründet werden. Bei kleineren Korpora bildet die Wortanzahl in etwa den Verlauf von F_1 nach. Bei größeren Korpora verliert sich diese Symmetrie. Bei ihnen treten die logisch widersprüchlichen Zustände von Sensibilität und Phonemkettenlänge als quasi-chaotische Zustände hervor. Der Rückgang der Segmentierungsleistung ab der dritten Periode liegt in dem *Embeddedness*-Problem der Morpheme begründet. Zunächst konnte festgestellt werden, dass ein großes Paralexikon diese Problemquelle abmildern würde, wenn es denn die morphemischen Varianten der Lexeme enthält.

Die genaue Form der Lexikonfunktion ζ ist schließlich weiter ergründet worden. Hier beschränkt sich die Erforschung allerdings auf eine Korpusgröße von 10.000 Zeichen. Nimmt man in jeder Periode unterschiedliche Korpora, lässt sich ein „Verzögerungseffekt“ nachweisen. Die Abnahme der F_1 -Werte erfolgt nicht in der zweiten, sondern in der vierten Periode. Dies ist unmittelbar für das Problem der *Embeddedness* relevant, weil offensichtlich zuerst weitere Einträge ins Paralexikon vorgenommen werden bevor die Zersplitterung der Wörter einsetzt. Somit sollte die Frage erforscht werden, wie viele Einträge noch vorgenommen werden können und ob der Beginn der Rekursion (μ) entscheidenden Einfluss auf die Segmentierung hat.

²⁰Die Lexikonfunktion, wenn konstant gehalten, ist gleich der Lexikongröße pro Periode bei $\psi \geq 5000$

Ein Vergleich verschiedener Paralexika, welche unter Manipulation der genannten Einflüsse, verglichen wurden, hat sowohl die optimale Lexikonfunktion als auch den richtigen Einsatz der Rekursion in Abhängigkeit von der Lexikonfunktion ζ bestimmt. So sollte bei einer Lexikonfunktion $\zeta = 10$ die Rekursion bei $\mu = 25$ Perioden beginnen, weil der Wortzuwachs danach unsignifikant ist (*Grenzwert* = 107). Für diese Kombinationen lassen sich abermals die maximalen F_1 -Werte durch eine Funktion $\mathit{argmax}(\mathit{LexFunc})$ ²¹ beschreiben. Bei einem Rekursionanfang nach 25 Perioden beginnen sich die F_1 -Werte der einzelnen Lexikonfunktionen ζ_i der $\mathit{argmax}(\mathit{LexFunc})$ anzuschmiegen. Die chaotische Zersetzung des Lautstroms hebt sich damit immer weiter auf, je mehr die Rekursion zeitlich verschoben wird. Für ein im Unendlichen beginnendes μ gilt somit

$$\mathit{argmax}(\mathit{LexFunc}) = \mathit{LexFunc}_\infty. \quad (6.10)$$

Schließlich sind die Eigenschaften des Paralexikons herausgestellt wurden. Vor dem Hintergrund, welche Informationen zuerst kognitiv verarbeitet werden müssten, konnte gezeigt werden, dass das erstellte Paralexikon als Ausgangspunkt ausreichend ist, um weitere Gesetzmäßigkeiten, die dem gleichen Prinzip unterliegen wie die Rekursionsmechanismen, genügt. Aus dem Wortmaterial des Paralexikons können phonotaktische und metrische Gesetzmäßigkeiten vorhergesagt werden. Diese Gesetzmäßigkeiten würden im sich anschließenden Fortgang der Segmentierungen genutzt werden, um die verbleibenden Segmentierungen vorzunehmen und die falschen Segmentierungen zu beseitigen.

²¹ $\mathit{LexFunc}$ umfasst alle möglichen Lexikonfunktionen ζ_i .

7 Schlussbetrachtung

7.1 Beitrag zur Anlage-Umwelt-Debatte

Das anfangs formulierte Ziel der Arbeit war, die nativistische Annahme der Reizarmut des Sprachsignals zu widerlegen. Das Argument der Reizarmut unterstellt, dass es keinen direkten Zusammenhang, keinen Mechanismus, zwischen dem Sprachinput und der erlernten Sprache, dem Output, geben kann, der erklärt wie aus dem Input der Output unter Verwendung allgemeiner kognitiver Gesetzmäßigkeiten, entstehen könne. Für den Bereich der Sprachsegmentierung kann diese Annahme nun nicht mehr aufrechterhalten werden.¹ Die Informationen, die im Sprachsignal statistisch kodiert sind, können eine Beziehung zwischen Input und Output herstellen. Damit verliert das Argument der Reizarmut seine Attraktivität, weil es zumindest theoretisch einen Mechanismus gibt, der es erlaubt, einen Sprachstrom zufriedenstellend zu segmentieren, ohne auf voreingestellte sprachspezifische Informationen zurückzugreifen und der die allgemeinen kognitiven Fähigkeiten der Sprachlerner dieses Alters berücksichtigt. Somit lässt sich kurz festhalten: Für die Sprachsegmentierung ist Chomskys Argument der Reizarmut nicht zutreffend, weil das Sprachsignal selbst so viele Informationen zur Verfügung stellt, wie zur Segmentierung eines Sprachstroms benötigt werden.²

Dieses Ergebnis stärkt also die Einflüsse der Umwelt auf die Sprachentwicklung. Wie zu Beginn erläutert, lässt das Ergebnis nicht die Aussage zu, dass der Sprachlerner den

¹Natürlich sagt die Arbeit nichts über das Argument in Anwendung auf die Syntax aus. Es ermutigt aber nach den versteckten Informationen und selbstorganisatorischen Prinzipien auch dort zu suchen.

²Chomsky ist in seinen diesbezüglichen Formulierungen sehr explizit, sodass hier die Bezugnahme auf seine Argumentation leicht fällt. Er beschreibt einen Wissenschaftler S, der unvoreingenommen zwischen zwei Ansätzen R und E wählen kann. R und E stehen für die Wissenschaftsparadigmen Empirismus und Rationalismus, wobei E die unbeschriebene Tafel, unter anderem mit der Fähigkeit

die Reaktionswahrscheinlichkeiten durch periphere Verstärkungsmechanismen zu modifizieren, die durch Rekurs auf den Stimulusraum definiert werden, ... (Chomsky, 1977, 175)

repräsentiert. Wie gewohnt, wird S mit einer Reihe von Sprachbeispielen konfrontiert, die ihm zeigen sollen, dass die Möglichkeiten in E nicht ausreichend sein können, die Strukturabhängigkeiten (*structure dependence*) zu erkennen und er deshalb auf R ausweichen muss:

Jedes dieser Beispiele liefert zusätzliche Belege für die Schlussfolgerung, daß ein Vorgehen nach E, in dem Segmentierung, Klassifikation, Analogie und Generalisierung eine grundlegende Rolle spielen, prinzipiell verfehlt ist. (Chomsky, 1977, 180f)

Eine entsprechende Widerlegung anhand der gegebenen Beispiele bei Zulassung von statistischen Informationen findet sich in Lewis und Elman (2001). An dieser Stelle kann man nur anmerken, dass Chomsky die Segmentierung aus der Aufzählung streichen sollte.

7 Schlussbetrachtung

vorgestellten Algorithmus benutzt, aber die Simulation entkräftet das nativistische Argument doch entscheidend. Die Schwäche des Arguments liegt ja darin, dass es eine Aussage über die Natur des Input-Output-Verhältnisses trifft, ohne selbst eine kritische Evaluierung der sehr generellen Annahmen des Input (*PLD*³) vorzunehmen (Pereira, 2000, 1243). Es basiert auf einer vielleicht anfangs intuitiv verständlichen und einleuchtenden, aber eben auf einer nicht ausreichend überprüften Annahme. Es ist möglich, diese Annahme als unzutreffend darzustellen, indem man zeigt, dass die empirische Prämisse (siehe Kap. 1.3) nicht generell wahr sein kann. Da die empirische Prämisse sehr allgemein ist – die Aussage, dass nicht genügend Informationen vorhanden sind, spezifiziert nicht die Art der Information – ist es auch möglich, die Annahme in ihrer breiten allgemeinen Aufstellung zu kritisieren. Schließlich umfasst die Annahme hier die Information im Allgemeinen und diese wiederum eben nicht nur offensichtliche, unilaterale und lineare Ausprägungen. Es muss also zulässig sein, auch alle Information einzubeziehen.

Wer das Datenproblem des *PLD* nicht schlüssig erklären kann, verfehlt das Ziel, die Hypothese der Reizarmut auf eine fundierte, plausible Basis zu stellen. Fehlt diese Basis, müssen auch die Versuche, den *LAD*⁴ zu rechtfertigen, letztendlich fehl schlagen. Die Literatur der generativen Linguistik beinhaltet aner kennenswerte Leistungen von Überlegungen über die genaue Ausgestaltung der Rahmenbedingungen innerhalb des *LAD*. Man kann kaum mehr Lehrbücher finden, in denen nicht zahlreiche Beispiele angeführt werden (vgl. Smith, 2004, 39). Selbst in Fachaufsätzen findet man dieses Muster durchgängig. Zum Beispiel mögen Crain und Pietroski (2001) eine sehr einleuchtende Herleitung für die Notwendigkeit von Rahmenbedingungen skizzieren. Der entscheidende Punkt liegt jedoch anderswo. Was wäre, wenn der *PLD* die notwendigen Informationen enthielte? Wozu würde man dann noch einem *LAD* benötigen? Sollte das Induktionsproblem für die Syntax ebenso gelöst werden können, wie nun im Ansatz für die Wortsegmentierung geschehen, dann werden schlagartig alle Argumente für einen *LAD* überflüssig, weil mit den Eigenschaften des Input der Output (Sprachkompetenz) erklärt wäre. In derartigen Disputen kann es doch nicht ausreichen, das Datenproblem mit (teilweise selbst konstruierten) Einzelbeispielen zu belegen und durch einfache Behauptungen eine Begründung dergleichen herzuleiten.

And it seems unlikely that such overgeneration would ever be noticed by adults, given the relative frequency of the sentence/contexts that would manifest the error. (Crain und Pietroski, 2001, 15f)

Hier muss sich ein Wissenschaftler fragen, welchen genauen Wert denn nun die relative Häufigkeit (*relative frequency*) hat. Wurde sie gemessen? Wo ist sie denn gegeben (*given*) wie die Autoren behaupten? Dies ist hier nur ein exemplarisches Beispiel, eine Art Muster, nach dem die Argumentation in den generativen Literaturen zu diesem Thema typischerweise aufgebaut wird. Offenbar werden Behauptungen als Begründungen angeführt. Der empirische Beweis wäre in diesem Fall recht einfach zu bewerkstelligen. Mit repräsentativen Korpora könnte die Aussage auf ein solides Fundament gestellt werden. Es mag dabei auch durchaus richtig sein, dass bestimmte Sätze auch in einem repräsen-

³Primary Linguistic Data

⁴Language Acquisition Device

tativen Korpus nicht vorhanden sind. Aber stimmt dies auch für einen Strukturvergleich der Syntax, für ein äußerlich nicht offensichtliches Muster (Seidenberg et al., 2002, 554) und für pragmatische oder semantische Symbolkonzepte, wie sie Tomasellos Kritik an Pinkers Buch *The Language Instinct* geäußert werden (vgl. Tomasello, 1995, 149)? Dies ist sicherlich nicht der Fall. Es überrascht daher kaum mehr, dass Korpusstudien, welche den Input des Kindes untersuchen und diesen mit dem von Erwachsenen vergleichen, die Annahmen im generativen Paradigma anzweifeln müssen (Lewis und Elman (2001); Reali und Christiansen (2003); vgl. Behrens (2006); Sampson und Postal (2005)). Fodor und Crowther (2002) versuchen sich zwar an empirischen Korpusdaten der CHILDES Datenbank, jedoch wird dies nicht dem Namen gerecht, da sie nur einen Fall aus CHILDES – „Adam“ – mit entsprechenden Daten der *Wall-Street-Journals*-Datenbank beziehungsweise der *Switchboard*-Telefonkonversationen vergleichen. Fodor und Crowther (2002) geht es in erster Linie um eine Kritik an der Feststellung von Pullum und Scholz (2002).⁵

a construction inaccessible to infants during the language acquisition process must be rare enough that it will be almost entirely absent from corpora of text quite generally. (Pullum und Scholz, 2002, 21)⁶

Für Fodor und Crowther (2002) gibt es Konstruktionen im Sprachwissen der Kinder, dessen Muster nicht aus dem Stimulusmaterial gewonnen werden können. Dazu versuchen die nativistisch geprägten Autoren zunächst eine Verschiebung zwischen dem Zeitpunkt der Aufnahme des Inputmaterials und dem beobachtbaren Sprachwissen des Kindes zu diskutieren. Sie bestreiten, dass beispielsweise ein zweijähriges Kind Sätze der Form „The chair is not strong enough to stand on (*it)“ und „He brought the chair in to stand on (it)“⁷ (Fodor und Crowther, 2002, 113f) gehört hat; die Kompetenz dieses Kindes enthielte aber schon das syntaktische Wissen dafür. Geschickt wird mit dem Alter des Kindes argumentiert, welches wenn nach unten korrigiert ja schon bestimmte Satzkonstruktionen kennt, obwohl diese noch nicht im Korpusmaterial für dieses Alter nachweisbar sind. Unabhängig davon, ob die Alterskorrekturen richtig sind oder überhaupt relevant sein werden, Fodor und Crowther (2002) sind ein typisches Beispiel für eine Einzelfallbetrachtung, atypisch aber insofern, dass die Autoren empirische Korpusdaten anfügen.⁸ Diese Einzelfälle können doch aber nur dann richtig beurteilt werden, wenn die Quantität derselben in ein rechtes Licht gerückt wird. Es könnte sich doch herausstellen, dass bei einer repräsentativen Stichprobe, tatsächlich zutreffende syntaktische Muster festzustellen sind oder dass bestimmte Konstruktionen zwar grammatikalisch sind, aber tatsächlich nicht verwendet werden. Sampson (2002) unternimmt einen solchen Versuch und kann für eine Reihe von Phänomenen ausreichende Quantitäten nachweisen.

Der Grund, ein umfangreiches Theoriegebäude, wie die generative Grammatik, auf einer nur oberflächlich untersuchten Annahme, wie dem Argument der Reizarmut, auf-

⁵Weiterhin gibt es Inputdaten, die irreführend (S. 117) oder mehrdeutig (S. 119) sind.

⁶Anders als der hier gewählte Ansatz versuchen Pullum und Scholz (2002) zu zeigen, dass durchaus *negative evidence* in Korpora zu finden ist.

⁷als Beispiel für Infinitivkomplementierung.

⁸Die Aussagefähigkeit der Untersuchung ist allerdings sehr begrenzt.

7 Schlussbetrachtung

zubauen, lässt sich auf das gleiche Motiv zurückführen, wie eine Stimulusarmut überhaupt zu konstruieren: Nämlich der Notwendigkeit die fehlenden Daten kompensieren zu müssen. Das Argument der Reizarmut erspart eine breit angelegte Untersuchung des Input für andere Weltsprachen⁹ hinsichtlich verschieden aufbereiteter Informationen (semantisch, pragmatische Kontextualität oder auch nichtlineare Zusammenhänge in den Symbolanordnungen). Für die generative Grammatik selbst existiert ebenso das Problem der mangelnden Daten. Es lässt sich kein empirischer Beweis für ihre Existenz anführen.¹⁰ Ihre Existenz wird in Teilen auf das Argument der Reizarmut zurückgeführt. In anderen Teilen geht sie auf sehr ähnliche Annahmen zurück, wovon keine eine klare empirische Datenlage aufweist (Tomasello, 1995). Unter diese Annahmen fallen die Modularität,¹¹ die Sprachuniversalien, die Einzigartigkeit der Sprache beim Menschen,

⁹Hiermit ist vorrangig nicht der indogermanische Sprachstamm gemeint.

¹⁰Levelt und Peter (1989), Jackendoff (1996) oder Pinker (1994) gelten als Arbeiten, in denen versucht wird, experimentelle Befunde mit der Chomskyschen Auffassung einer linguistischen Theorie in Einklang zu bringen (vgl. z.B. Lieberman, 2000, 7). Allerdings handelt es sich hierbei keinesfalls um eine empirische Basis, da immer nur die Ergebnisse für die generative Theorie hinterinterpretiert werden, ohne aber tatsächlich die Annahmen zu hinterfragen und ohne zu wissen, was Ursache und Wirkung ist, wie es von Calvin und Ojemann (2000, 296) so treffend formuliert wurde (siehe sich anschließende Fußnote). Ein weiteres Beispiel ist die Arbeit von Bickerton (1991), in der er die Evolution der Sprache als ein plötzliches Ereignis beschreibt, welches die einzigartige Syntax durch neuronale Verschaltungen hervorbrachte. Es lassen sich aber nur äußerliche Beobachtungen, die auf vielfältige Weise interpretierbar sind, ausmachen. Verwendbare Daten in einem strikten empirischen Sinne fehlen (vgl. hierzu auch Lieberman, 2000, 127). In Crain und Pietroski (2002, 168ff) werden empirische Details angeführt,

that illustrate the tight relation between the details of linguistic theory and the most impressive poverty-of-stimulus arguments (Crain und Pietroski, 2002, 167),

wobei

the empirical question is whether the proposed linguistic principles are fruitful in explaining the range of facts that natural languages exhibit (Crain und Pietroski, 2002, 166).

Diese empirischen Details sind nichts weiter als selbst erdachte Satzkonstruktionen, die grammatikalisch so gewählt sind, dass sie angeblich nur mit den zu Grunde liegenden UG-Prinzipien erklärt werden können. Kein Wort wird darüber verloren, wie es das Kind denn tatsächlich anstellt und welche Prozesse dort genau ablaufen, sodass man tatsächlich von empirischen Daten sprechen könnte. *Empirical details* kann doch nicht bedeuten, dass hier zwar eine tatsächlich in der Realität anzutreffende Satzkonstruktion gewählt und alles andere mit reinen Spekulationen hergeleitet wird. Auch diese Spekulationen müssen empirisch belegt werden. Nur das würde einer Behauptung über empirische Details gerecht werden.

¹¹Sie gründet sich auf Studien zu Savants und zu *Specific Language Impairment (SLI)*, zur Einzigartigkeit der Syntax und zu lokalisierbaren Gehirnregionen. Letzteres unterliegt leider oft einem Mißverständnis und ist nicht gut geeignet, den Disput wie viel Umwelt und Anlage im Spracherwerb notwendig ist, für sich zu entscheiden. Man möchte nämlich die Region im Gehirn finden, in welcher der *LAD* repräsentiert wird. Auf dieser Suche nach der Lokalisation der genetisch enkodierten Sprachregion, ziehen Forscher beider Parteien oft die Arbeiten aus der Psycholinguistik heran. Sie versuchen zu zeigen, welche Verhaltensweisen bei defekten Hirnfunktionen, beispielsweise bei Broca- oder Wernicke-Aphasien, auftreten und schließen nun wiederum auf die Eigenschaften der Existenz oder der Nichtexistenz des Sprachmoduls rück. Das Problem dabei formulieren Calvin und Ojemann sehr anschaulich. Sie sagen, dass diese Vorgehensweise Ursache und Zusammenhang von Beobachtungen verwechselt.

die Beobachtungen aus dem Erwerb der Zeichensprache und die Diskussion der Pidgins und Kreolen (Pinker, 1994). Bei (Pinker, 1994)¹² findet sich auch die Behauptung, dass der Mensch über keinen entsprechenden Lernmechanismus verfügt. Diese Aussage hängt allerdings direkt mit der Qualität des Input zusammen und lässt sich deshalb gut mit dem Argument der Reizarmut vereinen, denn beide Argumente unterstellen, dass die Eigenschaften des Inputs vollkommen bekannt sind. Für alle anderen eben angeführten Argumente kann diese Arbeit nicht den ihnen gebührenden Raum geben, obwohl man diese Annahmen zur Anlage-Umwelt-Debatte zuschreiben könnte. Weil sich die Arbeit in erster Linie auf die Wiederlegung des Arguments der Reizarmut konzentriert, soll nicht der Eindruck erweckt werden, dass auch die anderen Annahmen der generativen Idee von den Ergebnissen dieser Arbeit betroffen sind.

Um die Annahme des Arguments der Reizarmut zu spezifizieren, müsste man angeben, warum nur bestimmte Informationen vom Kind verwendet werden können. Das setzt aber voraus, dass man experimentell bestimmt hat, dass diese (und keine anderen Informationen) verarbeitet werden. Genau dieses Verfahren sollte aber letzten Endes mit dem Argument der Reizarmut umgangen werden, weil die Erhebung derartiger Daten sicher ebenso viel Aufwand oder Unmöglichkeit bedeutet, wie die Existenz des Sprachmoduls direkt zu beweisen. Da nun in der vorgeführten Simulation nicht nur die Bedingung einer generellen Annahme, sondern auch schon eine weitere, die Berücksichtigung der Fähigkeiten eines acht Monate alten Kindes, eingeflossen sind, würde selbst eine weitere Spezifizierung der Annahme ohnehin nicht weiter helfen.

Das Argument der Reizarmut baut auf der Annahme auf, dass es keinen erkennbaren Zusammenhang zwischen der zugeführten Sprache und der Sprachkompetenz gibt. Das wurde widerlegt. Es gibt einen derartigen Mechanismus. Der Nativist müsste jetzt beweisen, dass der Sprachlerner die vorgestellte Vorgehensweise eindeutig nicht nutzen kann, möchte er seine Annahme weiter aufrechterhalten. Dies dürfte jedoch schwerfallen, weil mindestens zwei Argumente dafür sprechen, die zusammen betrachtet, das Argument der Reizarmut als schwache Ausrede erscheinen lassen. Erstens kommt die Frage auf, warum der Mensch diesen doch recht ressourcenintensiven Mechanismus beherrscht und

Der 'Sprachkortex' ist nichts anderes als ein Stück Kortex, der neben anderen Funktionen offensichtlich auch Sprachfunktionen unterstützt. Wenn man seine Funktion durch das definiert, was ohne ihn nicht mehr klappt, verwechselt man einen Zusammenhang mit einer Ursache. Der Neurologe F.M.R. Walshe hat schon 1947 betont, daß es wahrscheinlich zu einer Fehlidentifikation kommt, wenn man die Funktion eines Gehirnbereichs anhand der Symptome definiert, die nach einer Schädigung des Bereichs auftreten. Er illustrierte das mit dem Beispiel eines weggebrochenen Zahns in einem Autogetriebe, was sich etwa in dem Symptom äußert, daß man bei jeder Umdrehung einen Klacks hört. (Calvin und Ojemann, 2000, 296)

Nativisten sowie Antinativisten sollten also sehr vorsichtig mit den gewonnenen Daten aus Läsionsstudien sein. Die intervenierenden Drittvariablen sind unbekannt und somit kann aus korrelierenden Eigenschaften kein sicherer Schluss gezogen werden. Zum jetzigen Stand der Forschung können neuroatomare Studien noch keine unstrittigen Ergebnisse für die Anlage-Umwelt-Debatte bereit stellen.

Vgl. zu dieser Thematik auch Ridley (2003, 124)

¹²sicherlich zurückgehend auf Chomsky (1977).

7 Schlussbetrachtung

entwickelt hätte, wenn er ihn nicht benötigt und zweitens steht keine annähernd plausible Theorie zur Verfügung, welche einen derart engen Bezug zu empirischen Daten und zur Einfachheit bietet.

Gegen das erste Argument wäre einzuwenden, dass die Natur eine ganze Reihe an opulenten Mechanismen hervorgebracht hat – die Plastizität des Gehirns durch unzählige Mehrfachverschaltungen und Doppelspeicherungen, kinderlose Mütter bei Herdentieren, Ersatzfunktionen von Genen aktivierten Proteinen, u.v.m. – oder in eine evolutionäre Sackgasse gelaufen ist. Ein gewisser Überfluss ist zugegebenermaßen nützlich, deshalb auch evolutionsstabil und möglicherweise gibt es auch kognitive Fähigkeiten oder Körperteile, die sich im Laufe der Evolution zuerst als vorteilhaft und später als überholt herausgestellt haben. Warum aber sollte dieses Gegenargument plausibler sein als die tatsächliche Nutzung eines Mechanismus, deren Fähigkeiten experimentell nachweisbar sind, zumal ein alternativer Vorschlag fehlt. Entscheidend dabei ist doch, dass es derzeit keine schlüssigere Erklärung gibt, denn es ist eben nicht so, dass die Stimulusarmut eine alternative Erklärung bieten würde, wie beispielsweise Crain und Pietroski (2002, 167) oder Lasnik und Uriagereka (2002, 150) behaupten.

For the empiricist approach to be considered a scientific alternative, it owes us the very *first step*: to show precisely how, starting *with no assumptions* about the organization of the grammar . . . is inferred by a learner on the basis of mere *positive data*. Meanwhile it is pointless to engage in mathematical analyses of part of the data. What could that possibly decide if there is only one theory under scrutiny? (Lasnik und Uriagereka, 2002, 150)¹³

Zunächst wird hier eine falsche Behauptung aufgestellt, die sehr stark an das Extrem der *tabula rasa* des Empirismus erinnert. Natürlich werden Annahmen über die Organisation der Grammatik getroffen. Diese Annahmen sind aber nicht speziell auf die Grammatik allein ausgerichtet, sondern es handelt sich um allgemeine Fähigkeiten, mit denen man auch andere Phänomene des menschlichen Geistes erklären kann. Zweitens ist *positive data* für den statistischen Ansatz grundsätzlich hinreichend (Lewis und Elman, 2001). Es kommt auf die Präsentation der Datenfülle im Zeitverlauf an (Newport (1990); Rohde und Plaut (1999)). Darum ist es vorteilhaft, dass sich die Gehirnstrukturen des Kindes sehr langsam entwickeln und daher auch nur sehr restriktiv bestimmte Informationen aufnehmen können. Dies liegt jedoch nicht an den Voreinstellungen einer vererbten Sprachvorrichtung, sondern an der Aufnahmekapazität des bis dato entwickelten Gehirns. Andere Informationsaufnahmeprozesse, wie das Sehen, unterliegen natürlich genau den gleichen Einschränkungen. Drittens wäre doch auch für den Rationalisten der *erste Schritt* zunächst einmal die angebliche Unzulänglichkeit des Inputs zu überprüfen und empirisch zu belegen. Ermächtigt denn eine höchst zweifelhafte Hypothese dazu, diesen ersten Schritt auszulassen und dann in ebenso höchst abstrakter Form generative Prinzipien aufzustellen? Lasnik und Uriagereka (2002) werfen doch wohl nicht ernsthaft dem Empiristen vor, diesen ersten Schritt nicht zu tun – die Hauptkritik am Argument der Reizarmut – ohne sich selbst diese Frage zu stellen! Wo ist denn dieser erste Schritt (*first step*, vgl. obiges Zitat von Lasnik und Uriagereka (2002, 150)) im rationalistischen Ansatz? Er wurde mit dem Argument der Reizarmut umgangen.

¹³kursive Hervorhebungen hinzugefügt.

Das Argument der Reizarmut wiederum soll der Idee des Sprachmoduls erst den Weg ebnen. Wenn die Annahme des Arguments aber falsch, somit auch der Schluss falsch ist, kann die darauf aufbauende alternative Erklärung aber auch nicht plausibel sein. Selbst wenn man diesen Gedanken nicht teilen möchte und sich zu einer Vermutung im Sinne eines Gedankenexperiments hinreißen lässt, so bleibt doch außer Frage, dass die Spezifikationen und möglichen Einstellungen in einem chomskyschen Sprachmodul selbst keine empirische Basis haben, sondern eher auf den Gesetzmäßigkeiten der formalen Logik beruhen und den Grundsatz der Sparsamkeit (*parsimony*) auch im jüngeren minimalistischen Programm offensichtlich stark verletzen.¹⁴

Diese Herangehensweise lässt sich oft auch in den Verhaltenswissenschaften entdecken, wenn aufgrund der Datenlage die Interpretationsbreite größer wird. Es ist erstaunlich zu beobachten, wie sich auch experimentelle Ergebnisse recht einseitig in ein bestimmtes, womöglich als vorherrschend erachtetes, Paradigma hineinpressen lassen.

Finally, infants' close attention to and retention of sound patterns that occur frequently in the input suggests that this aspect of language acquisition can be characterized as a form of innately guided learning: that is, infants are primed to learn some things very rapidly and to learn them in a particular way. (Jusczyk und Hohne, 1997, 1985)

Gerade die Experimente in Jusczyk und Hohne (1997), die der zitierten Quelle zu Grunde liegen, geben sehr wenig Anlass für eine derartige Vermutung. Das Gegenteil könnte man genauso annehmen: Habituation und Dishabituation haben dazu geführt, dass sich ganz bestimmte Signalausschnitte aufgrund ihrer Vorkommenshäufigkeit neural besser verdrahtet haben als andere (vgl. Breitenstein und Knecht, 2003). Dieser Prozess verläuft deshalb im Säuglingsalter schneller, weil noch keine oder nur wenige synaptische Verbindungen dergestalt verschaltet sind. Würden bereits Verbindungen bestehen, benötigt die Synapsenbildung aufgrund der räumlichen Enge mehr Zeit. Auf diese Weise kann man den endlosen Disput *ad absurdum* führen, wobei am Ende kein Erkenntnisgewinn steht. Marcus (2005)¹⁵ fordert vielleicht deswegen mehr Kooperation:

Im Wechselspiel von Anlage und Umwelt ist keine von beiden „besser“. Wir sollten nicht fragen, was Vorrang hat, sondern das Wie betrachten: Wie wirken Gene mit der Umwelt zusammen, sodass der menschliche Geist entsteht? (Marcus, 2005, 17)

Dieser Ansatz mag auch recht vielversprechend klingen, aber wie sich auch sehr schnell bei der Lektüre des Buches (Marcus, 2005) herausstellt, ist das genaue Zusammenspiel von Anlage und Umwelt doch nur möglich, wenn man weiß, was denn die Gene und was genau die Umwelt beiträgt. Damit bleibt die Ausgangsfrage unveränderbar bestehen. Wie viel trägt die Umwelt und wie viel die Anlage bei? Das *Wie viel* enthält bereits das *Was* und daraus lässt sich schließlich das angestrebte Ziel eines *Wie's* entwickeln. Umgekehrt macht es doch wohl kaum Sinn zu fragen, wie eine genetische Kodierung mit der Umwelt ein Verhalten hervorbringt, wenn man nur vermuten kann, was die Umwelt und was davon der Genpool auslöst. Eine genaue Erklärung wird man so nicht finden können. In diesem Sinne plädiert auch Pinker für eine kontroverse Fortsetzung des Streites:

¹⁴vgl. Putnam (1971).

¹⁵Dieser Gedanke ist auch bei Bronfenbrenner und Ceci (1993, 313) oder Wachs (1993) schon anzutreffen.

7 Schlussbetrachtung

some of the most provocative discoveries in twentieth-century psychology would have been impossible if there had not been a concerted effort to distinguish nature and nurture in human development. (Pinker, 2004, 10)

Es bleibt nun zu hoffen, dass sich der Disput zwischen Nativismus und Empirismus zuerst im Kräftegleichgewicht der Argumente löst, die hoffentlich als Reflektion der Wahrheit das *Wie viel* und das *Was* bekanntgeben. Sodann ergibt sich daraus ihre natürliche Verschmelzung ineinander. Bis dahin müssen Wege wie der hier beschriebene beschritten und die Kontroverse weitergeführt werden. Dies soll jedoch nicht bedeuten, dass die künstliche Trennung zwischen Empirismus und Rationalismus in Extrempositionen vertreten werden muss. Die Frage nach dem *Was* sollte natürlich gerade aus einer Kooperation von empiristischen und rationalistischen Ideen hervorgehen, die man aber nicht mit Anlage beziehungsweise Umwelt gleichsetzen darf.

7.2 Anknüpfungspunkte für weitergehende Forschung

Der oben beschriebene Algorithmus setzt voraus, dass diskrete Einheiten aus dem Sprachsignal gefiltert werden können und dass sich diese Einheiten auf eine sehr überschaubare Zahl von unter 40 Phonemen beschränken lassen. Zudem sind alle Repräsentanten einer Einheit völlig identisch. Allophonische Varianten treten also nicht auf. Dies ist eine starke Abstraktion der Wirklichkeit. In der Realität ist das akustische Sprachsignal so komplex, dass es bisher nicht gelungen ist, die Erkennungsleistung auf einzelne Phoneme automatisiert von Maschinen durchführen zu lassen (Ingram (2007, 93); Kewley-Port (1999, 336); Liberman et al. (1967)). Das Phonem mag ein sehr künstliches Produkt der linguistischen Analyse sein, das im Signal nicht nachweisbar ist. Dennoch ist es plausibel, eine abstrakte Einheit in dieser oder ähnlicher Form anzunehmen,¹⁶ welche die Schnittmenge aller Allophone repräsentiert (Caplan et al. (1999, 1496); Lahiri und Marslen-Wilson (1991)).¹⁷

Die Schnittmenge muss immer eindeutig ungleich der Schnittmenge aller anderen Phoneme sein.¹⁸ Das bedeutet, dass es wiederum (wie in Kapitel 1.4 dargelegt) mindestens eine gemeinsame Eigenschaft gibt, auf der sich ein Ordnungssystem gründen kann (vgl. Bischof, 1995, 12ff). Im Umkehrschluss heißt es: Gäbe es diese Schnittmenge nicht, könnte sich kein Ordnungssystem bilden und das Gehirn könnte auch keine abstrakten Repräsentationen von Lauten bilden. Dabei ist zunächst unwichtig, ob die Repräsentationen tatsächlich dem Konstrukt eines Phonems entsprechen, Kombinationen davon bilden oder weniger Informationen als ein Phonem beinhalten. Ohne Abstraktionsvermögen

¹⁶vgl. auch Cutler und Clifton (2000).

¹⁷Diese Schnittmenge definiert in etwa das, was Morgan (1994, 390) die *external individuality* nennt:

at a given level of analysis, the behavior of a group of elements that constitutes a unit is indistinguishable from the behavior of a single element.

¹⁸vgl. Remez (1994, 148f).

müsste jede bedeutungsunterscheidende Lautsequenz einzeln repräsentiert werden. Das ist schon allein deshalb unmöglich, weil ein gemeinsames Sprachsystem nicht entstehen könnte. Jedes Mitglied einer Sprachgemeinschaft würde für jeden Laut eines jeden Mitgliedes eine neue Repräsentation anlegen müssen, da sonst der Laut nicht verstanden werden kann. Kommt ein Fremder hinzu, könnte er nicht verstanden werden, weil es noch keine Repräsentationen seiner Laute gibt. Zudem kann man nicht erklären, wie die Repräsentationen zu bedeutungstragenden Einheiten kombiniert werden, wenn man sie nicht als „gleich“ klassifizieren kann. Es gäbe somit keine kombinatorische Strategie, nach der man die Laute in verschiedenen Anordnungen zu neuen Signifikaten zusammensetzt. Anstelle dessen gelangt man zu einem unendlichen Regress, weil immer nur neue Repräsentationen angelegt werden, ohne dass das Gehirn eine Chance hätte, Beziehungen zu den einzelnen Repräsentationen herzustellen. Da dies aber nicht der Beobachtung entspricht, muss auf theoretischer Ebene der Schluss erfolgen, dass es abstrakte Einheiten gibt. Wenn diese aber existieren, müssen sie sich auch nach einem Ordnungssystem verhalten. Jedem Ordnungssystem liegt mindestens ein gemeinsames Merkmal zu Grunde. Es gilt also, dieses Merkmal oder diese Merkmale zu finden.¹⁹

Auf diese Weise sind die gleichen Voraussetzungen gegeben, wie bei der Problemstellung dieser Arbeit. Es wäre also interessant, das Prinzip des Algorithmus auf das physische Sprachsignal zu übertragen, ähnlich wie von Hidden-Markoff-Ketten anfangs nur zur Phonemidentifizierung verwendet und nun auch in nachgelagerten Problembereichen Anwendung findet, würde der hiesige Algorithmus seine Anwendungsbereiche in umgekehrter Reihenfolge ausdehnen können. Die Phonemkettenlänge würde dem zeitlichen Verlauf entsprechen und so könnte man die Informationen verschiedener Spektren bei gegebener Zeit vergleichen. Da die Spektren jedoch nicht übereinstimmen, wie die Phonemrepräsentationen im hiesigen Fall, müssten eine weitere oder mehrere primäre Variablen die Abweichungen der einzelnen Spektren voneinander messen. Mehrere Variablen deshalb, weil das Frequenzspektrum des akustischen Signals durch verschiedene Eigenschaften (Tonhöhe, Länge, Intensität) bestimmt wird. Die Sensibilität würde hier definieren, wie viel Abweichung erlaubt ist, um eine Eigenschaft einer bestimmten Schnittmenge oder einem Rauschen zu zuschreiben. Weiterhin ist das hauptsächliche Charakteristikum der gesprochenen Sprache ihre Unvollständigkeit. Verschlucken von Endungen, Koartikulationen, Husten, Räuspern oder Lachen verändern das Klangbild entscheidend. Die Information ist unvollständig. Dies schlägt sich dann natürlich auf die intern gebildete Statistik der Verteilungen nieder. Hier muss die Frage beantwortet werden, wie stark diese Einflüsse sind.

Der Algorithmus gibt in seiner derzeitigen Implementierung eine recht zuverlässige Wortliste aus. Die Muster, die in dieser Liste enthalten sind, wurden in dieser Arbeit nur theoretisch beschrieben. Auch hier könnte man die rhythmischen und phonotaktischen Muster abermals mit dem gleichen Algorithmus extrahieren. So entstünde eine Rekursion innerhalb weiterer Rekursionen, weil auch die in der Liste gewonnenen Informationen

¹⁹Als kritischen Gegenpol zu dieser Sichtweise siehe beispielsweise Pisoni und Levi (2007, 7ff und 11ff). Für eine unterstützende neurowissenschaftliche Perspektive vgl. Price et al. (1999, 218).

7 Schlussbetrachtung

wiederum in den Folgeperioden genutzt werden könnten. Eine Simulation dazu wäre wünschenswert.

In diese Simulationen könnten nun auch weitere Informationen hinzugegeben und deren Wirkung gemessen werden. Die Ausweitung des Algorithmus erfolgt dann auch in die Richtung von syntaktischen, semantischen, gar pragmatischen Fragestellungen: Wie wirkt sich beispielsweise die Zuweisung von Informationen aus dem Weltwissen aus? Dies können auch logische Verknüpfungen sein oder die Betonung bestimmter Lautfolgen mit Gestiken und Mimiken. Kategorien wie Verb, Substantiv oder Adjektiv würden als Metainformation dann dazu einfließen. Daraus ergibt sich fast schon automatisch die Frage, ob einfache syntaktische Strukturen mit dem gleichen algorithmischen Prinzip erkannt werden könnten?²⁰ Hier besteht die Möglichkeit, die Interaktion der verschiedenen Kommunikationsebenen und deren Informationsinput zu simulieren. Die Wirkung von semantischen und syntaktischen Informationen würde schließlich zusammen und gesondert getestet werden können.²¹

Schließlich wäre auch eine Implementierung des Algorithmus als rein neuronales Netzwerk denkbar. Hierzu müssen allerdings noch zusätzliche Annahmen aufgestellt werden, die die genaue Architektur, Lernregel und die Anfangszustände der Knoten festlegen.

Selbst eine einfache Wiederholung der Simulation mit anderen Korpora würde weitreichende Erkenntnisgewinne erbringen. An erster Stelle wäre die Verwendung anderer Sprachen anzuführen. Hier bestände die Herausforderung in erster Linie in der Beschaffung beziehungsweise Anfertigung von entsprechenden Lautsprachelexika und der Auswertung der Daten, insofern man denn auch Sprachen, die nicht dem indogermanischen Sprachstamm angehören, untersuchen möchte. Selbst die Simulation an Korpora verschiedener Zeiten und Genres dürfte einige interessante Ergebnisse liefern.²² Es wäre doch gut zu wissen, ob sich die Gesetzmäßigkeiten auf Textkorpora unterschiedlicher Genres übertragen ließen. Dies bietet eine Möglichkeit, in der Zeit zurückzugehen, beispielsweise Texte aus dem 16. Jahrhundert zu Grunde zu legen und daraus abzuleiten, dass die funktionalen Beziehungen der Phonemanordnungen auch schon damals (vielleicht sogar schon immer) in der Sprache kodiert waren.

²⁰vgl. Jusczyk (2001, 371) als kurze Zusammenfassung über die Anwendung des statistischen Ansatzes auf die grammatische Organisation der Sprache. Erste experimentelle Befunde weisen in eine vielversprechende Richtung. Matthews et al. (2005) sehen die Häufigkeit der Verwendung von Wortklassen (Verben) im direkten Zusammenhang mit den Wortstellungsregeln der englischen Sprache. Mit *SCREEN* (Wermter und Weber, 1997) steht dazu schon eine Netzwerksimulation zur Verfügung, welche flache syntaktische Repräsentationen erkennen kann. Brents *Learner* (Brent, 1993) ist eine probabilistische, symbolische Modellierung, die aus einem englischen Korpus syntaktische Informationen einzelner Wörter auf Grundlage von lokalen probabilistisch kodierten morphosyntaktischen Hinweisen (*cues*) in Form von abstrakten Wortklassenzuweisungen extrahiert.

²¹Wolff (1976) hat seine Variante des Algorithmus auf nicht sprachliche Mechanismen der Konzeptbildung (*concept formation*) ausgeweitet und fand heraus, dass dies für bestimmte Typen der höheren Mustererkennung auch sinnvoll erscheint.

²²vgl. auch hier Wolff (1977) für eine der wohl ersten systematischen Untersuchungen dazu.

7.3 Ausblick zur wirtschaftlichen Nutzung der Ergebnisse

Die Computerlinguistik versteht sich als eine sowohl interdisziplinäre als auch angewandte Wissenschaft. Es soll also nicht verwundern, dass die Anwendungen der Computerlinguistik nicht nur auf die Überprüfung linguistischer Theorien beschränkt bleiben, sondern sich auch sehr pragmatisch vermarkten lassen. Für den Schlussteil der Arbeit sei diese Ausführung als Vorzug der computerlinguistischen Disziplin gestattet und resümierende Gedanken über dieses, von den Geisteswissenschaften oft tabuisierte Thema, erlaubt. Im Folgenden soll ein Bezug zum wirtschaftlichen und gesellschaftlichen Nutzen dieser Erkenntnisse hergestellt werden.

Offensichtlich können die hier vorgestellten Algorithmen in der angewandten Computerlinguistik für Sprachsoftware enorme Vorteile bringen, da sich die Algorithmen als automatisierte Wortsegmentierungsmechanismen herausgestellt haben, bei denen man auf eine speicherintensive Verarbeitung von sprachabhängigen Heuristiken verzichten kann. Dieser Punkt trägt einer anhaltenden Kritik Rechnung, geisteswissenschaftliche Arbeiten verfehlten das Ziel, ihre Erkenntnisgewinne der Gesellschaft als Ganzes zugänglich zu machen beziehungsweise ihren Nutzen explizit zu konkretisieren.

Im vorherigen Kapitel 7.2 wurden bereits mögliche Anknüpfungspunkte aufgezeigt, den Algorithmus auf die der Wortsegmentierung vorgelagerte, akustische Spracherkennung und vorhandene Informationen im Paralexikon nutzbar zu machen. Der Erfolg einer marktorientierten Umsetzung hängt dabei entscheidend von der vorgelagerten akustischen Spracherkennung ab. Gelingt es, die Grundstruktur des Algorithmus auf die Muster einzelner Phone oder Diphone²³ zu übertragen, wäre ein entscheidender Durchbruch auf vielen Sektoren der natürlichen Sprachverarbeitungssysteme im Bereich der Sprachanalyse gelungen.

Die hier erbrachten Ergebnisse können gewinnbringend in bestehende Spracherkennungssysteme eingebracht werden. Nach wie vor ist es problematisch, den Sprachstrom in einzelne Wörter zu unterteilen.²⁴

The recognition of isolated words is much easier than the recognition of continuous speech. The main fact contributing to the higher complexity of continuous recognition is the necessity to model word boundaries explicitly in the recognition model, while this can be neglected if the system can be certain that words are isolated. (Farghaly, 2003, 370)

An dieser Stelle bietet der entwickelte Algorithmus gute Dienste, wenn es sich um längere Texte des Englischen handelt (ab 2500 Wörter). Das Verfahren kann auch ergänzend eingesetzt werden, um bei unsicheren Segmentierungsentscheidungen, eine Gewichtung für eine der Alternativen zu errechnen.

Relativ einfache Anwendungen, wie Bürodiktiersysteme, die vom Markt bereits angenommen wurden, leiden unter dem schwerfälligen und langsamen Abgleich von Aus-

²³Vielversprechende Ansätze wurden bereits im Diphonbereich erzielt. Insbesondere müssen aber Sprecherunabhängige Eigenschaftsschnittmengen von sprachlichen Einheiten weiter ergründet werden.

²⁴vgl. Deshmukh et al. (1996)

7 Schlussbetrachtung

schnitten des physischen Sprachsignals.²⁵ Recht umständlich und zeitraubend müssen diese Systeme auf die Eigenheiten einer jeden Stimme trainiert werden.²⁶ Der hier vorgestellte Algorithmus ist einfach implementierbar und kann auf ein vorgegebenes Lexikon gänzlich verzichten. Dadurch könnten selbst Webapplikationen mit Applets die Segmentierungsaufgabe übernehmen – eine Anwendung, welche mit dem herkömmlichen Verfahren unmöglich zu sein scheint.

Entscheidende Vorteile würde auch die kurze Verarbeitungszeit der Segmentierung bringen. Herkömmliche Dialogsysteme zur mündlichen Sprachbearbeitung sind entweder auf die Verwendung sehr weniger Einwortsätze wie „ja“ und „nein“ beschränkt²⁷ oder der Dialog wirkt äußerst unnatürlich und zeitlich stark verzögert.²⁸ Der Grund dafür liegt natürlich wieder im Abgleich mit einem übergroßen Lexikon. Selbst die effizientesten Suchalgorithmen und viele weitere Leistungsverdopplungen der Prozessoren werden dieses Problem zunächst nicht beheben. Analog zur Funktionsweise unseres Gehirns wird man auf abstrakte Ebenen nicht verzichten können.

Die typische Lösung für derartige kommerzielle Anwendungen besteht dank der erhöhten Rechenkapazität mittlerweile auch in der Verwendung von Hidden-Markoff-Ketten, die noch vor wenigen Jahren lediglich für die Erkennung der einzelnen Phoneme verwendet werden konnten. Andere Ansätze gehen von Bayesianischen Statistiken oder neuronalen Netzen aus (*time delay neural networks*) und von Kombinationen aus beiden. In allen Spielarten dieser Ideen werden die statistisch kodierten Muster des Sprachsignals genutzt. Problematisch ist bei den Modellen jedoch oft die Festlegung des Anfangszustandes. Er muss im Prinzip willkürlich, per Zufall oder etwa nach dem Laplaceschen Gleichverteilungsprinzip erfolgen. Die Ansätze der Informationsentropie versuchen diese *a priori*-Wahrscheinlichkeiten besser abzuschätzen. Der hier vorgestellte Algorithmus gibt eine weitere Grundlage für derartige Anfangszustände.

Die Bayesformel folgt im Prinzip schon aus der Definition der bedingten Wahrscheinlichkeiten (siehe Kapitel 5.2.3), da man sie durch geschicktes Umformen der Gleichung erhält.²⁹ Diese Umformung bewirkt natürlich dann auch das Umkehren der Schlussfolgerungen von Ursache und Ereignis. Es könnte somit als Rückwärtsinduktion bezeichnet werden. Die *a priori*-Wahrscheinlichkeiten $P(A)$ und $P(B)$ können auf Grundlage der hier angestellten Berechnungen dann aber genauer festgelegt werden. Auf gleiche Weise können die Übergangswahrscheinlichkeiten bei verborgenen Markoff-Ketten bestimmt werden. Im 5-Tupel einer verborgenen Markoff-Kette müssen die Startzustände wie auch die Ausgabe- und Übergangswahrscheinlichkeiten per Zufall oder willkürlich festgelegt

²⁵Beispielsweise sei das System *Dragon NaturallySpeaking* der 4VoiceAG genannt, dass überraschend in guten Kritiken als bestes System dieser Art empfohlen wird (vgl. z.B. FAZ vom 5. September 2006). Es verfügt über spezielle Vokabularien im Bereich Medizin und Rechtswissenschaft.

²⁶Man möge die bereits sehr gute *Open Source Software* Sphinx der *Carnegie Mellon University* dazu testen (<http://cmusphinx.sourceforge.net/html/cmusphinx.php>).

²⁷Hier kann das automatische Auskunftssystem der Deutschen Bundesbahn oder das Telefonbanking von Banken angeführt werden.

²⁸Wie es auch noch bei dem Verbomobil deutlich zu Tage tritt.

²⁹Man erweitert die Verbundwahrscheinlichkeit $P(A|B)$ mit $P(A)$, wodurch man im Zähler nun $P(B|A)P(A)$ erhält beziehungsweise für endlich viele Ereignisse i $P(B|A_i)P(A_i)$

7.3 Ausblick zur wirtschaftlichen Nutzung der Ergebnisse

werden. Erst nach mehrfachen Durchläufen nähert sich das Modell dem Zustand, wo die Anpassungsalgorithmen³⁰ keine Verbesserung mehr erzielen. Die Durchläufe können in einer heuristischen Form von den hier in einer sichtbaren Markoffkette ergründeten Gesetzmäßigkeiten für Phonemübergänge festgelegt werden. Dies hat den Vorteil, die Komplexität der Markoff-Modelle leichter bewältigen zu können und nicht durch lokale Extrempunkte in den versteckten Zuständen falsche Wahrscheinlichkeiten zu erhalten, da man ja als Vergleich die typischen Durchschnittswerte des Phonemsystems als Übergangswahrscheinlichkeit zu Grunde legt.

Die hier skizzierten Anwendungsfelder hängen vorrangig von der Segmentierungsleistung des Inputs ab. Dies gilt immer dann, wenn gesprochene Sprache verarbeitet werden soll. Das Segmentierungsproblem ist also Grundbestandteil der mündlichen Kommunikation bei Menschen wie bei Maschinen gleichermaßen und damit wird der hier vorgestellte Algorithmus für alle Anwendungen relevant sein, bei denen mündliche Sprache von Maschinen analysiert werden soll. Im Zuge der weiter expandierenden Internetkommunikation sind noch weitere Anwendungen zu erwarten, die schnellere Verarbeitungsmechanismen benötigen. Traditionelle Eingabehilfen, wie Tastatur und Maus, werden schon in näherer Zukunft durch sprachliche Dialoge ersetzt werden können.

³⁰Baum-Welch, Jelinek- oder Kubiecs-Methoden (vgl. Manning und Schütze, 2003, 333ff und 357f).

8 Zusammenfassung

In der vorliegenden Arbeit wurde die Wortsegmentierung ohne anfänglichen Rückgriff auf ein Lexikon bei sechs bis acht Monate alten Babys untersucht und mittels einer Computersimulation gezeigt, dass durch die Berechnung von Übergangswahrscheinlichkeiten als *Bottom-up-Bootstrapping*-Verfahren in Verbindung mit rekursiven Strukturen als nachgelagertes *Top-down*-Verfeinerungsverfahren eine theoretische Möglichkeit zur Segmentierung eines Lautstroms besteht, die auf keine sprachspezifische Information zurückgreift. Die resultierende Wortliste kann zudem für die weitergehende Regelextraktion, von allophonischen, phonotaktischen und rhythmischen Mustern genutzt werden. Damit kann auch eine konkrete Alternative zum Argument, die anfängliche Wortliste entstehe durch die in der Sprache isoliert auftretenden Wörter, gegeben werden. Wie gezeigt werden konnte, ist das Argument der isolierten Wörter ein schlechter Ausweg aus dem Dilemma des Bootstrappingparadoxons, weil der Sprachlerner zu keinem Zeitpunkt wissen kann, ob es sich um ein einzelnes oder um mehrere Wörter handelt.

Am Anfang der Arbeit wurde in die Problematik eingeführt. Dabei konnte der Bezug zum Spracherwerb beibehalten werden. Nach der Beschreibung von generellen Überlegungen, Annahmen und Voraussetzungen ist die Einordnung der Arbeit aus verschiedenen Perspektiven erfolgt. Das Thema wurde sowohl im wissenschaftlichen Kontext mit anderen Wissensbereichen abgegrenzt sowie in den philosophischen Denktraditionen verortet. Von dort erfolgte schließlich der Übergang zu Forschungsausprägungen in jüngerer Zeit, welche auf die Linguistik, Psychologie und künstliche Intelligenz beschränkt blieb. Hier sollte zunächst die Nähe zwischen Symbolverarbeitung und nativistisch geprägter Arbeit einerseits sowie Konnektionismus und empiristischer Forschung andererseits aufgezeigt werden, jedoch nur deshalb, um dessen Dogmenhaftigkeit herauszustellen und eine Vermengung beider Ansätze – Symbolverarbeitung und Konnektionismus – als treffendere Beschreibung anzuerkennen sowie von den empiristischen und nativistischen Ideen abzukoppeln. Sodann wurde die Arbeit in der empiristischen Nähe positioniert.

Im weiteren Verlauf der Arbeit wurden die Voraussetzungen der Simulation vorbereitet. Das dritte Kapitel diente als vorbereitende Einführung in die Problematik der lexikalischen Erschließung. Im ersten Teil wurden alle gängigen Modelle, die ein Lexikon voraussetzen, beschrieben. Zum einen diente diese Abhandlung als Überblick über herkömmliche Lösungsmodelle des Zugriffproblems; zum anderen zeigte es, welches Modell in der Simulation am Besten verwandt werden konnte. Hierzu ist anzumerken, dass entsprechende Modelle für Paralexika wie sie möglicherweise von Kleinkindern zwischen sechs und acht Monaten verwendet werden, nicht vorhanden sind und deshalb auf die allgemeinen Theorien und Modelle zurückgegriffen werden musste. Im zweiten Teil des Kapitels konnten nun die speziellen Theorien für die Wortsegmentierung ohne Lexikon dargelegt werden. Für diese Ausführungen wurden spezielle Experimente für das ent-

8 Zusammenfassung

sprechende Alter mittlerweile durchgeführt. Diese konzentrieren sich auf prosodische, allophonische, phonotaktische und statistische Segmentierungsstrategien. In einer kritischen Auseinandersetzung mit den Theorien konnte eine mögliche zeitliche Abfolge vorgeschlagen werden. Dabei wurden die statistischen Verteilungen als Ausgangspunkt für alle weiteren Segmentierungsstrategien bekannt gegeben. Die Strategie der statistischen Verteilung stellten sich damit als erste Randbedingung der Simulation heraus.

Im vierten Kapitel wurde diese Strategie in ihrer Funktion als Randbedingung wiederholt betrachtet. Als weitere Randbedingungen erfolgte nun die wohl wichtigste Abhandlung über die linguistischen Einheiten, von denen die Übergangswahrscheinlichkeiten berechnet werden sollten. Auch hier stellte sich erst später mit den Ergebnissen der Simulation heraus, dass die anfangs wichtigste Entscheidung zum Design des Algorithmus keinen entscheidenden Einfluss auf die Segmentierungsleistung nehmen würde. Alle weiteren Modellannahmen, die Erinnerung an häufige Lautsegmente, die Fähigkeit zum Einfügen dieser Einheiten im *Top-down*-Verfahren und die Fähigkeit, die segmentierten Einheiten zu verwenden, verhielten sich wie erwartet als entscheidende Einflussfaktoren.

Mit diesem Wissen sollte endlich die Arbeit am Algorithmus erfolgen. Sie wurde im fünften Kapitel ausführlich dargelegt. Dazu gehörten die Herausarbeitung des Leistungsmaßes F_1 , das als Gleichgewichtung von Präzision und Vollständigkeit die segmentierten Lücken objektiv erfassen sollte, eine Beschreibung auf welche Weise die Übergangswahrscheinlichkeiten berechnet werden und eine Spezifizierung von primären und sekundären Variablen, welche die Leistung des Algorithmus auf verschiedenen Ebenen kontrollieren würden. Ausführungen zu den verwendeten Ausgangsdaten im dritten Unterkapitel garantieren die Replizierbarkeit der Simulation. Die Grundarchitektur wurde in vier Schritten beschrieben: Manipulation des Input als IPA transkribierten Lautstrom, Berechnung der Übergangswahrscheinlichkeiten, Anwenden der Häufigkeitsfunktion und Erstellen eines Paralexikons, Einfügen der Einheiten des Paralexikons in den manipulierten Input. Es folgten schließlich nur einige eher technische Randnotizen zum algorithmischen Ablauf der Simulation. Die direkte Verbindung zu einer erklärenden Theorie stellte eine zusammenhängende Darstellung am Ende des Kapitels her.

Das sechste Kapitel behandelte sodann die Ergebnisse der Simulation, welche einer Interpretation und weitergehenden Analyse unter Rückgriff auf die Erkenntnisse der theoretischen Abhandlungen der Kapitel 3 und 4 unterzogen wurden. Als wichtige Eckpunkte bleiben festzuhalten: Die Quantität des Inputs wirkt sich entscheidend auf das Ergebnis aus. Es lässt sich durchweg eine Zweiteilung von großen und kleinen Korpora begründen. Die kritische Masse liegt bei etwa 5000 Zeichen, sodass man einen entsprechend großen Input beim Kind annehmen müsste. Unerheblich ist dagegen die genaue Form der linguistischen Zugriffseinheiten. Grundsätzlich führt die Verwendung von Silben, einzelnen Phonemen oder Kombinationen aus beiden zu gleichen Segmentierungen. Der Sprachlerner sollte dabei jedoch recht „unsensibel“, d.h. großzügig mit den statistischen Verteilungen umgehen. Die Sensibilität ist bei entsprechendem Input so klein wie möglich zu wählen.

Unterstellt man lediglich die Berechnung der Übergänge dieser Einheiten kann keine

zufriedenstellende Segmentierung beobachtet werden. Bis hierher und dies ist bislang erfolgt erscheint das Argument der Reizarmut plausibel. Dabei hat man anscheinend eine sehr wichtige Qualität der segmentierten Einheiten übersehen, die auch nur durch eine automatisierte statistische Verarbeitung zu Tage tritt. Die entscheidende Beobachtung aus einer sehr detaillierten Analyse des Datenmaterials ist, dass die häufigsten Einheiten der statistischen Segmentierung immer auch lexikalische Einheiten sind. Legt man also eine weitere Häufigkeitsfunktion über den im Grunde fehlerhaft segmentierten Korpus, scheiden auf Dauer und im Durchschnitt die falschen Segmentierungen aus. Diese Eigenschaft der segmentierten Einheiten konnte letztendlich zum Aufbau eines Paralexikons genutzt werden. Dazu folgten Beschreibungen über die Eigenschaften des Paralexikons und seiner Entwicklung.

Die funktionalen Verläufe der richtigen Einträge im Paralexikon zeigten zueinander lineare Verläufe, sodass eine Masterfunktion ermittelt werden konnte, deren Grenzwert die maximale Größe des Paralexikons zu jeder gegebenen Lexikonfunktion ergibt. Aufgrund des sehr geringen Fehlerverhältnisses von richtigen und falschen Einträgen fiel die Entscheidung auf eine konstante Lexikonfunktion von durchschnittlich zehn Einträgen ins Paralexikon. Diese Entscheidung wurde auch aus Sicht der Leistungsmaße gestützt, denn selbst eine größere Wortanzahl brachte keine signifikante Verbesserung in der Segmentierung. Mit dem Beginn der Rekursion resultierte schließlich eine beträchtliche Verbesserung der richtigen Segmentierungen, da nun der Nachteil der Übergangswahrscheinlichkeiten durch die Einfügungen des Paralexikons kompensiert werden konnte und damit auch die Widerlegung des Arguments der Reizarmut für den Bereich der Wortsegmentierung greifbar wurde. Erst jetzt erschien die Information für eine nahezu vollständig richtige Segmentierung ausreichend. Eine Untersuchung der Eigenschaften des Paralexikons bestätigte dies, da die Qualität der Einträge eine derartige Form aufwies, dass mit den Fähigkeiten der Säuglinge, die prosodischen, phonotaktischen und allophonischen Strategien anzuwenden, nun eine vollständige Segmentierung eines Lautstroms möglich sein würde. Dieser letzte Schritt zeigte eine synergetische Zusammenarbeit von Experiment und Simulation, so wie sie in Kapitel 1.5 idealtypisch beschrieben wurde.

Im letzten Kapitel konnte der in der Einleitung gespannte Bogen entspannt werden. In einem kurzen Exkurs zur Anlage-Umwelt-Debatte wurde das Ergebnis der Simulation als Stärkung des Arguments, dass die Umwelt für die Sprachentwicklung bedeutend ist, interpretiert und das Argument der Reizarmut für den Bereich der anfänglichen Wortsegmentierung negiert. Ein Ausblick zur weiteren Forschung und wirtschaftlichen Nutzung stellte weiterführende Bezüge zur Anwendung aus praktischer und theoretischer Sicht her.

Anhang: Quellcodes, Wortlisten und Kodierungen

Inhaltsangabe

1	Anhang A: Quellcode	239
1.1	NextWord.java	239
1.2	IPA.java	239
1.3	WordCount2.java	245
1.4	LexiconDelayed.java	249
1.5	TransProbAutoCompareLexiconDelayed.java	251
1.6	Homonyme und Heteronyme des CMU: Homohete.pl	257
2	Anhang B: Wortlisten für $\zeta = 10$ und $\zeta = 30$	258
3	Anhang C: IPA und SAMBA Notationen	267

1 Anhang A: Quellcode

1.1 NextWord.java

```
1 public class NextWord
2 {
3     int position;
4     String word;
5
6     NextWord(String s,int i)
7     {
8         word = s;
9         position = i;
10    }
11 }
```

1.2 IPA.java

```
1 import java.util.*;
2 import java.io.*;
3 import java.util.regex.*;
4
```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
5
6 public class IPA
7 {
8     static String cmudict = "";
9     static
10    {
11        int cnt=0;
12        String line = "";
13        try
14        {
15            FileReader fr = new FileReader("cmudict");
16            BufferedReader br = new BufferedReader(fr);
17            char [] cmubuff = new char[3511906];
18            fr.read(cmubuff,0,3511906);
19            fr.close();
20            cmudict = String.valueOf(cmubuff);
21        } catch(IOException e){}
22    }
23
24    static String[] cmuChars = {"AA","AE","AH","AO","AW","AY","CH",
25                                "DH","EH","ER","EY","HH","IH","IY",
26                                "JH","NG","OW","OY","SH","TH","UH",
27                                "UW","ZH","B","D","F","G","K","L",
28                                "M","N","P","R","S","T","V","W",
29                                "Y","Z"};
30    static char[] ipaChars = {0x252, 0x259, 0x28C, 0x254, 0x263A,
31                                0x263B, 0x2A7, 0x0F0, 0x065, 0x25C,
32                                0x263C, 0x068, 0x26A, 0x069, 0x2A4,
33                                0x14B, 0x2640, 0x2642, 0x283, 0x3B8,
34                                0x28A, 0x075, 0x292, 0x062, 0x064,
35                                0x066, 0x067, 0x06B, 0x06C, 0x06D,
36                                0x06E, 0x070, 0x072, 0x073, 0x074,
37                                0x076, 0x077, 0x06A, 0x07A};
38
39    static String thisData;
40
41    static boolean isLetter(char c)
42    {
43        return
44        (
45            ((c>=65)&&(c<=90))
46            ||
47            ((c>=97)&&(c<=122))
48            ||
```

```

49         (c==0xe4)
50         ||
51         (c==0xf6)
52         ||
53         (c==0xfc)
54         ||
55         (c==0xc4)
56         ||
57         (c==0xd6)
58         ||
59         (c==0xdc)
60         ||
61         (c==0xdf)
62         ||
63         (c==39)
64     );
65 }
66
67 static boolean isIPAletter(char c)
68 {
69     boolean isIPA = false;
70     for(int i=0;i<ipaChars.length;i++)
71         if(c==ipaChars[i])
72             isIPA = true;
73
74     return(isIPA);
75 }
76
77 static NextWord findWord(int pos)
78 {
79     while((pos<thisData.length())&&!isLetter(thisData.charAt(pos)))
80     {
81         pos++;
82     }
83     String s = "";
84     while((pos<thisData.length())&&isLetter(thisData.charAt(pos)))
85     {
86         s=s.concat(String.valueOf(thisData.charAt(pos)));
87         pos++;
88     }
89     NextWord n = new NextWord(s,pos);
90     return(n);
91 }
92

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
93     static String convertTextToIPA(String data)
94     {
95         thisData = data;
96         String dataOut = "";
97         int pos = 0;
98         while(pos<data.length())
99         {
100             NextWord nextWord = findWord(pos);
101             nextWord.word = nextWord.word.toLowerCase();
102
103             String w = convertToIPA(nextWord.word);
104             if(w!=null)
105             {
106                 nextWord.word = w;
107                 dataOut+=nextWord.word;
108             }
109             pos = nextWord.position;
110         }
111         return(dataOut);
112     }
113
114     static String convertTextToIPASpace(String data)
115     {
116         thisData = data;
117         String dataOut = "";
118         int pos = 0;
119         while(pos<data.length())
120         {
121             NextWord nextWord = findWord(pos);
122             nextWord.word = nextWord.word.toLowerCase();
123             String w = convertToIPA(nextWord.word);
124             if(w!=null)
125             {
126                 nextWord.word = w;
127                 dataOut+=nextWord.word+" ";
128             }
129             pos = nextWord.position;
130         }
131         return(dataOut);
132     }
133
134     static String convertToIPA(String word)
135     {
136         int pos = cmudict.indexOf(String.valueOf((char)0x0a)).
```

```

137         concat(word.toUpperCase().concat(" ")));
138     if(pos>=0)
139     {
140         String cmuword = cmudict.substring(pos+3+word.length(),
141             cmudict.indexOf(String.valueOf((char)0x0a),
142             pos+3+word.length()));
143         cmuword = cmuword.replaceAll("\\d","");
144         for(int i=0;i<cmuChars.length;i++)
145         {
146             cmuword = cmuword.replace(cmuChars[i],String.
147                 valueOf(ipaChars[i]));
148         }
149         cmuword = cmuword.replaceAll("\\s","");
150         return(cmuword);
151     }
152     return(null);
153 }
154
155 static String convertToPlain(String word)
156 {
157     String cmuword=word;
158     for(int j=0;j<IPA.cmuChars.length;j++)
159     {
160         cmuword = (cmuword.replace(String.valueOf(IPA.ipaChars[j]),
161             IPA.cmuChars[j]+" ")).trim();
162     }
163
164     Pattern p = Pattern.compile("\\n(.*)\\s\\s"+cmuword+"\\n");
165     Matcher m = p.matcher(cmudict);
166     return(m.find() ? m.group(1) : null);
167 }
168
169
170 static String IPAToHTML(String word)
171 {
172     String html = "";
173     for(int i=0;i<word.length();i++)
174     {
175         if(word.charAt(i)>255)
176             html=html.concat("&#"+(int)word.charAt(i));
177         else
178             html=html.concat(String.valueOf(word.charAt(i)));
179
180         html=html.replace("\n","<br>");

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
181     /*int j=0;
182     while((j<Category.ipaChars.length()) && (word.charAt(i)!=
183                                             ipaChars[j])) j++;
184     html.concat(""+(int) ipaChars[j]);*/
185 }
186 return(html);
187 }
188
189 static String substituteDiphthongs(String data)
190 {
191     data=data.replace(String.valueOf((char)0x2640),
192                     String.valueOf((char)0x259)+String.valueOf((char)0x28A));
193     data=data.replace(String.valueOf((char)0x263b),
194                     String.valueOf((char) 0x61)+String.valueOf((char)0x26A));
195     data=data.replace(String.valueOf((char)0x263c),
196                     String.valueOf((char) 0x65)+String.valueOf((char)0x26A));
197     data=data.replace(String.valueOf((char)0x2642),
198                     String.valueOf((char)0x254)+String.valueOf((char)0x26A));
199     data=data.replace(String.valueOf((char)0x263a),
200                     String.valueOf((char) 0x61)+String.valueOf((char)0x28A));
201     return(data);
202 }
203
204 static String resubstituteDiphthongs(String data)
205 {
206     data=data.replace(String.valueOf((char)0x259)+
207                     String.valueOf((char)0x28A), String.valueOf((char)0x2640));
208     data=data.replace(String.valueOf((char) 0x61)+
209                     String.valueOf((char)0x26A), String.valueOf((char)0x263b));
210     data=data.replace(String.valueOf((char) 0x65)+
211                     String.valueOf((char)0x26A), String.valueOf((char)0x263c));
212     data=data.replace(String.valueOf((char)0x254)+
213                     String.valueOf((char)0x26A), String.valueOf((char)0x2642));
214     data=data.replace(String.valueOf((char) 0x61)+
215                     String.valueOf((char)0x28A), String.valueOf((char)0x263a));
216     return(data);
217 }
218
219 static String readFile(String file, boolean unicode)
220 {
221     String data = "";
222     String line = "";
223     char cData[] = new char[5*1048576];
224     int i = 0;
```

```

225     try
226     {
227         BufferedReader fr =
228             new BufferedReader(
229                 new InputStreamReader(new FileInputStream(file),
230                     unicode?"UnicodeLittle":"ISO8859_1"));
231         i = fr.read(cData);
232         data = new String(cData, 0, i);
233         fr.close();
234     } catch(IOException e){}
235     return(data);
236 }
237
238 static void writeFile(String file,String data, boolean unicode)
239 {
240     try
241     {
242         BufferedWriter bw = new BufferedWriter(unicode?new
243             OutputStreamWriter(new
244                 FileOutputStream(file),
245                 "UnicodeLittle"):new FileWriter(file));
246         bw.write(data);
247         bw.close();
248     } catch(IOException e){}
249 }
250
251 }

```

1.3 WordCount2.java

```

1  /*
2  *  Usage:
3  *    java WordCount2 <input file> <output File> [x] [ipa]
4  *
5  */
6
7
8  import java.util.*;
9  import java.io.*;
10
11
12  public class WordCount2
13  {

```

```
14
15 public static void main(String[] args)
16 {
17     boolean ipa = args.length>3 && "ipa".equals(args[3]);
18
19     int x = (args.length>2)?Integer.valueOf(args[2]):1000000;
20
21     Hashtable words = new Hashtable();
22     Hashtable oldWords = new Hashtable();
23
24     String oldLexicon = args[1];
25     int runIndex = oldLexicon.lastIndexOf("_".charAt(0),
26                 oldLexicon.lastIndexOf("_".charAt(0))-1)-1;
27     int runFirstIndex = oldLexicon.lastIndexOf(".",charAt(0),
28                 runIndex)+1;
29     System.out.println("===>"+runFirstIndex+" "+runIndex);
30     int oldRun = Integer.parseInt(oldLexicon.substring(runFirstIndex,
31                 runIndex+1))-1;
32     oldLexicon = oldLexicon.substring(0,runFirstIndex)+oldRun+
33                 oldLexicon.substring(runIndex+1,
34                 oldLexicon.length());
35     System.out.println("====> "+oldLexicon);
36     if((new File(oldLexicon)).exists())
37         oldWords = readVocabulary(oldLexicon,ipa);
38     else
39         System.out.println("Noooooooooooooooooooooo "+oldLexicon);
40
41     int fpos = 0;
42
43     String f = IPA.readFile(args[0],ipa)+" ";
44
45     while(fpos<f.length())
46     {
47         while(fpos<f.length() && isWhiteSpace(f.charAt(fpos)))
48         {
49             fpos++;
50             //System.out.println("-"+f.charAt(fpos));
51         }
52         int start=fpos;
53         while(fpos<f.length() && !isWhiteSpace(f.charAt(fpos)))
54         {
55             fpos++;
56             //System.out.println("+");
57         }
```

```

58     String w = f.substring(start,fpos);
59     //System.out.println("#"+w);
60     if(w.equals("")) break;
61     if(words.containsKey(w))
62     {
63         words.put(w, new Integer(((Integer)words.get(w)).intValue()+
64                                     1));
65     }
66     else
67     {
68         words.put(w, new Integer(1));
69     }
70 }
71
72 Vector v = new Vector(words.keySet());
73 for(int i=0;(i<v.size()-1) && (i<x);i++)
74 {
75     for(int j=i+1;j<v.size();j++)
76     {
77         if(((Integer)words.get((String)v.elementAt(i))).intValue()
78             < ((Integer)words.get((String)v.elementAt(j))).intValue())
79         {
80             Object o = v.elementAt(i);
81             v.setElementAt(v.elementAt(j),i);
82             v.setElementAt(o,j);
83         }
84     }
85 }
86
87 for(int i=v.size()-1;i>=x;i--)
88 {
89     if(((Integer)words.get((String)v.elementAt(i))).intValue()
90         <((Integer)words.get((String)v.elementAt(x-1))).intValue())
91         v.removeElementAt(i);
92 }
93
94 int countNew = 0;
95 for(int i=0;i<v.size();i++)
96 {
97     if(oldWords.containsKey((String)v.elementAt(i)))
98     {
99         oldWords.put((String)v.elementAt(i),
100             ((Integer)words.get((String)v.elementAt(i))).intValue()+
101             ((Integer)oldWords.get((String)v.elementAt(i))).intValue());

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
102     }
103     else
104     {
105         countNew++;
106         oldWords.put((String)v.elementAt(i),
107                     words.get((String)v.elementAt(i)));
108     }
109 }
110
111 try
112 {
113     FileWriter fw = new FileWriter("wordcount",true);
114     BufferedWriter bw = new BufferedWriter(fw);
115     bw.write(args[1]+" "+countNew+" new words\n");
116     bw.close();
117     fw.close();
118 } catch(IOException e){}
119
120 String dataOut = "";
121 v = new Vector(oldWords.keySet());
122 for(int i=0;(i<v.size()-1);i++)
123 {
124     for(int j=i+1;j<v.size();j++)
125     {
126         if((((Integer)oldWords.get((String)v.elementAt(i))).intValue()
127         < ((Integer)oldWords.get((String)v.elementAt(j))).intValue())
128         {
129             Object o = v.elementAt(i);
130             v.setElementAt(v.elementAt(j),i);
131             v.setElementAt(o,j);
132         }
133     }
134 }
135
136 for(int i=0;(i<v.size());i++)
137 {
138     dataOut+=(((String)v.elementAt(i)) + " " +
139              ((Integer)oldWords.get((String)v.elementAt(i))).intValue())
140              +"\n";
141 }
142
143 IPA.writeFile(args[1],IPA.substituteDiphthongs(dataOut),ipa);
144 }
145
```

```

146     static boolean isWhiteSpace(char c)
147     {
148         return(!(IPA.isLetter(c) || IPA.isIPALetter(c)));
149     }
150
151     static Hashtable readVocabulary(String file, boolean ipa)
152     {
153         String s = IPA.resubstituteDiphthongs(IPA.readFile(file, ipa));
154         Hashtable words = new Hashtable();
155         String lines[] = s.split("\n");
156         for(String l:lines)
157         {
158             //System.out.println(l);
159             if(l.matches("[a-zA-Z'+String.copyValueOf(IPA.ipaChars)+'"]+
160                                     "\\d+"))
161             {
162                 //System.out.println("!");
163                 int index = l.indexOf(" ");
164                 words.put(l.substring(0, index),
165                           Integer.valueOf(l.substring(index+1)));
166             }
167         }
168         return(words);
169     }
170 }

```

1.4 LexiconDelayed.java

```

1  import java.util.*;
2  import java.io.*;
3
4
5  public class LexiconDelayed
6  {
7      public static void main(String[] args) throws IOException
8      {
9
10         String file = args[0];
11         int maxlength = 2;
12         int raster = 2;
13         int runs = 3;
14         int delay = 2;
15

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
16     for (int i = 1; i < args.length; i++)
17     {
18         switch(i)
19         {
20             case 1:
21                 maxlength = (Integer.parseInt(args[i]));
22                 break;
23             case 2:
24                 raster = (Integer.parseInt(args[i]));
25                 break;
26             case 3:
27                 runs = (Integer.parseInt(args[i]));
28                 break;
29             case 4:
30                 delay = (Integer.parseInt(args[i]));
31                 break;
32         }
33     }
34
35     for(int run=1;run<=runs;run++)
36     {
37         copy(new File(file+"."+run),new File(file));
38         String ar[] = {file, String.valueOf(maxlength),
39                       String.valueOf(raster), String.valueOf(run),
40                       String.valueOf(delay), "ipa"};
41         TransProbAutoCompareLexiconDelayed.main(ar);
42         File dir = new File(".");
43         String[] subfile = dir.list();
44         System.out.println(file+"\\."+run+"_*\\.substitute");
45         for(String f:subfile)
46         {
47             if(f.matches(file+"\\."+run+"_*\\.substitute"))
48             {
49                 String lexfile=f.replace(".substitute",".lexicon");
50                 System.out.println(lexfile);
51                 String ar2[] = {f, lexfile,
52                                 String.valueOf(grenzfunktion(run)), "ipa"};
53                 WordCount2.main(ar2);
54             }
55         }
56     }
57 }
58
59 static int grenzfunktion(int run)
```

```

60     {
61         return((int)(10*Math.pow(run,1.2)));
62     }
63
64     static void copy(File src, File dst) throws IOException
65     {
66         InputStream in = new FileInputStream(src);
67         OutputStream out = new FileOutputStream(dst);
68
69         byte[] buf = new byte[1024];
70         int len;
71         while ((len = in.read(buf)) > 0) out.write(buf, 0, len);
72         in.close();
73         out.close();
74     }
75 }
76
77

```

1.5 TransProbAutoCompareLexiconDelayed.java

```

1   import java.util.*;
2   import java.io.*;
3
4
5   public class TransProbAutoCompareLexiconDelayed {
6
7       static int len = 2;
8       static int maxlength = 2;
9       static int raster = 10;
10      static int run = 3;
11      static int delay = 2;
12
13      static double limit = 0.1;
14
15      static String thisData;
16
17      public static void main(String[] args)
18      {
19          boolean ipa = false;
20
21          for (int i = 1; i < args.length; i++)
22              {

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
23     switch(i)
24     {
25         case 1:
26             maxlength = (Integer.parseInt(args[i]));
27             break;
28         case 2:
29             raster = (Integer.parseInt(args[i]));
30             break;
31         case 3:
32             run = (Integer.parseInt(args[i]));
33             break;
34         case 4:
35             delay = (Integer.parseInt(args[i]));
36             break;
37         case 5: if("ipa".equals(args[i])) ipa = true; break;
38     }
39 }
40
41 String dataIn = IPA.readFile(args[0],false);
42 if(ipa) IPA.writeFile(args[0]+".ipa",
43     IPA.convertTextToIPAWhiteSpace(dataIn),ipa);
44 if(ipa) dataIn = IPA.convertTextToIPA(dataIn);
45 dataIn = dataIn.toLowerCase();
46 dataIn = removeWhiteSpace(dataIn);
47 int gaps[] = new int[dataIn.length()];
48
49 double limitStep = 1.0/raster;
50 for(int len=1;len<=maxlength;len++)
51 {
52     System.out.println("len "+len);
53     for(double limit=1;limit>=0;limit=java.lang.Math.round(1000*
54         (limit-limitStep))/1000.0)
55     {
56         System.out.println("limit "+limit);
57         String outFile = args[0]+". "+run+"_"+len+"_"+limit;
58         Vector phonemeChains1 = new Vector();
59         if(run>1 && (run>delay))
60         {
61             String chainsfile1 = args[0]+". "+(run-1)+"_"+len+"_"+
62                 limit+".lexicon";
63             String data = IPA.resubstituteDiphthongs(IPA.readFile
64                 (chainsfile1,ipa));
65             System.out.println("==>"+chainsfile1+" "+data.length());
66             String c[] = data.split("[^a-zA-Z"]+
```

```

67         String.copyValueOf(IPA.ipaChars)+"");
68     for(int i=0;i<c.length;i++)
69     {
70         for(int j=i+1;(i<c.length-1)&&(j<c.length);j++)
71         {
72             if(c[i].length()<c[j].length())
73             {
74                 String tmp = c[i];
75                 c[i]=c[j];
76                 c[j]=tmp;
77             }
78         }
79         phonemeChains1.addElement(c[i]);
80         //System.out.println("1 "+i+c[i]);
81     }
82 }
83
84
85 String dataOut = berechneTransProb(dataIn,len,limit,
86     phonemeChains1);
87 System.out.println("#"+dataOut);
88 //System.out.println("##"+
89     dataOut.charAt(dataOut.length()-1));
90
91 IPA.writeFile(outFile,dataOut,ipa);
92 if(ipa) IPA.writeFile(outFile+".substitute",
93     IPA.substituteDiphthongs(dataOut),true);
94 Vector compareResult = Compare(args[0],outFile,ipa);
95 if(limit==1)
96 {
97     try
98     {
99         FileWriter fw = new FileWriter(args[0]+". "+run+
100             ".result",len!=1);
101         BufferedWriter bw = new BufferedWriter(fw);
102         bw.write("laenge = "+len+"\n");
103         bw.close();
104         fw.close();
105     } catch(IOException e){}
106 }
107 try
108 {
109     FileWriter fw = new FileWriter(args[0]+". "+run+
110         ".result",true);

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
111         BufferedWriter bw = new BufferedWriter(fw);
112         bw.write(""+limit+" ");
113         for(int r=2;r>=0;r--)
114             bw.write(""+((Integer)compareResult.elementAt(r)).
115                                     intValue()+" ");
116         bw.write("\n");
117         bw.close();
118         fw.close();
119     } catch(IOException e){}
120 }
121 }
122 }
123
124 static String removeWhiteSpace(String data)
125 {
126     return(data.replaceAll("[^a-zA-ZäöüÄÖÜß"+
127                             String.valueOf(IPA.ipaChars)+""]",""));
128 }
129
130 static String berechneTransProb(String data,int len,
131                                 double limit,Vector phonemeChains1)
132 {
133     boolean marks[] = new boolean[data.length()];
134     for(int i=0;i<=data.length()-2*len;i++)
135     {
136         String sDouble = data.substring(i,i+2*len);
137         int index = 0;
138         int x = 0;
139         while((index = data.indexOf(sDouble,index))>=0)
140         {
141             x++;
142             index++;
143         }
144         String sSingle = data.substring(i,i+len);
145         index = 0;
146         int y = 0;
147         while((index = data.indexOf(sSingle,index))>=0)
148         {
149             y++;
150             index++;
151         }
152         if(x/(double)y<limit)
153         {
154             marks[i+len-1]=true;
```

```

155     }
156   }
157
158   for(int j=0;j<data.length()-1;j++)
159     //remove marks that are immediately followed by a mark
160   {
161     if(marks[j] && marks[j+1])
162       marks[j]=false;
163   }
164
165   String dataOut="";
166
167   for(int i=0;i<data.length();i++)
168   {
169     System.out.println(i);
170     int chains1 = -1;
171     for(int j=0;(j<phonemeChains1.size()) && (chains1<0);j++)
172     {
173       String chain = (String)phonemeChains1.elementAt(j);
174       if(data.regionMatches(i,chain,0,chain.length()))
175       {
176         chains1=j;
177       }
178     }
179     if(chains1>=0)
180     {
181       String chain = (String)phonemeChains1.elementAt(chains1);
182       System.out.println("1 "+ chain);
183       if((dataOut.length()>0) && (dataOut.charAt(dataOut.length()
184         -1)!=32)) dataOut=dataOut.concat(" ");
185       dataOut=dataOut.concat(chain+" ");
186       i+=chain.length()-1;
187     }
188     else
189     {
190       dataOut=dataOut.concat(data.substring(i,i+1));
191       if(marks[i])
192       {
193         dataOut=dataOut.concat(" ");
194       }
195     }
196   }
197
198   return(dataOut.trim());

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```
199     }
200
201
202     static String f[] = new String[2];
203     static int fpos[] = new int[2];
204     static int fmiss[] = new int[2];
205
206     static Vector Compare(String inFile, String outFile, boolean ipa)
207     {
208
209         f[0] = IPA.readFile(inFile,false);
210         f[1] = IPA.readFile(outFile,ipa);
211         fpos[0]=-1;
212         fpos[1]=-1;
213         fmiss[0]=0;
214         fmiss[1]=0;
215         if(ipa) f[0] = IPA.convertTextToIPAWhiteSpace(f[0]);
216
217         int hit=0;
218         while(((fpos[0] = nextLetter(0))>=0) &&
219             (fpos[1] = nextLetter(1))>=0)
220         {
221             //System.out.println(""+fpos[0]+" "+fpos[1]+" "+
222             //f[0].charAt(fpos[0])+" "+f[1].charAt(fpos[1]));
223             if(((fpos[0]+1<f[0].length())&&(fpos[1]+1<f[1].length())))
224             {
225                 if(isWhiteSpace(f[0].charAt(fpos[0]+1)))
226                 {
227                     if(isWhiteSpace(f[1].charAt(fpos[1]+1)))
228                     {
229                         hit++;
230                     }
231                     else
232                     {
233                         fmiss[0]++;
234                     }
235                 }
236                 else
237                 {
238                     if(isWhiteSpace(f[1].charAt(fpos[1]+1)))
239                     {
240                         fmiss[1]++;
241                     }
242                 }

```

```

243     }
244 }
245
246     Vector results = new Vector();
247     results.addElement(new Integer(fmiss[0]));
248     results.addElement(new Integer(fmiss[1]));
249     results.addElement(new Integer(hit));
250
251     return(results);
252 }
253
254     static boolean isWhiteSpace(char c)
255     {
256         return(!(IPA.isLetter(c) || IPA.isIPALetter(c)));
257     }
258     static int nextLetter(int i)
259     {
260         int pos = fpos[i]+1;
261         while((pos<f[i].length())&&(!IPA.isLetter(f[i].charAt(pos)))&&
262             (!IPA.isIPALetter(f[i].charAt(pos))))
263             pos++;
264         if(pos<f[i].length())
265             return(pos);
266         return(-1);
267     }
268 }

```

1.6 Homonyme und Heteronyme des CMU: Homohete.pl

```

1  open (IN, "<cmudict");
2  open (HETE, ">hete");
3  open (HOMO, ">homo");
4
5  %hete = ();
6  %homo = ();
7
8  while(<IN>)
9  {
10     $_ =~ /(.*) (.*)/;
11     $clear = $1;
12     $samba = $2;
13     $clear =~ /([\^\\(]*)/;
14     $clear = $1;

```

Anhang: Quellcodes, Wortlisten und Kodierungen

```

15     $samba =~ s/\d//g;
16     #print("$clear - $samba\n");
17     $hete{$clear}++;
18     $homo{$samba}++;
19 }
20
21 $heteCnt=0;
22 foreach $c (sort keys(%hete))
23 {
24     if($hete{$c}>1)
25     {
26         $heteCnt++;
27         print HETE (" $c  ".$hete{$c}."\n");
28     }
29 }
30
31 $homoCnt=0;
32 foreach $c (sort keys(%homo))
33 {
34     if($homo{$c}>1)
35     {
36         $homoCnt++;
37         print HOMO (" $c  ".$homo{$c}."\n");
38     }
39 }
40 print("$heteCnt Heteronyme.\n");
41 print("$homoCnt Homophone.\n");

```

2 Anhang B: Wortlisten für $\zeta = 10$ und $\zeta = 30$

$\zeta = 30$			$\zeta = 10$		
IPA	Normalform	Anz.	IPA	Normalform	Anz.
Korrekte Einträge im Paralexikon (nach <i>CMU Pronouncing Dictionary</i> -Abgleich)					
ju	you	553	ju	you	543
jɔr	your	533	jɔr	your	482
ənd	and	394	ənd	and	359
ðət	that	327	ðət	that	238
wʌt	what	276	hɪr	hear/here	189
hɪr	hear/here	268	mɒmi	mommy	185
mɒmi	mommy	232	wʌt	what	174
ðʌ	the	205	ðer	their	139

2 Anhang B: Wortlisten für $\zeta = 10$ und $\zeta = 30$

ðer	their	205	ðΛ	the	123
ðəts	that's	195	sΛm	some	88
tu	to	178	mərgΛn	morgan	87
get	get	154	ɔl	all	80
kən	can	150	ðəts	that's	79
sΛm	some	146	kən	can	67
mər	more	126	dɪlΛn	dhillon	66
mərgΛn	morgan	122	tu / two	to /two	65
ɔl	all	122	θɪŋ	thing	65
ðɪs	this	115	mændi	mandie	64
lets	let's	111	lets	let's	59
mændi	mandie	108	mər	more	59
θɪŋ	thing	107	get	get	47
fər	for	107	həv	have	47
ɒn	on	102	ɒn	on	43
həv	have	101	wi	we	40
ɔr	are	98	wətɜ	water	37
dɪlΛn	dhillon	96	əʊkeɪ	ok	35
ɪts	it's	86	ɪts	it's	34
lɪtl	little	82	ɔr	are	28
ɪt	it	67	jə	yeah	28
wətɜ	water	65	wer	where	25
wer	where	65	dʒʊs	juice	24
lʊk	look	61	kΛm	come	23
kΛm	come	59	lɪtl	little	23
wΛn	One / won	58	fər	for	22
wɒ	wa	58	dʊŋ	doing	22
Λn	an	55	lʊk	look	21
wi	we	54	tʃɪkΛn	chicken	18
si	see	52	ɪt	it	18
ʌp	up	46	red	red /read	16
ɔr	or	45	dəʊnt	don't	16
ɪz	is	43	Λn	an	15
nɒt	not	41	traɪ	try	15
hænd	hand	40	ɪz	is	14
jə	yeah	40	her	hair	13
gɒt	got	38	bək	back	13
wʌts	what's	37	pɪti	pretty	13
əʊkeɪ	ok	35	Λm	um	13
mɒmɪz	mommies/'s	34	wΛn	one /won	12
dʒʌst	just	32	si	see	12
gʊd	good	32	ʌp	up	12
dʊŋ	doing	30	kɪti	kitty	12
mi	me	30	wɒnt	want	12

Anhang: Quellcodes, Wortlisten und Kodierungen

red	red /read	30	hænd	hand	12
bək	back	24	gəʊ	go	11
tʃɪkən	chicken	24	bɛɪbi	baby	11
dʒʊs	juice	24	wɒ	wa	11
wɑnt	want	23	nɔɪz	noise	10
kju	cue	22	mɒmɪz	mommies/'s	10
ɪtəl	it'll	21	ðɪs	this	10
swɪti	sweetie	20	dʒʌst	just	10
bək	book	19	ʃu	shoe	9
drɪŋk	drink	19	tən	tan	9
let	let	19	hɪz	his	9
ʌm	um	18	bʌbəl	bubble	8
gɒtə	gotta	18	ʃʌt	shut	8
pʊt	put	18	nəʊ	no /know	8
wɒnə	wanna	18	ɒf	off	8
flɔr	floor	18	rəʊnd	round	8
plɪz	please	18			
gɒnə	gonna	18			
dəʊnt	don't	16			
li	lea	16			
bɔl	ball	16			
traɪ	try	15			
hɪz	his	15			
kʌp	cup	15			
bʊm	boom	14			
mɪlk	milk	14			
θɪŋk	think	13			
her	hair	13			
ən	ahn	13			
pɪtɪ	pretty	13			
tɪkəl	tickle	12			
kʊl	cool	12			
jʌmi	yummy	12			
kɪti	kitty	12			
du	do	11			
bɒtəl	bottle	11			
sɪŋ	sing	11			
kənt	can't	11			
ɪt	eat	10			
end	end	10			
səmi	sami	10			
nɔɪz	noise	10			
kɪt	kit	10			
ʃu	shoe	9			

2 Anhang B: Wortlisten für $\zeta = 10$ und $\zeta = 30$

dɒli	dalai()	9
bɹɔklɒn	brooklyn	9
dʌk	duck	9
ʌweɪ	away	9
tən	tan	9
fəʊn	phone	9
gəʊɪŋ	going	9
aɪd	I'd	8
kliːn	clean	8
əʊvə	over	8
gəʊ	go	8
bʌbəl	bubble	8
græmə	grandma	8
hʌni	honey	8
əpl	apple	8
raʊnd	round	8
ɔf	off	8
ʃʌt	shut	8
peɪdʒ	page	7
peəz	pairs	7
weəz	where's	7
pænd	panned	7
əʊkeɪ	ok	7
bɔlz	ball's/balls	7
bæskɪtbɔl	basketball	7
hɒt	hot	7
ʒəɡəʊ	jago	7
weɪ	way	7
aʊt	out	7
ðeəz	theirs	7
bɒtəm	bottom	7
kliːnɒp	cleanup	7
bæt	bat	7
bəʊnsɪŋ	bouncing	7
ləʊʃən	lotion	7
daʊnstəz	downstairs	7
seɪ	say	7
bʌbəlz	bubbles	7
ʌv	of	7
bart	byte/bite	7
nəʊ	now	7
edi	eadie	7
tʃɪriəʊs	cheerios	7
bɪɡ	big	6

Anhang: Quellcodes, Wortlisten und Kodierungen

bəʊl	bowl	6
kəʊld	cold	6
jus	use	6
nɪd	need	6
hi	he	6
dɪŋ	dring	6
kɪtʃɪn	kitchen	6
tʃɪriəʊ	cheerio	6
pleɪ	play	6
stɛəz	stairs	6
tʃiːz	cheese	6
bɜː	bir	6
θʌm	thumb	6
dædi	daddy	6
raɪt	right	6
daʊn	down	6
tə	to(3)	6
naɪt	night	6
ðiːz	these	6
aɪsi	icy / I see	6
skrʌb	scrub	6
re	reh	6
ɪtsi	itsy	6
dəʊnt	don't	6
tʃɪl	chill	6
gɜːl	girl	6
hələʊ	hello	6
aʊts	out's	6
taɪm	time	6
eg	egg	6
sɪt	sit	6
laɪk	like	6
nəʊ	know	6
bɪtsi	bitsy	6
ɔːlaɪt	alright	6
pɒp	pop	5
əʊnd	around	5
fɪʃ	fish	5
wɪð	with	5
nɪŋ	ning	5
raɪ	rye	5
daɪpɜː	diaper	5
haʊ	how	5
bɛɪbi	baby	5

brekfʌst	breakfast	5
hɪrz	hears	5
pɪg	pig	5
gəʊɪŋ	going	5
kɪstə	christa	5
bi	be	5
sʌn	son	5
bʌnənʌz	bananas	5
hʌg	hug	5
ðəm	them	5
əʊpən	open	5
gəʊ	go	5
ədi	addie	5
kɪs	kiss	5
mʌ	ma	5
twɪŋkl	twinkle	5
tiθ	teeth	5
bɛr	bear	5
maɪ	my	5

Problematisch, aber im *CMU Pronouncing Dictionary* enthalten

ɪŋ	ing	143	ɪŋ	ing	68
el	ehle	39			
dɪ	d	12			
eɪbi	ab	5			
er	aer	5			
ti	t	5			
le		5			

Zusammengesetzte Phrasen (mit interner Rekursion lösbar)

ɔrju	are you	109	ɔrju	are you	60
dʊju	do you	36	kʌmɒn	come on	14
kənju	can you	27	əʊju	oh you	13
kʌmhɪr	come here	19	kʌdlɪtedɪ	cuddly teddy	13
kʌmɒn	come on	19	nəʊju	no you	12
gʊdɡɜl	good girl	19	kənjuːsɪ	can you say	12
dɪdju	did you	20	draɪjɔr	dry your	12
lʊkət	look at	20	ðɒlbɔl	the ball	12
dʊjuwɒnt	do you want	18	ðɛrjuːgəʊ	there you go	12
kʊkʌdʊdʌldʊ		14	wɒtɔrjuːdʊɪŋ	what are you doing	11
əʊju	oh you	13	ðətsraɪt	that's right	10
kʌdlɪtedɪ	cuddly teddy	13	kənju	can you	10

Anhang: Quellcodes, Wortlisten und Kodierungen

nəʊju	no you	12
ðʌbəl	the ball	12
draɪjər	dry your	12
kənʤuseɪ	can you say	12
wɛrʤjər	where's your	10
siðʌ	see the	9
wʌtɔrʤudɪŋ	what are you doing	11
aɪgɔtʤu	I got you	9
bɪgtoʊ	big toe	9
aɪdəʊnt	I don't	8
mɔmɪzɡɔnʌ	mommy 's gonna	8
tedɪbɛr	teddy bear	8
wɔʃjər	watch your	8
jəjə	yeah yeah	8
juwɔnt	you want	7
juwɔnʌ	you wanna	7
ʤuseɪ	you say	7
ɡetsʌm	get some	7
ʌnðʌ	on the	7
sɪtdaʊn	sit down	7
ðɛrʤugəʊ	there you go	7
əndʌ	and a	7
ðətsraɪt	that's right	7
aɪgɔtʤu	I got you	7
ɡetʌ	get a	6
əndʤu	and you	6
aɪləʊvʤu	I love you	6
aməʊ	I no	6
ʌʃeɪpʌv	a shape of	6
ɡethɪm	get him	6
teɪkʌ	take a	6
spɑdʒdɪdʌn	spider did an	6
aɪləʊv	I love	6
braʊnbɛr	brown bear	5
pʊtʤɔr	put your	5
brʌʃjər	brush your	5
ʌbəl	a ball	5
ðɪsmɔr	this more	5
wɪðʌ	with a	5
hɪrsɪt	here sit	5
əʊju	oh you	5
wɪgəʊ	we go	7

Falsche Einträge

hə	82	hə	12
zjər	21	zənd	9
zɪt	14		
zənd	14		
husez	12		
zðʌ	8		
ʌnðer	8		
zðət	8		
ɔtɜ	8		
tðət	7		
ɔktʌp	7		
fʌl	7		
seɪb	7		
fɪŋ	7		
we	7		
zəndðʌ	6		
itam	6		
pʌvwɔtɜ	6		
tr	6		
puber	6		
uðʌ	6		
iwerzmənd	6		
brɪsk	6		
iʌl	6		
letst	6		
tʌd	6		
zɪtʌl	6		
dɪʌn	6		
ɪgʊd	5		
ɜfʌl	5		
gəʊldfɪ	5		
hɪ	6		
st	5		
ɜnðʌ	5		
rɪ	5		
lɪpɪbɪbɪ	5		

Anhang: Quellcodes, Wortlisten und Kodierungen

3 Anhang C: IPA und SAMBA Notationen

SAMBA	IPA	Hexadezimal
Vokale		
AA	ɒ	0252
AE	ə	0259
AH	ʌ	028C
AO	ɔ	0254
EH	e	0065
ER	ɜ	025C
IH	ɪ	026A
IY	i	0069
UH	ʊ	028A
UW	u	0075
Halbvokale		
W	w	0077
Y	j	006A
Diphtonge		
AW	aʊ	☺ → 263A
AY	aɪ	☹ → 263B
OW	əʊ	♀ → 2640
OY	ɔɪ	♂ → 2642
EY	eɪ	☼ → 263C
Konsonanten		
B	b	0062
P	p	0070
D	d	0064
T	t	0074
G	g	0067
K	k	006B
JH	dʒ	02A4
CH	tʃ	02A7
V	v	0076
F	f	0066
M	m	006D
N	n	006E
NG	ŋ	014B
DH	ð	00F0
TH	θ	03B8
S	s	0073
SH	ʃ	0283
Z	z	007A
ZH	ʒ	0292
HH	h	0068
L	l	006C
R	r	0072

Literaturverzeichnis

- Abney, Steven: Statistical methods and linguistics. In: Klavans, Judith L. und Resnik, Philip (Hg.) *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, MIT Press, Cambridge, S. 1–26. 1996.
- Albert, Marilyn, Diamond, Adele D., Holly Fitch, Roslyn, Neville, Helen J., Rapp, Peter R. und Tallal, Paula A.: Cognitive development. In: Zigmond, Michael J., Bloom, Floyd E., Roberts, James L., Landis, Story C. und Squire, Larry R. (Hg.) *Fundamental Neuroscience*, Academic Press, New York, S. 1313–1337. 1999.
- Altmann, Gerry T. M.: *The ascent of Babel: An exploration of language, mind, and understanding*. Oxford University Press, Oxford, 1997.
- Altmann, Gerry T. M. und Dienes, Zoltán: Technical comments: Rule learning by seven-month-old infants. In: *Science*, Band 284:S. 875, 1999.
- Anderson, John M.: *Linguistic representation: structural analogy and stratification*. Trends in linguistics / Studies and monographs, 67. Mouton de Gruyter, Berlin, 1992.
- Anderson, John R.: *Kognitive Psychologie*. Spektrum Akademischer Verlag, Heidelberg, 2001.
- Arbib, Michael A.: A guided tour of brain theory and neural networks. In: Arbib, Michael A. (Hg.) *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, Band Psychology, Linguistics, and Artificial Intelligence, S. 26–80. 2003.
- Aristoteles: *Über die Seele*. Felix Meiner Verlag, 1995.
- Aslin, Richard N. und Pisoni, David B.: Some developmental processes in speech perception. In: Yeni-Komishian, Grace H., Kavanagh, James F. und Ferguson, Charles A. (Hg.) *Child Phonology, Volume 2: Perception*, Academic Press, New York, Band 2, S. 67–96. 1980.
- Aslin, Richard N., Saffran, Jenny R. und Newport, Elissa L.: Computation of conditional probability statistics by 8-month-old infants. In: *Psychological Science*, Band 9(4):S. 321–324, 1998.
- Aslin, Richard N., Saffran, Jenny R. und Newport, Elissa L.: Statistical learning in linguistic and nonlinguistic domains. In: MacWhinney, Brian (Hg.) *The Emergence of Language*, Erlbaum, London, S. 359–380. 1999.

Literaturverzeichnis

- Aslin, Richard N., Woodward, Julide, LaMendola, Nicholas und Bever, Thomas: Models of word segmentation in fluent maternal speech to infants. In: Morgan, James und Demuth, Katherine (Hg.) *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, Erlbaum, Providence, S. 117–134. 1996.
- Baayen, Rolf Harald: On frequency, transparency and productivity. In: *Yearbook of Morphology, 1993*, Band 2:S. 181–208, 1992.
- Baayen, Rolf Harald und Schreuder, Robert: Towards a psycholinguistic computational model for morphological parsing. In: *Philosophical Transactions*, Band 358(1769):S. 1281–1292, 2000.
- Baroni, Marco: Distribution-driven morpheme discovery. In: *Yearbook of Morphology, 2003*, Band 13:S. 213–248, 2002.
- Baroni, Marco: Distributions in text. In: Lüdeling, A. und Kytö, M. (Hg.) *Corpus Linguistics: An international Handbook*, de Gruyter, Berlin. 2005.
- Batchelder, Eleanor O.: Bootstrapping the lexicon: A computational model of infant speech segmentation. In: *Cognition*, Band 83(2):S. 167–206, 2002.
- Bates, Elizabeth und Elman, Jeffrey: Learning rediscovered. In: *Science*, Band 274(5294):S. 1849–1850, 1996.
- Bates, Elizabeth und MacWhinney, Brian: A functionalist approach to the acquisition of grammar. In: *Functionalism in Linguistics*, S. 209–264, 1987.
- Bechtel, William: Contemporary connectionism: Are the new parallel distributed processing models cognitive or associationist? In: *Behaviorism*, Band 13(1):S. 53–61, 1985.
- Bechtel, William und Abrahamsen, Adele: *Connectionism and the mind: An Introduction to Parallel Processing in Networks*. Blackwell, Cambridge, 1991.
- Behrens, Heike: The input-output relationship in first language acquisition. In: *Language and Cognitive Processes*, Band 21(1-3):S. 2–24, 2006.
- Bendor, Daniel und Wang, Xiaoqin: The neuronal representation of pitch in primate auditory cortex. In: *Nature*, Band 436(7054):S. 1161–1165, 2005.
- Berk, Laura E.: *Entwicklungspsychologie*. Pearson Studium, München, 2005.
- Berns, Gregory S., Cohen, Jonathan D. und Mintun, Mark A.: Brain regions responsive to novelty in the absence of awareness. In: *Science*, Band 276(5316):S. 1272–1275, 1997.
- Bernstein Ratner, Nan: From signal to syntax: But what is the nature of the signal? In: Morgan, James und Demuth, Katherine (Hg.) *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, Erlbaum, Providence, S. 135–150. 1996.

- Bertoncini, Josiane und Mehler, Jacques: Syllables as units in infant speech perception. In: *Infant Behavior & Development*, Band 4:S. 247–260, 1981.
- Bertoncini, Josiane; Caroline Floccia; Thierry Nazzi; Jacques Mehler: Morae and syllables: Rhythmical basis of speech representations in neonates. In: *Language and Speech*, Band 38(4):S. 311–329, 1995.
- Bertoncini, Josiane; Ranka Bijeljac-Babic; Peter W. Jusczyk; Lori J. Kennedy; Jacques Mehler: An investigation of young infants' perceptual representations of speech sounds. In: *Journal of Experimental Psychology; General*, Band 117(1):S. 21–33, 1988.
- Best, Catherine T., McRoberts, Gerald W. und Sithole, Nomathemba M.: Examination of perceptual reorganization for non-native speech contrasts. In: *Journal of Experimental Psychology: Human Perception and Performance*, Band 14(3):S. 345–360, 1988.
- Best, Karl-Heinz: *Quantitative Linguistik*. Peust und Gutschmidt, Göttingen, 2003.
- Biber, Douglas: Representativeness in corpus design. In: *Linguistic and Literary Computing*, Band 8(4):S. 243–257, 1993.
- Biber, Douglas: Using register-diversified corpora for general language studies. In: *Using Large Corpora*, S. 179–201, 1994.
- Biber, Douglas: Representativeness in corpus design. In: Sampson, G. und McCarthy, D. (Hg.) *Corpus Linguistics: Readings in a Widening Perspective*, Continuum, London, Band 174-197. 2004.
- Bickerton, Derek: *Language and species*. University of Chicago Press, Chicago, 1991.
- Bijeljac-Babic, Ranka, Bertoncini, Josiane und Mehler, Jacques: How do 4-day-old infants categorize multisyllabic utterances? In: *Developmental Psychology*, Band 29(4):S. 711–721, 1993.
- Bischof, Norbert: *Struktur und Bedeutung*. Huber, Bern, 1995.
- Blakemore, Sarah-Jayne und Frith, Uta: *Wie wir lernen*. Dt. Verl.-Anst., München, 2006.
- Blevins, Juliette: The syllable in phonological theory. In: *The Handbook of Phonological Theory*, S. 206–244, 1995.
- Bloom, L.: The intentionality model of word learning: How to learn a word, any word. In: Golinkoff, R. M., Hirsh-Pasek, K., Akhtar, N., Bloom, L., Hollich, G. und Smith, L. (Hg.) *Becoming a Word Learner: A Debate on Lexical Acquisition*, Oxford University Press, New York, S. 19–50. 2001.
- Bortfeld, Heather: Which came first: Infants learning language or motherese? In: *Behavioral and Brain Sciences*, Band 27(4):S. 505, 2004.

Literaturverzeichnis

- Bortfeld, Heather, Morgan, James L. und Golinkoff, Roberta Michnick: Mommy and me. In: *Psychological Science*, Band 16(4):S. 298–304, 2005.
- Breitenstein, C. und Knecht, S.: Spracherwerb und Statistisches Lernen. In: *Der Nervenarzt*, Band 2:S. 133–143, 2003.
- Brent, Michael R.: From grammar to lexicon: Unsupervised learning of lexical syntax. In: *Computational Linguistics*, Band 19(2):S. 243–262, 1993.
- Brent, Michael R.: Advances in the computational study of language acquisition. In: Brent, Michael R. (Hg.) *Computational Approaches to Language Acquisition*, MIT Press, Cambridge, Band 1-38. 1997.
- Brent, Michael R.: An efficient, probabilistically sound algorithm for segmentation and word discovery. In: *Machine Learning*, Band 34(1-3):S. 71–105, 1999a.
- Brent, Michael R.: Speech segmentation and word discovery. In: *Trends in Cognitive Sciences*, Band 3(8):S. 294–301, 1999b.
- Brent, Michael R. und Cartwright, Timothy A.: Distributional regularity and phonotactic constraints are useful for segmentation. In: *Cognition*, Band 61(1-2):S. 93–125, 1996.
- Brent, Michael R. und Siskind, Jeffrey M.: The role of exposure to isolated words in early vocabulary development. In: *Cognition*, Band 81(2):S. B33–B44, 2001.
- Broadbent, Donald E.: Word-frequency effect and response bias. In: *Psychological Review*, Band 74(1):S. 1–15, 1967.
- Bronfenbrenner, Urie und Ceci, Stephen J.: Heredity, environment, and the question „how?“ – a first approximation. In: Plomin, Robert und McClearn, Gerald E. (Hg.) *Nature, Nurture, and Psychology*, American Psychological Association, Washington/DC, S. 313–324. 1993.
- Brooks, Roger A.: Intelligence without representation. In: *Artificial Intelligence*, Band 47:S. 139–159, 1991.
- Brown, Christian M.: Audition. In: Zigmond, Michael J., Bloom, Floyd E., Roberts, James L., Landis, Story C. und Squire, Larry R. (Hg.) *Fundamental Neuroscience*, Academic Press, New York, S. 791–820. 1999.
- Brown, Roger: *Words and things*. Free Press, New York, 1968.
- Brown, Roger: *A first language: The early stages*. Allen & Unwin, London, 1973.
- Bruner, Jerome S.: *Wie das Kind sprechen lernt*. Huber, Bern, 1987.
- Bußmann, Hadumod: *Lexikon der Sprachwissenschaft*. Kröner, Stuttgart, 2002.

- Butzkamm, Wolfgang und Butzkamm, Jürgen: *Wie Kinder sprechen lernen*. Francke, Tübingen, 2004.
- Bybee, Joan und Hopper, Paul: Introduction to frequency and the emergence of linguistic structure. In: Bybee, Joan und Hopper, Paul (Hg.) *Frequency and the Emergence of Linguistic Structure*, John Benjamins Publishing Company, Amsterdam, S. 1–24. 2001.
- Cairns, Paul, Shillcock, Richard, Chater, Nick und Levy, Joe: Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. In: *Cognitive Psychology*, Band 33(2):S. 111–153, 1997.
- Calvin, William H. und Ojemann, George A.: *Einsicht ins Gehirn*. Dt. Taschenbuch-Verl., München, 2000.
- Caplan, David, Carr, Thomas, Gould, James und Martin, Randi: Language and communication. In: Zigmond, Michael J., Bloom, Floyd E., Roberts, James L., Landis, Story C. und Squire, Larry R. (Hg.) *Fundamental Neuroscience*, Academic Press, New York, S. 1487–1519. 1999.
- Carlyon, Robert P.: How the brain separates sounds. In: *Trends in Cognitive Sciences*, Band 8(10):S. 465–471, 2004.
- Carpenter, Gail A. und Grossberg, Stephen: Adaptive resonance theory. In: Arbib, Michael A. (Hg.) *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge. 2003.
- Carroll, John Bissell: *Word retrieval latencies as a function of frequency and age-of-acquisition priming, repeated trial, and individual differences*. Educational Testing Service, Princeton, 1976.
- Carstensen, Kai-Uwe, Ebert, Christian, Endriss, Cornelia, Jekat, Susanne, Klabunde, Ralf und Langer, Hagen: *Computerlinguistik und Sprachtechnologie*. Elsevier, München, 2004.
- Chomsky, Noam: *Syntactic structures*. Mouton de Gruyter, Den Haag, 1957.
- Chomsky, Noam: *Aspects of the theory of syntax*. MIT Press, Cambridge, 1965.
- Chomsky, Noam: *Reflexionen über die Sprache*. Suhrkamp, Frankfurt am Main, 1977.
- Chomsky, Noam: *Rules and representations*. Columbia University Press, New York, 1980.
- Chomsky, Noam: *Language and problems of knowledge*. MIT Press, Cambridge, 1988.
- Chomsky, Noam und Halle, Morris: *The sound pattern of English*. Harper & Row, New York, 1968.

Literaturverzeichnis

- Chomsky, Noam A.: A review of B. F. Skinner's verbal behavior. In: *Language*, Band 35(1):S. 26–58, 1959.
- Christiansen, Joseph Allen und Seidenberg, Mark S.: Learning to segment speech using multiple cues: A connectionist model. In: *Language and Cognitive Processes*, Band 13(2-3):S. 221–268, 1998.
- Christiansen, Morten H. und Curtin, Suzanne: Transfer of learning: rule acquisition or statistical learning? In: *Trends in Cognitive Sciences*, Band 3(8):S. 289–290, 1999.
- Christophe, Anne, Dupoux, Emmanuel und Bertoncini, Josiane: Do infants perceive word boundaries? In: *Journal of the Acoustical Society of America*, Band 95(3):S. 1570–1580, 1994.
- Church, Kenneth W.: Phonological parsing and lexical retrieval. In: *Cognition*, Band 25(1-2):S. 53–69, 1987.
- Church, Kenneth W.; Robert L. Mercer: Introduction to the special issue on computational linguistics using large corpora. In: *Computational Linguistics*, Band 19(1):S. 1–24, 1993.
- Clahsen, Harald: Lexical entries and rules of language. In: *Behavioral and Brain Sciences*, Band 22(6):S. 991–1060, 1999.
- Clahsen, Harald und Rothweiler, Monika: Inflectional rules in children's grammars. In: *Yearbook of Morphology, 1993*, Band 2:S. 1–34, 1992.
- Clark, Eve V.: *First language acquisition*. Cambridge University Press, Cambridge, 2003.
- Clark, Eve V.: How language acquisition builds on cognitive development. In: *Trends in Cognitive Sciences*, Band 8(10):S. 472–478, 2004.
- Clark, Robin, Gleitman, Lila und Kroch, Anthony: Acquiring language. In: *Science*, Band 276:S. 1179, 1997.
- Clements, G. N. und Hume, Elizabeth V.: The internal organization of speech sounds. In: Goldsmith, John A. (Hg.) *The Handbook of Phonological Theory*, Blackwell, Cambridge, S. 243–306. 1995.
- Coady, Jeffrey A. und Aslin, Richard N.: Young children's sensitivity to probabilistic phonotactics in the developing lexicon. In: *Journal of Experimental Child Psychology*, Band 89(3):S. 183–213, 2004.
- Collins, John: Cowie on the poverty of stimulus. In: *Synthese*, Band 136(2):S. 159–190, 2003.
- Connine, Cynthia M.: It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. In: *Psychonomic Bulletin & Review*, Band 11(6):S. 1084–1089, 2004.

- Cowie, Fiona: *What's within? Nativism reconsidered*. Oxford University Press, New York, 1999.
- Crain, Stephen und Pietroski, Paul M.: Nature, nurture and universal grammar. In: *Linguistics and Philosophy*, Band 24(2):S. 139–186, 2001.
- Crain, Stephen und Pietroski, Paul M.: Why language acquisition is a snap. In: *The Linguistic Review*, Band 19:S. 163–183, 2002.
- Culicover, Peter W.: *Principles and parameters*. Oxford University Press, Oxford, 1997.
- Curtin, Suzanne und Werker, Janet F.: The perceptual foundations of phonological development. In: Gaskell, Gareth (Hg.) *The Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford. 2007.
- Cutler, Anne: Phoneme-monitoring reaction as a function of preceding intonation contour. In: *Perception & Psychophysics*, Band 20(1):S. 55–60, 1976.
- Cutler, Anne: Exploiting prosodic probabilities in speech segmentation. In: Altmann, Gerry T.M. (Hg.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, MIT Press, Cambridge, S. 105–121. 1995.
- Cutler, Anne und Carter, David M.: The predominance of strong initial syllables in the English vocabulary. In: *Computer Speech and Language*, Band 2(3-4):S. 133–142, 1987.
- Cutler, Anne und Clifton, Charles: Comprehending spoken language: A blueprint of the listener. In: Brown, Colin M. und Hagoort, Peter (Hg.) *The Neurocognition of Language*, Oxford University Press, Oxford, S. 123–166. 2000.
- Cutler, Anne, McQueen, James M. und Norris, Dennis: The roll of the silly ball. In: *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, S. 181–194, 2001.
- Cutler, Anne, Mehler, Jacques und Norris, Dennis: The syllable's differing role in the segmentation of French and English. In: *Journal of Memory and Language*, Band 25(4):S. 385–400, 1986.
- Cutler, Anne, Mehler, Jacques, Norris, Dennis und Segui, Juan: A language-specific comprehension strategy. In: *Nature*, Band 304:S. 159–160, 1983.
- Cutler, Anne und Norris, Dennis: The role of strong syllables in segmentation for lexical access. In: *Journal of Experimental Psychology: Human Perception and Performance*, Band 14(1):S. 113–121, 1988.
- Cutler, Anne; Takashi Otake: Mora or phoneme – further evidence for language-specific listening. In: *Journal of Memory and Language*, Band 33(6):S. 824–844, 1994.

Literaturverzeichnis

- Darwin, Christopher J.: Auditory processing and speech perception. In: Bourma, Herman und Bouwhuis, Don G. (Hg.) *Attention and Performance X: Control of Language Processes*, Erlbaum, London, S. 197–209, 1984.
- Davis, Matthew H.: Connectionist modelling of lexical segmentation and vocabulary acquisition. In: Quinlan, Philip (Hg.) *Connectionist Models of Development*, Psychology Press, Hove, 2003.
- de Marcken, Carl: *Unsupervised Language Acquisition*. Dissertationsschrift, MIT, Cambridge, 1996.
- Dehaene-Lambertz, Ghislaine und Baillet, S.: A phonological representation in the infant brain. In: *Neuroreport*, Band 9(8):S. 1885–1888, 1998.
- Dehaene-Lambertz, Ghislaine und Gliga, T.: Common neural basis for phoneme processing in infants and adults. In: *Journal of Cognitive Neuroscience*, Band 16(8):S. 1375–1387, 2004.
- Dehaene-Lambertz, Ghislaine, Hertz-Pannier, Lucie und Dubois, Jessica: Nature and nurture in language acquisition: Anatomical and functional brain-imaging studies in infants. In: *Trends in Neurosciences*, Band 29(7):S. 367–373, 2006.
- Descartes, René: *Abhandlung über die Methode, die Vernunft richtig zu gebrauchen*. Marixverlag, Wiesbaden, 2006.
- Deshmukh, Neeraj, Duncan, Richard J., Ganapathiraju, Aravind und Picone, Joseph: Benchmarking human performance for continuous speech recognition. In: *Proceedings*, Band 4th ICSLP:S. 2486–2489, 1996.
- Devlin, Keith J.: *Das Mathe-Gen*. Dt. Taschenbuch-Verl., München, 2004.
- Dietrich, Rainer: *Psycholinguistik*. Metzler, Stuttgart, 2007.
- Dijkstra, Ton; Koenraad de Smedt: *Computational Psycholinguistics*. Taylor & Francis, London, 1996.
- Dirven, Vilém, René; Fried: *Functionalism in Linguistics*, Band 20 von *Linguistic & Literary Studies in Eastern Europe*. John Benjamins, Amsterdam, 1987.
- Dittmann, Jürgen: *Der Spracherwerb des Kindes*. Beck, München, 2006.
- Dominey, Peter F. und Dodane, Christelle: Indeterminacy in language acquisition. In: *Journal of Neurolinguistics*, Band 17(2-3):S. 121–145, 2004.
- Dominey, Peter F.; Franck Ramus: Neural network processing of natural language: I. sensitivity to serial, temporal and abstract structure of language in the infant. In: *Language and Cognitive Processes*, Band 15(1):S. 87–127, 2000.

- Echols, Catharine H., Crowhurst, Megan J. und Childers, Jane B.: The perception of rhythmic units in speech by infants and adults. In: *Journal of Memory and Language*, Band 36(2):S. 202–225, 1997.
- Eimas, Peter D.: Speech perception in early infancy. In: Cohen, Leslie B. und Salaptek, Philip (Hg.) *Infant Perception: From Sensation to Cognition: Perception of Space, Speech, and Sound*, Academic Press, New York, Band 2, S. 193–231. 1975.
- Eimas, Peter D.: Do infants learn grammar with algebra or statistics? In: *Science*, Band 284(5413):S. 435–436, 1999.
- Eimas, Peter D., Siqueland, Einar R., Jusczyk, Peter W. und Vigorito, James: Speech perception in infants. In: *Science*, Band 171:S. 303–306, 1971.
- Elman, Jeffrey L.: Connectionist approaches to acoustic/phonetic processing. In: Marslen-Wilson, William (Hg.) *Lexical Representation and Process*, MIT Press, Cambridge, S. 227–260. 1989.
- Elman, Jeffrey L.: Finding structure in time. In: *Cognitive Science*, Band 14(2):S. 179–211, 1990.
- Elman, Jeffrey L.: The emergence of language: A conspiracy theory. In: McWhinney, Brain (Hg.) *The Emergence of Language*, Erlbaum, London, S. 1–27. 1999.
- Elman, Jeffrey L.; Elizabeth A. Bates: Acquiring language. In: *Science*, Band 276(5316):S. 1175–1183, 1276, 1997.
- Elman, L., Jeffrey, Bates, Elizabeth A., Johnson, Mark H., Karmiloff-Smith, Annette, Parisi, Domenico und Plunkett, Kim: *Rethinking innateness*. MIT Press, Cambridge, 1997.
- Elsen, Hilke: *Erstspracherwerb*. Deutscher Universitätsverlag, Wiesbaden, 1991.
- Engfer, Hans-Jürgen: *Empirismus versus Rationalismus?* Schöningh, Paderborn, 1996.
- Ewert, Jörg-Peter: *Neurobiologie des Verhaltens*. Huber, Bern, 1998.
- Falk, Dean: Prelinguistic evolution in early hominins. In: *Behavioral and Brain Sciences*, Band 27(4):S. 491–541, 2004.
- Farghaly, Ali: *Handbook for language engineers*. CSLI Publ., Stanford, 2003.
- Fetzer, James H.: *Computers and cognition*. Kluwer, Dordrecht, 2001.
- Fischer-Jørgensen, Eli: What can the new techniques of acoustic phonetics contribute to linguistics? In: Salporta, Sal (Hg.) *Psycholinguistics: A Book of Readings*, Holt, Rinehart and Winston, New York, S. 112–142. 1961.

Literaturverzeichnis

- Fitzpatrick, Jennifer und Wheeldon, Linda: Phonology and phonetics in psycholinguistic models of speech perception. In: *Phonological Knowledge*, S. 131–160, 2000.
- Fodor, Janet Dean und Crowther, Carrie: Understanding stimulus poverty arguments. In: *The Linguistic Review*, Band 19(1):S. 105–146, 2002.
- Fodor, Jerry A.: *The modularity of mind*. MIT Press, Cambridge, 1983.
- Fodor, Jerry A.: *The mind doesn't work that way*. MIT Press, Cambridge, 2001.
- Fodor, Jerry A. und Pylyshyn, Zenon W.: Connectionism and cognitive architecture. In: *Cognition*, Band 28(1-2):S. 3–71, 1988.
- Forster, Kenneth I.: Accessing the mental lexicon. In: Wales, R. J. und Walker, Edward C. T. (Hg.) *New Approaches to Language Mechanisms*, North-Holland Publishing Company, Amsterdam, S. 257–287. 1976.
- Forster, Kenneth I.: Levels of processing and the structure of the language processor. In: Cooper, William E. und Walker, Edward C. T. (Hg.) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, Erlbaum, Hillsdale, S. 27–85. 1979.
- Foss, Donald J. und Blank, Michelle A.: Identifying the speech codes. In: *Cognitive Psychology*, Band 12(1):S. 1–31, 1980.
- Foss, Donald J. und Gernsbacher, Morton Ann: Cracking the dual code. In: *Journal of Verbal Learning and Verbal Behavior*, Band 22(6):S. 609–632, 1983.
- Foss, Donald J. und Lynch, R. H.: Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. In: *Journal of Verbal Learning and Verbal Behavior*, Band 8(4):S. 145–148, 1969.
- Foss, Donald J. und Swinney, David A.: On the psychological reality of the phoneme: Perception, identification, and consciousness. In: *Journal of Verbal Learning and Verbal Behavior*, Band 12(3):S. 246–257, 1973.
- Frauenfelder, Uli H. und Floccia, Caroline: The recognition of spoken words. In: *Language Comprehension*, S. 1–40, 1999.
- Friederici, Angela D. und Wessels, Jeanine M. I.: Phonotactic knowledge of word boundaries and its use in infant speech perception. In: *Perception & Psychophysics*, Band 54(3):S. 287–295, 1993.
- Gambell, Timothy und Yang, Charles D.: Scope and limits of statistical learning in word segmentation. 2003.
- Gambell, Timothy und Yang, Charles D.: Statistics learning and universal grammar: Modeling word segmentation. In: Sakas, W. G. (Hg.) *COLING 2004: Psycho-Computational Models of Human Language Acquisition*, COLING, New York, S. 51–53. 2004.

- Gambell, Timothy und Yang, Charles D.: Mechanisms and constraints in word segmentation. ms, 2005.
- Gardner, Howard: *Dem Denken auf der Spur*. Klett-Cotta, Stuttgart, 1989.
- Gaskell, Gareth M.: Statistical and connectionist models of speech perception and word recognition. In: Gaskell, Gareth M. (Hg.) *The Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford, S. 55–69. 2007.
- Gegenfurtner, Karl R.: *Gehirn and Wahrnehmung*. Fischer-Taschenbuch-Verlag, Frankfurt am Main, 2005.
- Gelman, Rochel und Cordes, Sara: Counting in animals and humans. In: Dupoux, Emmanuel (Hg.) *Language, Brain, and Cognitive Development*, MIT Press, Cambridge, S. 279–301. 2001.
- Gentner, Dedre und Medina, José: Similarity and the development of rules. In: Sloman, Steven A. und Rips, Lance J. (Hg.) *Similarity and Symbols in Human Thinking*, MIT Press, Cambridge, S. 177–211. 1998.
- Gerken, LouAnn und Aslin, Richard N.: Thirty years of research on infant speech perception. In: *Language Learning and Development*, Band 1(1):S. 5–21, 2005.
- Gleitman, Lila: The structural sources of verb meanings. In: *Language Acquisition*, Band 1(1):S. 3–55, 1990.
- Gleitman, Lila R. und Wanner, Eric: The state of the art of the state of the art. In: Wanner, Eric und Lila R, Gleitman (Hg.) *Language acquisition: The state of the art*, Cambridge, University Press, Cambridge,, S. 3–48. 1982.
- Gómez, Rebecca L.: Statistical learning in infant language development. In: Gaskell, Gareth M., Altmann, Gerry T. M. und Bloom, Paul (Hg.) *The Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford, S. 601–616. 2007.
- Gómez, Rebecca L. und Gerken, LouAnn: Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. In: *Cognition*, Band 70(2):S. 109–135, 1999.
- Gold, E. Mark: Language identification in the limit. In: *Information and Control*, Band 10:S. 447–474, 1967.
- Goldinger, Stephen D.: Words and voices. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Band 22(5):S. 1166–1183, 1996.
- Goldinger, Stephen D.: *Talker variability in speech processing*. Talker Variability in Speech Processing. Academic Press, San Diego, 1997.
- Goldinger, Stephen D.: Echoes of echoes? In: *Psychological Review*, Band 105(2):S. 251–279, 1998.

Literaturverzeichnis

- Goldinger, Stephen D. und Azuma, T.: Puzzle-solving science: The quixotic quest for units in speech perception. In: *Journal of Phonetics*, Band 31(3-4):S. 305–320, 2003.
- Goldinger, Steven D., Pisoni, David B. und Logan, John S.: On the nature of talker variability effects on recall of spoken word lists. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Band 17(1):S. 152–162, 1991.
- Goldsmith, John A.: Unsupervised learning of the morphology of a natural language. In: *Computational Linguistics*, Band 27(2):S. 153–198, 2001.
- Golinkoff, Roberta Michnick und Hirsh-Pasek, Kathy: *How babies talk*. Plume Books, New York, 2000.
- Goodsitt, Jan V., Morgan, James L. und Kuhl, Patricia K.: Perceptual strategies in prelingual speech segmentation. In: *Journal of Child Language*, Band 20(2):S. 229–252, 1993.
- Gordon, Ian und Slater, Alan: Nativism and empiricism: The history of two ideas. In: Slater, Alan, Hainline, Louise und Moseley, Merrick J. (Hg.) *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, Psychology Press, Hove, Band 73-103. 1998.
- Goswami, Usha: *Analogical reasoning in children*. Erlbaum, Hove, 1992.
- Grossberg, Stephen: How does a brain build a cognitive code? In: *Psychological Review*, Band 87(1):S. 1–51, 1980.
- Görz, Günther und Nebel, Bernhard: *Künstliche Intelligenz*. Fischer-Taschenbuch-Verlag, Frankfurt am Main, 2003.
- Gummer, A. W. und Zenner, Hans-Peter: Central processing of auditory information. In: *Comprehensive Human Physiology*, Springer, Berlin, S. 729–756. 1996.
- Haase, Frank: Abschied von der Medienkompetenz. In: Schade, Sigrid, Sieber, Thomas und Tholen, Georg Christoph (Hg.) *Schnittstellen*, Schwabe, Basel, Band 121-131. 2005.
- Halle, Morris: Review of manual of phonology. In: *Journal of the Acoustical Society of America*, Band 28:S. 509–510, 1956.
- Harnish, Robert M.: *Minds, brains, computers*. Blackwell, Malden, 2002.
- Harris, John und Lindsey, Geoff: There is no level of phonetic representation. In: *UCL Working Papers in Linguistics*, University College London, Band 5, S. 355–373. 1993.
- Harris, Zellig S.: Distributional structure. In: *Word*, Band 10:S. 146–162, 1954.
- Harris, Zellig S.: From phoneme to morpheme. In: *Language*, Band 31(2):S. 190–222, 1955.

- Harris, Zellig S.: *Morpheme Boundaries Within Words: Report on a Computer Test, Transformations and Discourse Analysis Papers 73*, Band 1 von *Papers in Structural and Transformational Linguistics*. University of Pennsylvania, Philadelphia, 1967.
- Hasher, Lynn und Chromiak, Walter: The processing of frequency information: an automatic mechanism? In: *Journal of Verbal Learning and Verbal Behavior*, Band 16(2):S. 173–184, 1977.
- Hasher, Lynn und Zacks, Rose T.: Automatic processing of fundamental information: The case of frequency of occurrence. In: *American Psychologist*, Band 39(12):S. 1372–1388, 1984.
- Hasher, Lynn, Zacks, Rose T. und Rose, Karen C.: Truly incidental encoding of frequency information. In: *The American Journal of Psychology*, Band 100(1):S. 69–91, 1987.
- Haspelmath, Martin: Frequency vs. iconicity in explaining grammatical asymmetries. In: *Cognitive Linguistics*, Band 19(1):S. 1–33, 2008.
- Hauser, Marc D.: What's so special about speech? In: Dupoux, Emmanuel (Hg.) *Language, Brain, and Cognitive Development*, MIT Press, Cambridge, S. 417–433. 2001.
- Hayes, John R. und Clark, Herbert H.: Experiments on the segmentation of an artificial speech analogue. In: Hayes, John R. (Hg.) *Cognition and the Development of Language*, John Wiley & Son, New York, S. 221–234. 1970.
- Healy, Alice F. und Cutting, James E.: Units of speech perception: phoneme and syllable. In: *Journal of Verbal Learning and Verbal Behavior*, Band 15(1):S. 73–83, 1976.
- Hebb, Donald Olding: *The Organization of Behavior*. Erlbaum Associates, 2002.
- Hecox, Kurt: Electrophysiological correlates of human auditory development. In: Cohen, Leslie B. und Salaptek, Philip (Hg.) *Infant Perception: From Sensation to Cognition: Perception of Space, Speech, and Sound*, Academic Press, New York, Band 2, S. 151–191. 1975.
- Hickok, Gregory und Poeppel, David: Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. In: *Cognition*, Band 92(1-2):S. 67–99, 2004.
- Hillenbrand, James M.: Perceptual organization of speech sounds by infants. In: *Journal of Speech and Hearing Research*, Band 26(2):S. 268–282, 1983.
- Hillenbrand, James M.: Speech perception by infants: Categorization based on nasal consonant place of articulation. In: *Journal of the Acoustical Society of America*, Band 75:S. 1613–1622, 1984.
- Hillenbrand, James M.: Perception of feature similarities by infants. In: *Journal of Speech and Hearing Research*, Band 28:S. 317–318, 1985.

Literaturverzeichnis

- Hirsh-Pasek, Kathy, Kemler Nelson, Deborah G. und Jusczyk, Peter W.: Clauses are perceptual units for young infants. In: *Cognition*, Band 26(3):S. 269–286, 1987.
- Hockett, Charles F.: *A manual of phonology*. Indiana University Press, Bloomington, 1955.
- Hohne, Elizabeth A. und Jusczyk, Peter W.: 2-month-old infants sensitivity to allophonic differences. In: *Perception & Psychophysics*, Band 56(6):S. 613–623, 1994.
- Hollich, George, Newman, Rochelle S. und Jusczyk, Peter W.: Infants' use of synchronized visual information to separate streams of speech. In: *Child Development*, Band 76(3):S. 598–613, 2005.
- Holyoak, Keith J. und Thagard, Paul: *Mental leaps*. MIT Press, Cambridge, 1995.
- Hopfield, John J.: Neural networks and physical systems with emergent collective computational abilities. In: *Proceedings of the National Academy of Sciences of the United States of America*, Band 79:S. 2554–2558, 1982.
- Householder, Fred W.: *Linguistic Speculations*. Cambridge University Press, Cambridge, 1971.
- Houston, Derek M.: Speech perception in infants. In: *The Handbook of Speech Perception*, S. 417–448, 2005.
- Houston, Derek M., Jusczyk, Peter W., Kuijpers, Cecile, Coolen, Riet und Cutler, Anne: Cross-language word segmentation by 9-month-olds. In: *Psychonomic Bulletin & Review*, Band 7(3):S. 504–509, 2000.
- Houston, Derek M., Santelmann, Lynn und Jusczyk, Peter W.: English-learning infants' segmentation of trisyllabic words from fluent speech. In: *Language and Cognitive Processes*, Band 19(1):S. 97–136, 2004.
- Hüther, Gerald und Krens, Inge: *Das Geheimnis der ersten neun Monate*. Walter, Düsseldorf, 2005.
- Hume, David: *A treatise of human nature*. Everyman's Library, London, 1968.
- Hume, David: *Enquiries concerning human understanding and concerning the principles of morals*. Clarendon Press, Oxford, 2007.
- Huttenlocher, Peter R. und Dabholkar, Arun S.: Regional differences in synaptogenesis in human cerebral cortex. In: *The Journal of Comparative Neurology*, Band 387:S. 167–178, 1997.
- Hövel, Jörg auf dem: *Abenteuer künstliche Intelligenz*. Discorsi, Hamburg, 2002.
- Ingram, David: *First language acquisition*. Cambridge University Press, Cambridge, 1989.

- Ingram, John C.L.: *Neurolinguistics*. Cambridge University Press, Cambridge, 2007.
- Jackendoff, Ray: *in the mind: Language and human nature*. BasicBooks, New York, 1996.
- Jacobs, Arthur M.: Simulative Methoden. In: Rickheit, Gert, Herrmann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Band 24 von *Handbücher zur Sprach- und Kommunikationswissenschaft*, S. 125–142. 2003.
- Jakobson, Roman: *Child language aphasia and phonological universals*. Mouton, The Hague, 1972.
- Jakobson, Roman, Fant, C. G. und Halle, Morris: *Preliminaries to speech analysis*. MIT Press, Cambridge, 1967.
- Jenkins, Lyle und Maxam, Allan: Acquiring language. In: *Science*, Band 276:S. 1179, 1997.
- Johnson, Elizabeth K. und Jusczyk, Peter W.: Word segmentation by 8-month-olds: When speech cues count more than statistics. In: *Journal of Memory and Language*, Band 44(4):S. 548–567, 2001.
- Johnson, Elizabeth K., Jusczyk, Peter W., Cutler, Anne und Norris, Dennis: Lexical viability constraints on speech segmentation by infants. In: *Cognitive Psychology*, Band 46(1):S. 65–97, 2003.
- Jurafsky, Daniel und Martin, James H.: *Speech and language processing*. Prentice Hall, New Jersey, 2000.
- Jusczyk, Peter W.: Auditory versus phonetic coding of speech signals during infancy. In: Mehler, Jacques, Walker, Edward C. T. und Garrett, Merrill (Hg.) *Perspectives on Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities*, Erlbaum, London, S. 339–360. 1982.
- Jusczyk, Peter W.: *The discovery of spoken language*. MIT Press, Cambridge, 1997.
- Jusczyk, Peter W.: Constraining the search for structure in the input. In: *Lingua*, Band 106(1-4):S. 197–218, 1998.
- Jusczyk, Peter W.: How infants begin to extract words from speech. In: *Trends in Cognitive Sciences*, Band 3(9):S. 323–328, 1999.
- Jusczyk, Peter W.: Learning language. In: Dupoux, Emmanuel (Hg.) *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, MIT Press, Cambridge, S. 363–377. 2001.

Literaturverzeichnis

- Jusczyk, Peter W.: Some critical developments in acquiring native language sound organization during the first year. In: *Annals of Otology Rhinology and Laryngology*, Band 111(5):S. 11–15, 2002.
- Jusczyk, Peter W. und Aslin, Richard N.: Infants' detection of the sound patterns of words in fluent speech. In: *Cognitive Psychology*, Band 29(1):S. 1–23, 1995.
- Jusczyk, Peter W., Cutler, Anne und Redanz, Nancy J.: Infants' preference for the predominant stress patterns of english words. In: *Child Development*, Band 64(3):S. 675–687, 1993a.
- Jusczyk, Peter W. und Derrah, Carolyn: Representation of speech sounds by young infants. In: *Developmental Psychology*, Band 23(5):S. 648–654, 1987.
- Jusczyk, Peter W., Friederici, Angela D. und Wessels, Jeanine M. I.: Infants' sensitivity to the sound patterns of native language words. In: *Journal of Memory and Language*, Band 32(3):S. 402–420, 1993b.
- Jusczyk, Peter W., Goodman, Mara B. und Baumann, Angela: Nine-month-olds' attention to sound similarities in syllables. In: *Journal of Memory and Language*, Band 40(1):S. 62–82, 1999a.
- Jusczyk, Peter W. und Hohne, Elizabeth A.: Infants' memory for spoken words. In: *Science*, Band 277:S. 1984–1986, 1997.
- Jusczyk, Peter W., Hohne, Elizabeth A. und Baumann, Angela: Infants' sensitivity to allophonic cues for word segmentation. In: *Perception & Psychophysics*, Band 61(8):S. 1465–1476, 1999b.
- Jusczyk, Peter W., Houston, Derek M. und Goodman, Mara B.: Speech perception during the first year. In: Slater, Alan (Hg.) *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, Psychology Press, Hove, S. 357–388. 1998.
- Jusczyk, Peter W., Houston, Derek M. und Newsome, Mary R.: The beginnings of word segmentation in english-learning infants. In: *Cognitive Psychology*, Band 39(3-4):S. 159–207, 1999c.
- Jusczyk, Peter W., Kennedy, Lori J. und Jusczyk, Ann M.: Young infants' retention of information about syllables. In: *Infant Behavior and Development*, Band 18:S. 27–41, 1995.
- Jusczyk, Peter W. und Luce, Paul A.: Speech perception and spoken word recognition: Past and present. In: *Ear and Hearing*, Band 23(1):S. 2–40, 2002.
- Jusczyk, Peter W., Luce, Paul A. und Luce, Jan Charles: Infants sensitivity to phonotactic patterns in the native language. In: *Journal of Memory and Language*, Band 33(5):S. 630–645, 1994.

- Jusczyk, Peter W. und Thompson, Elizabeth: Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. In: *Perception & Psychophysics*, Band 23(2):S. 105–109, 1978.
- Kandel, Eric R.: *Auf der Suche nach dem Gedächtnis: die Entstehung einer neuen Wissenschaft des Geistes*. Siedler, München, 2006.
- Karmiloff Smith, Annette: *A functional approach to child language: A study of determiners and reference*. Cambridge Studies in Linguistics, 24. Cambridge University Press, Cambridge, 1979.
- Karmiloff, Kyra und Karmiloff-Smith, Annette: *Pathways to language: from fetus to adolescent*. Harvard University Press, Cambridge, 2001.
- Karmiloff-Smith, Annette: *Beyond modularity: A developmental perspective on cognitive science*. MIT Press, Cambridge, 1995.
- Keil, Frank: Nativism. In: Wilson, Robert A. und Keil, Frank C. (Hg.) *The MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge, S. 583–586. 2001.
- Kemp, David T.: Stimulated acoustic emissions from within the human auditory system. In: *Journal of the Acoustical Society of America*, Band 64(5):S. 1386–1391, 1978.
- Kennedy, Graeme D.: *An introduction to corpus linguistics*. Longman, London, 1999.
- Kent, Ray D. und Miolo, Giuliana: Phonetic abilities in the first year of life. In: Fletcher, Paul und MacWhinney, Brian (Hg.) *The Handbook of Child Language*, Blackwell, Oxford, S. 303–331. 2004.
- Kewley-Port, Diane: Speech recognition by machine. In: Pickett, James M. (Hg.) *The Acoustics of Speech Communication*, Allyn and Bacon, Boston. 1999.
- Kim, Sahyang: The role of post-lexical tonal contours in word segmentation. In: *Proceedings for 15th International Congress of Phonetic Sciences*, Band Barcelona (Spanien), 2003.
- Klabunde, Ralf: *Formale Grundlagen der Linguistik : ein Arbeitsbuch*. Narr, Tübingen, 1998.
- Klann-Delius, Gisela: *Spracherwerb*. Sammlung Metzler, 321. Metzler, Stuttgart, 1999.
- Klatt, Dennis H.: Speech perception. In: *Journal of Phonetics*, Band 7(3):S. 279–312, 1979.
- Klatt, Dennis H.: Review of selected models of speech perception. In: Marslen-Wilson, William (Hg.) *Lexical Representation and Process*, MIT Press, Cambridge, S. 169–226. 1989.

Literaturverzeichnis

- Klenk, Ursula: Auf der Suche nach einer natürlichen Grammatik. In: Willee, Gerd, Schröder, Bernhard und Schmitz, Hans-Christian (Hg.) *Computerlinguistik: Was geht, was kommt?*, Gardez! Verlag, Sankt Augustin, Reihe: Sprachwissenschaft, Computerlinguistik und neue Medien, S. 180–183. 2002.
- Kluender, Keith R.: Speech perception as a tractable problem in cognitive science. In: *Handbook of Psycholinguistics*, S. 173–217, 1994.
- Knobloch, Clemens: Geschichte der Psycholinguistik. In: Rickheit, Gert, Herrmann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Handbücher zur Sprach- und Kommunikationswissenschaft, S. 15–33. 2003.
- Kohonen, Teuvo: Self-organized formation of topologically correct feature maps. In: *Biological Cybernetics*, Band 43:S. 59–69, 1982.
- Kuhl, Patricia K.: Perceptual constancy for speech-sound categories in early infancy. In: Yeni-Komishian, Grace H., Kavanagh, James F. und Ferguson, Charles A. (Hg.) *Child Phonology: Perception*, Academic Press, New York, Band 2, S. 41–66. 1980.
- Kuhl, Patricia K.: Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. In: *Perception & Psychophysics*, Band 50(2):S. 93–107, 1991.
- Kuhl, Patricia K.: A new view of language acquisition. In: Papers from the National Academy of Sciences Colloquium on Auditory Neuroscience: Development, Transduction und Integration (Hg.) *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Science, Washington/DC, Band 97, S. 11850–11857. 2000.
- Kuhl, Patricia K., Conboy, Barbara T., Padden, Denise, Nelson, Tobey und Pruitt, Jessica: Early speech perception and later language development. In: *Language Learning and Development*, Band 1(3-4):S. 237–264, 2005.
- Kuhl, Patricia K., Lacerda, Williams und Lindblom, Steven: Linguistic experiences alter phonetic perception in infants by 6 month of age. In: *Science*, Band 255:S. 606–608, 1992.
- Kurzweil, Ray: *KI: das Zeitalter der künstlichen Intelligenz*. Hanser, München, 1993.
- Ladefoged, Peter: *A course in phonetics*. Harcourt Brace Jovanovich, New York, 1975.
- Lahiri, Aditi und Marslen-Wilson, William D.: The mental representation of lexical form. In: *Cognition*, Band 38(3):S. 245–294, 1991.
- Lalonde, Chris E. und Werker, Janet F.: Cognitive influences on cross-language speech perception in infancy. In: *Infant Behavior & Development*, Band 18(4):S. 459–475, 1995.

- Langacker, Ronald W.: *Concept, image and symbol*. Mouton de Gruyter, Berlin, 1990.
- Langacker, Ronald W.: *Grammar and conceptualization*. Cognitive Linguistics Research, 14. Mouton de Gruyter, Berlin, 1999.
- Lappin, Shalom und Shieber, Stuart M.: Machine learning theory and practice as a source of insight into universal grammar. In: *Journal of Linguistics*, Band 43(2):S. 393–427, 2007.
- Lasnik, Howard und Uriagereka, Juan: On the poverty of the challenge. In: *The Linguistic Review*, Band 19(1):S. 147–150, 2002.
- Lecanuet, Jean-Pierre: Foetal responses to auditory and speech stimuli. In: Slater, Alan (Hg.) *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy*, Psychology Press, Hove, S. 317–355. 1998.
- Legate, Julie Anne und Yang, Charles D.: Empirical re-assessment of stimulus poverty arguments. In: *The Linguistic Review*, Band 19(1):S. 151–162, 2002.
- Leibniz, Gottfried Wilhelm: *Leibniz*. Deutscher Taschenbuchverlag, München, 2000.
- Levelt, Willem J. M. und Peter, W. Jusczyk: *Speaking*. MIT Press, Cambridge, 1989.
- Levitt, Andrea, Jusczyk, Peter W. und Murray, Janice: Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. In: *Journal of Experimental Psychology: Human Perception and Performance*, Band 14(3):S. 361–368, 1988.
- Lewis, John D. und Elman, Jeffrey L.: Lernability and the statistical structure of language: Poverty of the stimulus arguments revisited. In: *Proceedings of the 26th Annual Boston University Conference on Language Development*, 2001.
- Liberman, Alvin. M., Cooper, F. S., Schankweiler, D. P. und Studdert-Kennedy, Michael: Perception of the speech code. In: *Psychological Review*, Band 74(6):S. 431–461, 1967.
- Liberman, Alvin. M. und Mattingly, Ignatius G.: A specialization for speech perception. In: *Science*, Band 243:S. 489–494, 1989.
- Licklider, Joseph C. R.: On the process of speech perception. In: *Journal of the Acoustical Society of America*, Band 24(590-594), 1952.
- Lieberman, Philip: *Human language and our reptilian brain: the subcortical bases of speech, syntax, and thought*. Harvard University Press, Cambridge, 2000.
- Lieberman, Philip: *Toward an evolutionary biology of language*. Harvard University Press, Cambridge, 2006.
- Lively, Scott E., Pisoni, David B. und Goldinger, Stephen D.: Spoken word recognition. In: *Handbook of Psycholinguistics*, S. 265–301, 1994.

Literaturverzeichnis

- Locke, John: *An essay concerning human understanding*. Clarendon Press, Oxford, reprinted Auflage, 1987.
- Luce, Paul A. und Large, Nathan R.: Phonotactics, density, and entropy in spoken word recognition. In: *Language and Cognitive Processes*, Band 16(5-6):S. 565–581, 2001.
- Luce, Paul A. und Pisoni, David B.: Recognizing spoken words: The neighborhood activation model. In: *Ear and Hearing*, Band 19(1):S. 1–36, 1998.
- Macken, Marlys A.: Phonological acquisition. In: *The Handbook of Phonological Theory*, S. 671–696, 1995.
- MacWhinney, Brian: The competition model. In: MacWhinney, Brian (Hg.) *Mechanisms of Language Acquisition*, Erlbaum, Hillsdale, S. 249–308. 1987.
- MacWhinney, Brian: *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum, Hillsdale, 1995.
- MacWhinney, Brian: Emergentist approaches to language. In: Bybee, Joan und Hopper, Paul (Hg.) *Frequency and the Emergence of Linguistic Structure*, John Benjamins, Amsterdam, S. 449–470. 2001.
- Manning, Christopher D. und Schütze, Hinrich: *Foundations of statistical natural language processing*. MIT Press, Cambridge, 2003.
- Marcus, Gary F.: Can connectionism save constructivism? In: *Cognition*, Band 66(2):S. 153–182, 1998.
- Marcus, Gary F.: Do infants learn grammar with algebra or statistics? response. In: *Science*, Band 284:S. 436–437, 1999a.
- Marcus, Gary F.: Reply to Christiansen and Curtin. In: *Trends in Cognitive Sciences*, Band 3(8):S. 290–291, 1999b.
- Marcus, Gary F.: Reply to Seidenberg and Elman. In: *Trends in Cognitive Sciences*, Band 3(8):S. 289, 1999c.
- Marcus, Gary F.: Technical comment summaries: Rule learning by seven-month-old infants. In: *Science*, Band 284:S. 875, URL am 5. Juli 2008: <http://www.sciencemag.org/cgi/content/full/284/5416/875a>, 1999d.
- Marcus, Gary F.: *The algebraic mind: Integrating connectionism and cognitive science*. Bradford Books, Cambridge, 2001.
- Marcus, Gary F.: *Der Ursprung des Geistes*. Patmos Verlag, Düsseldorf, 2005.
- Marcus, Gary F., Vijayan, S., Bandi Rao, S. und Vishton, P. M.: Rule learning by seven-month-old infants. In: *Science*, Band 283(5398):S. 77–79, 1999.

- Marcus, Stephen Michael: Eris – context sensitive coding in speech perception. In: *Journal of Phonetics*, Band 9(2):S. 197–200, 1981.
- Marcus, Stephen Michael: Recognizing speech: On the mapping from sound to word. In: Bouma, Herman und Bouwhuis, Don G. (Hg.) *Attention and Performance X: Control of Language Processes*, Erlbaum, London, S. 151–163. 1984.
- Marslen-Wilson, William D.: Function and process in spoken word recognition: A tutorial review. In: Bouma, Herman und Bouwhuis, Don G. (Hg.) *Attention and Performance X: Control of Language Processes*, Erlbaum, London, S. 125–150. 1984.
- Marslen-Wilson, William D. und Warren, Paul: Levels of perceptual representation and process in lexical access. In: *Psychological Review*, Band 101(4):S. 653–675, 1994.
- Marslen-Wilson, William D. und Welsh, Alan: Processing interaction and lexical access during recognition of continuous speech. In: *Cognitive Psychology*, Band 10:S. 29–63, 1978.
- Massaro, Dominic: Psychological aspects of speech perception. In: Gernsbacher, Morton Ann (Hg.) *Handbook of Psycholinguistics*, Academic Press, New York, S. 219–263. 1994.
- Massaro, Dominic: *Perceiving talking faces: From speech perception to a behavioral principle*. MIT Press, Cambridge, 1998.
- Matthews, Danielle, Lieven, Elena, Theakston, Anna und Tomasello, Michael: The role of frequency in the acquisition of english word order. In: *Cognitive Development*, Band 20(1):S. 121–136, 2005.
- Matthews, Robert J.: The case for linguistic nativism. In: Staiton, Robert J. (Hg.) *Contemporary Debates in Cognitive Science*, Blackwell, Malden, S. 81–96. 2006.
- Mattys, Sven L. und Clark, Jamie H.: Lexical activity in speech processing: evidence from pause detection. In: *Journal of Memory and Language*, Band 47(3):S. 343–359, 2002.
- Mattys, Sven L. und Jusczyk, Peter W.: Phonotactic cues for segmentation of fluent speech by infants. In: *Cognition*, Band 78(2):S. 91–121, 2001.
- Mattys, Sven L., Jusczyk, Peter W., Luce, Paul A. und Morgan, James L.: Phonotactic and prosodic effects on word segmentation in infants. In: *Cognitive Psychology*, Band 38(4):S. 465–494, 1999.
- Maye, Jessica, Werker, Janet F. und Gerken, LouAnn: Infant sensitivity to distributional information can affect phonetic discrimination. In: *Cognition*, Band 82(3):S. B101–B111, 2002.

Literaturverzeichnis

- McClelland, James L. und Elman, Jeffrey L.: The trace model of speech perception. In: *Cognitive Psychology*, Band 18(1):S. 1–86, 1986.
- McClelland, James L. und Plaut, Davis C.: Does generalization in infant learning implicate abstract algebra-like rules? In: *Trends in Cognitive Sciences*, Band 3(5):S. 166–168, 1999.
- McCulloch, Warren und Pitts, Walter: A logical calculus of the ideas immanent in nervous activity. In: *Bull Math Biophysics*, Band 5:S. 115–133, 1943.
- McEnery, Tony und Wilson, Andrew: *Corpus linguistics*. Edinburgh University Press, Edinburgh, 1996.
- McLean, James E. und McLean, Lee S.: *How children learn language*. Singular Publ., San Diego, 1999.
- McNeill, David und Lindig, Karen: The perceptual reality of phonemes, syllables, words and sentences. In: *Journal of Verbal Learning and Verbal Behavior*, Band 12(4):S. 419–430, 1973.
- McQueen, James M.: Segmentation of continuous speech using phonotactics. In: *Journal of Memory and Language*, Band 39(1):S. 21–46, 1998.
- McQueen, James M., Norris, Dennis und Cutler, Anne: Competition in spoken word recognition – spotting words in other words. In: *Journal of Experimental Psychology-Learning Memory and Cognition*, Band 20(3):S. 621–638, 1994.
- McQueen, James M. und Pitt, Mark A.: Transitional probability and phoneme monitoring. In: *Proceedings, 4th ICSLP*, S. 2502–2505, 1996.
- Medler, David A.: A brief history of connectionism. URL am 5. Juli 2008: http://neuron-ai.tuke.sk/NCS/VOL1/P3_html/vol1_3.html, 1998.
- Mehler, Jacques: The role of syllables in speech processing: Infant and adult data. In: *The Psychological Mechanisms of Language*, Band Symposium of the Royal Society and the British Society:S. 333–352, 1981.
- Mehler, Jacques, Christophe, Anne und Ramus, Franck: How infants acquire language. In: *Image, Language, Brain*, S. 51–75, 2000.
- Mehler, Jacques, Dommergues, Jean Yves und Frauenfelder, Uli: The syllable's role in speech segmentation. In: *Journal of Verbal Learning and Verbal Behavior*, Band 20(3):S. 298–305, 1981.
- Mehler, Jacques, Dupoux, Emmanuel und Segui, Juan: Constraining models of lexical access: The onset of word recognition. In: Altmann, Gerry T.M. (Hg.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, MIT Press, Cambridge, S. 236–262. 1995.

- Mehler, Jacques, Jusczyk, Peter W., Lambertz, Ghislaine, Halsted, Nilofar, Bertoncini, Josiane und Amiel-Tison, Claude: A precursor of language acquisition in young infants. In: *Cognition*, Band 29(2):S. 143–178, 1988.
- Mehler, Jacques, Pallier, Christophe und Christophe, Anne: Language and cognition. In: Sbourin, M., Craik, F. und Robert, M. (Hg.) *Advances in Psychological Science: Biological and Cognitive Aspects*, Volume 2, S. 381–398. 1998.
- Meschyan, Gayane und Hernandez, Arturo: Age of acquisition and word frequency: Determinants of object-naming speed and accuracy. In: *Memory & Cognition*, Band 30(2):S. 262–269, 2002.
- Miller, George A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. In: *Psychological Review*, Band 63:S. 81–97, 1956.
- Miller, Joanne L. und Dexter, Emily R.: Effects of speaking rate and lexical status on phonetic perception. In: *Journal of experimental psychology / Human perception and performance*, Band 14(3):S. 369–378, 1988.
- Miller, Joanne L. und Eimas, Peter D.: Organization in infant speech perception. In: *Canadian Journal of Psychology*, Band 33:S. 353–367, 1979.
- Miller, Joanne L. und Eimas, Peter D.: Studies on the categorization of speech by infants. In: *Cognition*, Band 13(2):S. 135–165, 1983.
- Minsky, Marvin Lee und Papert, Seymour: *Perceptrons*. MIT Press, Cambridge, 1969.
- Mitkov, Ruslan: *The Oxford handbook of computational linguistics*. Oxford University Press, Oxford, 2003.
- Müller, Horst M. und Rickheit, Gert: Die Erforschung der neurokognitiven Grundlagen der Sprache. In: Müller, Horst M. und Rickheit, Gert (Hg.) *Neurokognition der Sprache*, Stauffenburg, Tübingen, S. 1–4. 2003.
- Monahan, Philip: We need more than statistics to segment speech. University of Maryland, ms., 2006.
- Moore, Jean K.: Maturation of human auditory cortex: Implications for speech perception. In: *Annals of Otology Rhinology and Laryngology*, Band 111(5):S. 7–10, 2002.
- Morais, José, Cary, Luz, Alegria, Jesús und Bertelson, Paul: Does awareness of speech as a sequence of phones arise spontaneously? In: *Cognition*, Band 7(4):S. 323–331, 1979.
- Morais, José und Kolinsky, Régine: Perception and awareness in phonological processing – the case of the phoneme. In: *Cognition*, Band 50(1-3):S. 287–297, 1994.
- Morgan, James L.: Converging measures of speech segmentation in preverbal infants. In: *Infant Behavior & Development*, Band 17(4):S. 389–403, 1994.

Literaturverzeichnis

- Morgan, James L.: Extracting sentence structure from infant-directed speech. In: *Infant Behavior & Development*, S. 632, 1996a.
- Morgan, James L.: A rhythmic bias in preverbal speech segmentation. In: *Journal of Memory and Language*, Band 35(5):S. 666–688, 1996b.
- Morgan, James L. und Saffran, Jenny R.: Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. In: *Child Development*, Band 66(4):S. 911–936, 1995.
- Morton, John: Interaction of information in word recognition. In: *Psychological Review*, Band 76(2):S. 165–178, 1969.
- Morton, John und Long, John: Effect of word transitional probability on phoneme identification. In: *Journal of Verbal Learning and Verbal Behavior*, Band 15(1):S. 43–51, 1976.
- Moss, Helen, Tyler, Lorraine K. und Taylor, Kirsten I.: Conceptual structure. In: Gaskell, Gareth M., Altmann, Gerry T. M. und Bloom, Paul (Hg.) *The Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford, S. 217–234. 2007.
- Mutter, Sharon A. und Hashtroudi, Shahin: Cognitive effort and the word frequency effect in recognition and lexical decision. In: *The American Journal of Psychology*, Band 100(1):S. 93–116, 1987.
- Myers, James, Jusczyk, Peter W., Kemler-Nelson, Deborah G., Charles Luce, J., Woodward, A. L. und HirshPasek, K.: Infants' sensitivity to word boundaries in fluent speech. In: *Journal of Child Language*, Band 23(1):S. 1–30, 1996.
- Nazzi, Thierry, Bertoncini, Josiane und Mehler, Jacques: Language discrimination by newborns: Toward an understanding of the role of rhythm. In: *Journal of Experimental Psychology: Human Perception and Performance*, Band 24(3):S. 756–766, 1998.
- Nazzi, Thierry, Jusczyk, Peter W. und Johnson, Elizabeth K.: Language discrimination by english-learning 5-month-olds: Effects of rhythm and familiarity. In: *Journal of Memory and Language*, Band 43(1):S. 1–19, 2000a.
- Nazzi, Thierry, Kemler-Nelson, Deborah G., Jusczyk, Peter W. und Jusczyk, Ann M.: Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. In: *Infancy*, Band 1(1):S. 123–147, 2000b.
- Negishi, Michiro: Do infants learn grammar with algebra or statistics? In: *Science*, Band 284(5413):S. 435; author reply 436–7, 1999.
- Neisser, Ulric: An experimental distinction between perceptual process and verbal response. In: *Journal of Experimental Psychology*, Band 47:S. 399–402, 1954.

- Newell, Allen: *Unified theories of cognition*. The William James lectures ; 1987. Harvard Univ. Press, Cambridge, 1990.
- Newport, Elissa L.: Maturational constraints on language learning. In: *Cognitive Science*, Band 14:S. 11–28, 1990.
- Newport, Elissa L. und Aslin, Richard N.: Innately constrained learning: Blending old and new approaches to language acquisition. In: Howell, S. C., Fish, S. A. und Keith-Lucas, T. (Hg.) *Proceedings of the 24th Annual Boston University Conference on Language Development*. Cascadilla Press, Somerville, 2000, S. 1–21.
- Newsome, Mary R. und Jusczyk, Peter W.: Do infants use stress as a cue in segmenting fluent speech? In: *Conference on Language Development: Proceedings of the 19th Annual Boston University Conference on Language Development*, Cascadilla Press, Somerville, Band 2, S. 415–426. 1995.
- Niyogi, Partha: *The computational nature of language learning and evolution*. Current studies in linguistics; 43. MIT Press, Cambridge, 2006.
- Norris, Dennis: Shortlist – a connectionist model of continuous speech recognition. In: *Cognition*, Band 52(3):S. 189–234, 1994.
- Norris, Dennis und Cutler, Anne: The relative accessibility of phonemes and syllables. In: *Perception & Psychophysics*, Band 43(6):S. 541–550, 1988.
- Norris, Dennis, McQueen, James M. und Cutler, Anne: Competition and segmentation in spoken-word recognition. In: *Journal of Experimental Psychology: Learning Memory and Cognition*, Band 21(5):S. 1209–1228, 1995.
- Norris, Dennis, McQueen, James M., Cutler, Anne und Butterfield, Sally: The possible-word constraint in the segmentation of continuous speech. In: *Cognitive Psychology*, Band 34(3):S. 191–243, 1997.
- Norris, Dennis, McQueen, James M., Cutler, Anne, Butterfield, Sally und Kearns, Ruth: Language-universal constraints on speech segmentation. In: *Language and Cognitive Processes*, Band 16(5/6):S. 637–660, 2001.
- Nusbaum, Howard C. und Goodman, Judith C.: Learning to hear speech as spoken language. In: *The Development of Speech Perception*, S. 299–338, 1994.
- Oden, Gregg C. und Massaro, Dominic: Integration of featural information in speech perception. In: *Psychological Review*, Band 85(3):S. 172–191, 1978.
- Oerter, Rolf und Montada, Leo: *Entwicklungspsychologie: ein Lehrbuch*. Beltz, Psychologie-Verlags-Union, Weinheim, 1998.
- Olivier, D. C.: *Stochastic grammars and language acquisition mechanisms*. Harvard Universität, Cambridge, 1968.

Literaturverzeichnis

- Pallier, Christophe: Phonemes and syllables in speech perception: Size of attentional focus in French. In: *Proceedings of Eurospeech '97*, S. 67–74, 1998.
- Pallier, Christophe, Sebastian-Gallés, Núria, Felguera, Teodora, Christophe, Anne und Mehler, Jacques: Attentional allocation within the syllabic structure of spoken words. In: *Journal of Memory and Language*, Band 32(3):S. 373–389, 1993.
- Papoušek, Mechthild: *Vom ersten Schrei zum ersten Wort: Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation*. Huber, Bern, 1994.
- Paprotté, Wolf: Korpuslinguistik. In: Müller, Horst M. (Hg.) *Arbeitsbuch Linguistik*, Ferdinand Schöningh, Paderborn, S. 364–381. 2005.
- Peña, Marcela, Bonatti, Luca L., Nespors, Marina und Mehler, Jacques: Signal-driven computations in speech processing. In: *Science*, Band 298(5593):S. 604–607, 2002.
- Pence, Khara L. und Justice, Laura M.: *Language development from theory to practice*. Pearson Merrill Prentice Hall, Upper Saddle River, 2008.
- Pereira, Fernando: Formal grammar and information theory: Together again? In: *Philosophical Transactions*, Band 358(1769):S. 1239–1253, 2000.
- Perruchet, Pierre und Vinter, Anne: Parser: A model for word segmentation. In: *Journal of Memory and Language*, Band 39(2):S. 246–263, 1998.
- Pesetsky, David Michael, Wexler, Kenneth und Fromkin, Victoria: Acquiring language. In: *Science*, Band 276(5316):S. 1177, 1997.
- Peters, Ann M.: *The units of language acquisition*. Cambridge University Press, Cambridge, 1983.
- Piaget, Jean: *The language and thought of the child*. Kegan Paul, London, 1923.
- Piaget, Jean: *Biologie und Erkenntnis*. Fischer, Frankfurt am Main, 1974.
- Pickett, James M.: *The Acoustics of Speech Communication*. Allyn and Bacon, Boston, 1999.
- Pierce, Patsy und Profio, Andrea: From cooing to conversation to the carrot seed: Oral and written language connections. In: Rosenkotter, Sharon E. und Knapp-Philo, Joanne (Hg.) *Learning to Read the World: Language and Literacy in the First Three Years*, Zero to Three Press, Washington/DC. 2006.
- Pierrehumbert, Janet B.: Phonological and phonetic representation. In: *Journal of Phonetics*, Band 18(3):S. 375–394, 1990.
- Pierrehumbert, Janet B.: Phonetic diversity, statistical learning, and acquisition of phonology. In: *Language and Speech*, Band 46:S. 115–154, 2003.

- Pinker, Steven: *Language learnability and language development*. Harvard University Press, Cambridge, 1984.
- Pinker, Steven: *Learnability and cognition*. MIT Press, Cambridge, 1989.
- Pinker, Steven: *The language instinct*. Morrow, New York, 1994.
- Pinker, Steven: Acquiring language. In: *Science*, Band 276:S. 1178, 1997.
- Pinker, Steven: Out of the minds of babes. In: *Science*, Band 283(5398):S. 40–41, 1999.
- Pinker, Steven: Four decades of rules and associations, or whatever happened to the past tense debate? In: *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, S. 157–179, 2001.
- Pinker, Steven: *The blank slate: The modern denial of human nature*. Lane, Penguin Books, London, 2002.
- Pinker, Steven: On human nature – why nature and nurture won't go away. In: *Daedalus*, Band 133(4):S. 5–17, 2004.
- Pinker, Steven und Prince, Alan: On language and connectionism. In: *Cognition*, Band 28(1-2):S. 73–193, 1988. Analysis of a parallel distributed processing model of language acquisition.
- Pisoni, David B.: Some current theoretical issues in speech perception. In: *Cognition*, Band 10(1-3):S. 249–259, 1981.
- Pisoni, David B. und Levi, Susannah V.: Representations and representational specificity in speech perception and spoken word recognition. In: Gaskell, M. Gareth (Hg.) *The Oxford Handbook of Psycholinguistics*, Oxford University Press, Oxford, S. 3–17. 2007.
- Pisoni, David B. und Luce, Paul A.: Acoustic-phonetic representations in word recognition. In: *Cognition*, Band 25:S. 21–52, 1987.
- Pitt, Mark A. und McQueen, James M.: Is compensation for coarticulation mediated by the lexicon? In: *Journal of Memory and Language*, Band 39(3):S. 347–370, 1998.
- Pitt, Mark A., Samuel, Arthur G. und Brian, MacWhinney: Attentional allocation during speech perception. In: *Journal of Memory and Language*, Band 29(5):S. 611–632, 1990.
- Platon: Menon. In: Loewenthal, Erich (Hg.) *Sämtliche Werke in drei Bänden*, Wissenschaftliche Buchgesellschaft, Darmstadt, Band 1. 2004.
- Plomin, Robert: *Nature and nurture*. Brooks/Cole, Pacific Grove, 1990.
- Plunkett, Kim: Language acquisition and connectionism. In: *Language and Cognitive Processes*, Band 13(2-3):S. 97–104, 1998.

Literaturverzeichnis

- Polka, Linda und Werker, Janet F.: Developmental changes in perception of nonnative vowel contrasts. In: *Journal of Experimental Psychology: Human Perception and Performance*, Band 20(2):S. 421–435, 1994.
- Pospeschill, Markus: *Konnektionismus und Kognition: eine Einführung*. Kohlhammer, Stuttgart, 2004.
- Price, Cathy, Indefrey, Peter und van Turenhout, Miranda: The neural architecture underlying the processing of written and spoken word forms. In: Brown, Colin M. und Hagoort, Peter (Hg.) *Neurocognition of Language*, Oxford University Press, Oxford, S. 212–240. 1999.
- Pullum, Geoffrey K. und Scholz, Barbara C.: Empirical assessment of stimulus poverty arguments. In: *The Linguistic Review*, Band 19(1):S. 9–50, 2002.
- Putnam, Hilary: The innateness hypothesis and explanatory models in linguistics. In: Searle, John R. (Hg.) *The Philosophy of Language*, Oxford University Press, Glasgow. 1971.
- Pylyshyn, Zenon W.: *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press, Cambridge, 1984.
- Ramus, Franck: Language discrimination by newborns. In: *Annual Review of Language Acquisition*, Band 2:S. 85–115, 2002.
- Ramus, Franck und Mehler, Jacques: Language identification with suprasegmental cues: A study based on speech resynthesis. In: *Journal of the Acoustical Society of America*, Band 105(1):S. 512–521, 1999.
- Reali, Florencia und Christiansen, Morton H. (Hg.): *Reappraising Poverty of Stimulus Argument: A Corpus Analysis Approach*. URL am 5. Juli 2008: <http://cnl.psych.cornell.edu/papers/fr-bucl2003.pdf>, Proceedings of the 28th Boston University Conference on Language Development, 2003.
- Redington, Martin und Chater, Nick: Probabilistic and distributional approaches to language acquisition. In: *Trends in cognitive sciences*, Band 1(7):S. 273–281, 1997.
- Reetz, Henning: Spezielle Verfahren V: Analyse und Synthese von Sprachlauten. In: Rickheit, Gert, Herrmann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Handbücher zur Sprach- und Kommunikationswissenschaft, S. 202–212. 2003.
- Remez, Robert E.: A guide to research on the perception of speech. In: *Handbook of psycholinguistics*, S. 145–172, 1994.
- Ridley, Matt: *Nature via nurture: genes, experience, and what makes us human*. Harper Perennial, London, erste Auflage, 2003.

- Rischel, Jørgen: What is phonetic representation? In: *Journal of phonetics*, Band 18(3):S. 395–410, 1990.
- Rissanen, Jorma: *Stochastic complexity in statistical inquiry*. World Scientific, Singapore, 1989.
- Ritter, Nancy A.: Introduction. In: *The Linguistic Review*, Band 19:S. 1–7, 2002.
- Rohde, Douglas L. T. und Plaut, David C.: Language acquisition in the absence of explicit negative evidence: How important is starting small? In: *Cognition*, Band 72(1):S. 67–109, 1999.
- Rosenblatt, Frank: The perceptron: A probabilistic model for information storage and organization in the brain. In: *Psychological Review*, Band 65:S. 386–408, 1958.
- Roth, Gerhard: *Das Gehirn und seine Wirklichkeit: Kognitive Neurobiologie und ihre philosophischen Konsequenzen*. Suhrkamp, Frankfurt am Main, 1996.
- Roth, Gerhard und Prinz, Wolfgang: *Kopf-Arbeit: Gehirnfunktionen und kognitive Leistungen*. Spektrum Akademischer Verlag, Heidelberg, 1996.
- Roy, Deb K. und Pentland, Alex P.: Learning words from sights and sounds: a computational model. In: *Cognitive Science*, Band 26(1):S. 113–146, 2002.
- Rumelhart, David E. und McClelland, James L.: On learning the past tenses of english verbs. In: McClelland, James L. und Rumelhart, David E. (Hg.) *Parallel Distributed Processing*, MIT, Cambridge, Band 2, S. 216–271. 1986a.
- Rumelhart, David E. und McClelland, James L.: PDP models and general issues in cognitive science. In: McClelland, James L. und Rumelhart, David E. (Hg.) *Parallel Distributed Processing*, MIT, Cambridge, Band 1, S. 110–149. 1986b.
- Saffran, Jenny R.: The use of predictive dependencies in language learning. In: *Journal of Memory and Language*, Band 44(4):S. 493–515, 2001a.
- Saffran, Jenny R.: Words in a sea of sounds: The output of infant statistical learning. In: *Cognition*, Band 81(2):S. 149–169, 2001b.
- Saffran, Jenny R.: Constraints on statistical language learning. In: *Journal of Memory and Language*, Band 47(1):S. 172–196, 2002.
- Saffran, Jenny R.: Absolute pitch in infancy and adulthood: The role of tonal structure. In: *Developmental Science*, Band 6(1):S. 35–43, 2003a.
- Saffran, Jenny R.: Birds do it – why not babies? In: *Developmental Science*, Band 6(1):S. 46–47, 2003b.
- Saffran, Jenny R.: Musical learning and language development. In: *New York Academy of Sciences: Annals of the New York Academy of Sciences*, Band 999:S. 397–401, 2003c.

Literaturverzeichnis

- Saffran, Jenny R.: Statistical language learning: Mechanisms and constraints. In: *Current Directions in Psychological Science*, Band 12(4):S. 110–114, 2003d.
- Saffran, Jenny R., Aslin, Richard N. und Newport, Elissa L.: Statistical learning by 8-month-old infants. In: *Science*, Band 274(5294):S. 1926–1928, 1996a.
- Saffran, Jenny R., Aslin, Richard N. und Newport, Elissa L.: Acquiring language – Response. In: *Science*, Band 276(5316):S. 1180, 1997a.
- Saffran, Jenny R. und Griepentrog, Gregory J.: Absolute pitch in infant auditory learning: Evidence for developmental reorganization. In: *Developmental Psychology*, Band 37(1):S. 74–85, 2001.
- Saffran, Jenny R., Johnson, Elizabeth K., Aslin, Richard N. und Newport, Elissa L.: Statistical learning of tone sequences by human infants and adults. In: *Cognition*, Band 70(1):S. 27–52, 1999.
- Saffran, Jenny R., Newport, Elissa L. und Aslin, Richard N.: Word segmentation: The role of distributional cues. In: *Journal of Memory and Language*, Band 35(4):S. 606–621, 1996b.
- Saffran, Jenny R., Newport, Elissa L., Aslin, Richard N., Tunick, Rachel A. und Barrueco, Sandra: Incidental language learning: Listening (and learning) out of the corner of your ear. In: *Psychological Science*, Band 8(2):S. 101–105, 1997b.
- Saffran, Jenny R., Reeck, Karelyn, Niebuhr, Aimee und Wilson, Diana: Changing the tune: the structure of the input affects infants' use of absolute and relative pitch. In: *Developmental Science*, Band 8(1):S. 1–7, 2005.
- Saffran, Jenny R., Senghas, Ann und Trueswell, John C.: The acquisition of language by children. In: *Proceedings of the National Academy of Sciences of the United States of America*, Band 98(23):S. 12874–12875, 2001.
- Saffran, Jenny R. und Thiessen, Erik D.: Pattern induction by infant language learners. In: *Developmental Psychology*, Band 39(3):S. 484–494, 2003.
- Saffran, Jenny R., Werker, Janet F. und Werker, Lynne A.: The infant's auditory world: Hearing speech, and the beginnings of language. In: Siegler, R. und Kuhn, D. (Hg.) *Handbook of Child Development*, John Wiley & Sons, New York, S. 58–108. 2006.
- Saffran, Jenny R. und Wilson, Diana P.: From syllables to syntax: Multilevel statistical learning by 12-month-old infants. In: *Infancy*, Band 4(2):S. 273–284, 2003.
- Saint-Aubin, Jean und Poirier, Marie: Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. In: *Memory*, Band 13(3–4):S. 325–332, 2005.

- Sampson, Geoffrey: Exploring the richness of the stimulus. In: *The Linguistic Review*, Band 19(1):S. 73–104, 2002.
- Sampson, Geoffrey und Postal, Paul M.: *The 'language instinct' debate*. Continuum, London, 2005.
- Savin, H.B. und Bever, T.G.: The nonperceptual reality of the phoneme. In: *Journal of Verbal Learning and Verbal Behavior*, Band 9(3):S. 295–302, 1970.
- Scharlau, Ingrid, Ansorge, Ulrich und Neumann, Odmir: Spezielle Verfahren IV: Reaktionszeitmessung, Grundlagen und Anwendungen. In: Rickheit, Gert, Herrmann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Band 24 von *Handbücher zur Sprach- und Kommunikationswissenschaft*, S. 190–202. 2003.
- Scheerer, Eckart: *Symbolverarbeitung und Konnektionismus: von der Konfrontation zur Koexistenz*. Mind and Brain: Perspectives in Theoretical Psychology and the Philosophy of Mind. Institut für Kognitionsforschung, Universität Bielefeld, 1990.
- Scholz, Barbara C. und Pullum, Geoffrey K.: Searching for arguments to support linguistic nativism. In: *The Linguistic Review*, Band 19(1):S. 185–223, 2002.
- Scholz, Barbara C. und Pullum, Geoffrey K.: Irrational nativist exuberance. In: Staiton, Robert J. (Hg.) *Contemporary Debates in Cognitive Science*, Blackwell, Malden, S. 59–80. 2006.
- Schukat-Talamazzini, Ernst Günter: *Automatische Spracherkennung: Grundlagen, statistische Modelle und effiziente Algorithmen*. Vieweg, Wiesbaden, 1995.
- Sebastian-Gallés, Núria, Dupoux, Emmanuel, Segui, Juan und Mehler, Jacques: Contrasting syllabic effects in catalan and spanish. In: *Journal of Memory and Language*, Band 31(1):S. 18–32, 1992.
- Segui, Juan: The syllable: A basic perceptual unit in speech processing? In: Bouma, Herman und Bouwhuis, Don G. (Hg.) *Attention and Performance X: Control of Language Processes*, Erlbaum, London, S. 165–181. 1984.
- Segui, Juan, Dupoux, Emmanuel und Mehler, Jacques: The role of the syllable in speech segmentation, phoneme identification, and lexical access. In: Altmann, Gerry T. M. (Hg.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, MIT Press, Cambridge, S. 263–280. 1995.
- Segui, Juan, Frauenfelder, Ulrich H. und Halle, Pierre A.: Phonotactic constraints shape speech perception. In: *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, S. 195–208, 2001.
- Seidenberg, Mark S. und Elman, Jeffrey L.: Do infants learn grammar with algebra or statistics? In: *Science*, Band 284(5413):S. 434–435, 1999.

Literaturverzeichnis

- Seidenberg, Mark S., MacDonald, Maryellen C. und Saffran, Jenny R.: Does grammar start where statistics stop? In: *Science*, Band 298(5593):S. 553–554, 2002.
- Seidl, Horst: Kommentar. In: Aristoteles (Hg.) *Über die Seele*, Felix Meiner Verlag, Hamburg, S. 211–277. 1995.
- Sendlmeier, Walter F. und Sendlmeier, Una Maria: Vom lallen zum sprechen. In: *Sprache & Kognition*, Band 10(3):S. 162–170, 1991.
- Shannon, Claude E. und Weaver, Warren: *The mathematical theory of communication*. University of Illinois Press, Urbana, 1971.
- Shi, Rushen S., Cutler, Anne, Werker, Janet und Cruickshank, Marisa: Frequency and form as determinants of functor sensitivity in english-acquiring infants. In: *Journal of the Acoustical Society of America*, Band 119(6):S. E161–E167, 2006.
- Sichelschmidt, Lorenz und Carbone, Elena: Experimentelle Methoden. In: Rickheit, Gert, Herrmann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Band 24 von *Handbücher zur Sprach- und Kommunikationswissenschaft*, S. 115–124. 2003.
- Siegmüller, Julia: Sprachentwicklung. In: Kaufmann, Liane, Nuerk, Hans-Christoph, Konrad, Kerstin und Willmes, Klaus (Hg.) *Kognitive Entwicklungsneuropsychologie*, Hofgrefe, Göttingen, S. 119–136. 2007.
- Simon, Herbert A. und Newell, Allen: Heuristic problem solving: The next advance in operations research. In: *Operations Research*, Band 6, 1958.
- Singer, Wolf: Time as coding space in neocortical processing: A hypothesis. In: Gazzaniga, Michael S. (Hg.) *The Cognitive Neuroscience*, MIT-Press, Cambridge, S. 91–104. 1995.
- Skinner, Burrhus F.: *Verbal behavior*. Appleton-Century-Crofts, New York, 1957.
- Slobin, Dan Isaac: *The crosslinguistic study of language acquisition*, Band 1-5. Erlbaum, Hillsdale, 1985-1997.
- Smith, Neilson V.: *Chomsky: Ideas and ideals*. Cambridge University Press, Cambridge, 2004.
- Smolensky, Paul und Legendre, Géraldine: Harmony optimization and the computational architecture of the mind/brain. In: Smolensky, Paul und Legendre, Géraldine (Hg.) *The Harmonic Mind*, MIT-Press, Cambridge, Band 3-61. 2006a.
- Smolensky, Paul und Legendre, Géraldine: Principles of the integrated connectionist/symbolic cognitive architecture. In: Smolensky, Paul und Legendre, Géraldine (Hg.) *The Harmonic Mind*, MIT-Press, Cambridge, S. 63–97. 2006b.

- Soderstrom, Melanie, Kemler-Nelson, Deborah G. und Jusczyk, Peter W.: Six-month-olds recognize clauses embedded in different passages of fluent speech. In: *Infant Behavior & Development*, Band 28(1):S. 87–94, 2005.
- Soltis, J.: The signal functions of early infant crying. In: *Behavioral and Brain Sciences*, Band 27(4):S. 443–+, 2004.
- Sommer, Bruce: The shape of kunjen syllables. In: Goyvaerts, D.L. (Hg.) *Phonology in the 1980s*, Story-Scientia, Ghent, S. 231–244. 1981.
- Spitzer, Manfred: *Geist im Netz : Modelle für Lernen, Denken und Handeln*. Spektrum Akademischer Verlag, Heidelberg, 2000.
- Spitzer, Manfred: *Lernen: Gehirnforschung und Schule des Lebens*. Spektrum Akademischer Verlag, Heidelberg, 2002.
- Sproat, R.: *Morphology and Computation*. MIT-Press, Cambridge, 1992.
- Stampe, David: The acquisition of phonetic representation. In: *First language acquisition*, S. 307–315, 2004.
- Stevens, Kenneth N: Models of phonetic recognition II: A feature-based model of speech recognition. In: Mermelstein, P. (Hg.) *12th International Congress on Acoustics*. 1986, Band Proceedings of the Montreal Satellite Symposium of Speech Recognition.
- Strömquist, Sven: Language acquisition in early childhood. In: Rickheit, Gert, Hermann, Theo und Deutsch, Werner (Hg.) *Psycholinguistik: Ein internationales Handbuch*, de Gruyter, Berlin, Band 24 von *Handbücher zur Sprach- und Kommunikationswissenschaft*, S. 790–801. 2003.
- Studdert-Kennedy, Michael: The perception of speech. In: Sebeok, Thomas A. (Hg.) *Current trends in linguistics*, de Gruyter, The Hague, Band 12, S. 2349–2385. 1974.
- Swingley, Daniel: Statistical learning and the contents of the infant vocabulary. In: *Cognitive Psychology*, Band 50:S. 86–132, 2005.
- Swingley, Daniel und Aslin, Richard N.: Spoken word recognition and lexical representation in very young children. In: *Cognition*, Band 76(2):S. 147–166, 2000.
- Szagun, Gisela: *Wie Sprache entsteht: Spracherwerb bei Kindern mit normalem und beeinträchtigtem Hören*. Beltz-Taschenbuch; 103: Psychologie. Beltz, Weinheim, 2001.
- Szagun, Gisela: *Sprachentwicklung beim Kind*. Beltz-Taschenbuch; 62: Psychologie. Beltz, Weinheim, 2006.
- Taft, Marcus: Recognition of affixed words and the word frequency effect. In: *Memory and Cognition*, Band 7(4):S. 263–272, 1979.

Literaturverzeichnis

- Taft, Marcus und Forster, Kenneth I.: Lexical storage and retrieval of prefixed words. In: *Journal of Verbal Learning and Verbal Behavior*, Band 14(6):S. 638–647, 1975.
- Taft, Marcus und Forster, Kenneth I.: Lexical storage and retrieval of polymorphic and polysyllabic words. In: *Journal of Verbal Learning and Verbal Behavior*, Band 15(6):S. 607–620, 1976.
- Thiessen, Erik D., Hill, Emily A. und Saffran, Jenny R.: Infant-directed speech facilitates word segmentation. In: *Infancy*, Band 7(1):S. 53–71, 2005.
- Thiessen, Erik D. und Saffran, Jenny R.: When cues collide: Use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. In: *Developmental Psychology*, Band 39(4):S. 706–716, 2003.
- Thiessen, Erik D. und Saffran, Jenny R.: Spectral tilt as a cue to word segmentation in infancy and adulthood. In: *Perception & Psychophysics*, Band 66(5):S. 779–791, 2004.
- Thomas, Margaret: Development of the concept of „the poverty of the stimulus?“. In: *The Linguistic Review*, Band 19(1):S. 51–72, 2002.
- Tomasello, Michael: Language is not an instinct. In: *Cognitive Development*, Band 10:S. 131–146, 1995.
- Tomasello, Michael: *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press, Cambridge, 2003.
- Trehub, Sandra E.: Infants' sensitivity to vowel and tonal contrasts. In: *Developmental Psychology*, Band 9(1):S. 91–96, 1973.
- Treiman, Rebecca: The structure of spoken syllables. In: *Cognition*, Band 15(1-3):S. 49–74, 1983.
- Vihman, Marilyn M.: *Phonological development: The origins of language in the child*. Blackwell, Cambridge, 1996.
- Vroomen, J., van den Bosch, A. und de Gelder, B.: A connectionist model for bootstrap learning of syllabic structure. In: *Language and Cognitive Processes*, Band 13(2-3):S. 193–220, 1998.
- Vygotsky, Lev S.: *Mind in society*. Harvard University Press, Cambridge, 1978.
- Wachs, Theodore D.: The nature-nurture gap: What we have here is a failure to collaborate. In: Plomin, Robert und McClearn, Gerald E. (Hg.) *Nature, Nurture, and Psychology*, American Psychological Association, Washington/DC, S. 375–391. 1993.
- Waddington, Conrad H.: *The evolution of an evolutionist*. University Press, Edinburgh, 1975.

- Walley, Amanda C., Smith, Linda B. und Jusczyk, Peter W.: The role of phonemes and syllables in the perceived similarity of speech sounds for children. In: *Memory & Cognition*, Band 14(3):S. 220–229, 1986.
- Wallmannsberger, Josef: *Virtuelle Textwelten: theoretische Grundlagen und Implementationsperspektiven der anglistischen Computerphilologie*. Anglistische Forschungen, 226. Winter, Heidelberg, 1994.
- Weijer, Joost Christian van de: *Language input for word discovery*. MPI series in psycholinguistics. Kath. University, Nijmegen, 1999.
- Werker, Janet F.: Exploring developmental changes in cross-language speech perception. In: Osherson, Daniel N. (Hg.) *An invitation to cognitive science*, MIT Press, Cambridge, S. 87–106. 1995.
- Werker, Janet F. und Desjardins, Rennee N.: Listening to speech in the 1st year of life - experiential influences on phoneme perception. In: *Current Directions in Psychological Science*, Band 4(3):S. 76–81, 1995.
- Werker, Janet F., Gilbert, John H. V., Humphrey, Keith und Tees, Richard C.: Developmental aspects of cross-language speech perception. In: *Child development*, Band 52(1):S. 349–355, 1981.
- Werker, Janet F. und Lalonde, Chris E.: Cross-language speech perception: Initial capabilities and developmental change. In: *Developmental Psychology*, Band 24(5):S. 672–683, 1988.
- Werker, Janet F. und Tees, Richard C.: Developmental changes across childhood in the perception of non-native speech sounds. In: *Canadian Journal of Psychology*, Band 37(2):S. 278–286, 1983.
- Werker, Janet F. und Tees, Richard C.: Cross-language speech perception: Evidence for perceptual reorganization during the first year. In: *Infant Behavior & Development*, Band 7:S. 49–63, 1984.
- Wermter, Stefan und Weber, Volker: Screen: Learning a flat syntactic and semantic spoken language analysis using artificial neural networks. In: *Journal of Artificial Intelligence Research*, Band 6:S. 35–85, 1997.
- Wickelgren, Wayne A.: Context-sensitive coding, associative memory and serial order in (speech) behaviour. In: *Psychological Review*, Band 76(1):S. 1–15, 1969.
- Wiener, Norbert: *Cybernetics or Control and Communication in the animal and the machine*. John Wiley & Sons, New York, 1949.
- Wildgen, Wolfgang: *Catastrophe theoretic semantics*. Benjamins, Amsterdam, 1982.
- Wildgen, Wolfgang: *Archetypensemantik*. Narr, Tübingen, 1985.

Literaturverzeichnis

- Wildgen, Wolfgang und Laurent, Mottron: *Dynamische Sprachtheorie: Sprachbeschreibung und Spracherklärung nach den Prinzipien der Selbstorganisation und der Morphogenese*. Studienverlag Brockmeyer, Bochum, 1987.
- Wolff, J. G.: The acquisition of linguistic structures. 1975a.
- Wolff, J. G.: An algorithm for the segmentation of an artificial language analogue. In: *British Journal of Psychology*, Band 66:S. 79–90, 1975b.
- Wolff, J. G.: Frequency, conceptual structure and pattern recognition. In: *British Journal of Psychology*, Band 67:S. 377–390, 1976.
- Wolff, J. G.: The discovery of segments in natural language. In: *British Journal of Psychology*, Band 68(1):S. 97–106, 1977.
- Yang, Charles D.: Universal grammar, statistics or both? In: *Trends in Cognitive Sciences*, Band 8(10):S. 451–456, 2004.
- Yost, William A.: *Fundamentals of hearing : an introduction*. Academic Press, San Diego, fünfte Auflage, 2007.
- Zatorre, Robert J.: Neuroscience: Finding the missing fundamental. In: *Nature*, Band 436(7054):S. 1093–1094, 2005.
- Zenner, Hans-Peter: *Hören: Physiologie, Biochemie, Zell- und Neurobiologie*. Thieme, Stuttgart, 1994. 5 Tabellen.
- Zhang, Jialu: On the syllable structures of chinese relating to speech recognition. In: *Proceedings*, Band 4th ICSLP:S. 2450–2453, 1996.
- Zimmer, Dieter E.: *So kommt der Mensch zur Sprache: Über Spracherwerb, Sprachentstehung und Sprache & Denken*. Heyne, München, 2003.
- Zipf, George Kingsley: Relative frequency as a determinant of phonetic change. In: *Harvard Studies in Classical Philology*, Band 40:S. 1–95, 1929.
- Zipf, George Kingsley: *The psycho-biology of language: an introduction to dynamic philology*. Houghton Mifflin, Boston, 1935.