Linguistische Arbeiten 437

Herausgegeben von Hans Altmann, Peter Blumenthal,
Hans Jürgen Heringer, Ingo Plag, Heinz Vater und Richard Wiese

Sabine Reich

Struktur und Erwerb der englischen Nominalphrase

Max Niemeyer Verlag
Tübingen 2001

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Reich, Sabine: Struktur und Erwerb der englischen Nominalphrase / Sabine Reich. – Tübingen : Niemeyer, 2001

(Linguistische Arbeiten ; 437)

Zugl.: Köln, Univ., Diss., 1999 u. d. T.: Reich, Sabine: Funktionale Kategorien im Minimalist Program und im Erstspracherwerb

ISBN 3-484-30437-5 ISSN 0344-6727

Gedruckt auf alterungsbeständigem Papier.

Druck: Weihert-Druck GmbH, Darmstadt

Einband: Industriebuchbinderei Nädele, Nehren

Inhaltsverzeichnis

Vorwort

Die folgende Arbeit ist eine stellenweise überarbeitete Fassung meiner Dissertation "Funktionale Kategorien im Minimalist Program und im Erstspracherwerb", die ich von 1996 bis 1999 an der Universität zu Köln und an der Technischen Universität Chemnitz verfaßt habe und die 1999 von der Philosophischen Fakultät der Universität zu Köln angenommen wurde. Gutachter waren Prof. Dr. Wolf-Dietrich Bald (erster Referent) und Prof. Dr. Jon Erickson (zweiter Referent). Die Disputation fand am 29.11.1999 statt.

Die Arbeit wäre sicher ohne die Unterstützung einer Vielzahl von Leuten nicht verfaßt worden. Als erstes möchte ich mich herzlich bei dem Betreuer meiner Arbeit, Prof. Dr. Wolf-Dietrich Bald, und bei Professor Dr. Jon Erickson am Englischen Seminar in Köln für ihre langjährige Unterstützung und für ihr Vertrauen in mich bedanken. An die Zeit am Englischen Seminar, insbesondere die Arbeit am Lehrstuhl Bald, denke ich gerne mit Freude zurück. Dem Herausgeber der Reihe Linguistische Arbeiten, Prof. Dr. Heinz Vater, danke ich dafür, die Veröffentlichung bei Niemeyer möglich gemacht zu haben, und für hilfreiche Ratschläge bei der Überarbeitung. Dank geht auch an die Graduiertenförderung des Landes Nordrhein-Westfalen, die mich von August 1996 bis April 1997 mit einem Promotionsstipendium gefördert hat.

Ferner bin ich einer Reihe von Freunden und Kollegen in Köln und Chemnitz zu großem Dank verpflichtet: jetzigen und ehemaligen Mitgliedern der "LinguistInnenbande" in Köln (Marion Gymnich, Monika Klages-Kubitzki, Katja Lenz, Ludger Mainka, Ruth Möhlig, Inge Molitor, Sybil Scholz, Marion Zenthöfer) und dem wechselnden Team in Chemnitz (Birgit Ahlemeyer, Claudia Claridge, Claire Cowie, Christian Eckhardt, Angela Hahn, Naomi Hallan, Eva Hertel, Diana Hudson-Ettle, Anne Schröder, Andrew Wilson). Ohne den Teamgeist sowohl in Köln als auch in Chemnitz hätte mir die Auseinandersetzung mit linguistischen Fragestellungen sicher nicht so viel Spaß gemacht. Insbesondere möchte ich Marion Gymnich, Angela Hahn, Monika-Klages-Kubitzki und Inge Molitor herzlich danken, die sich mit mir auf viele Diskussionen zum Thema Minimalist Program und Spracherwerb eingelassen haben. Dank geht auch an Marion Gymnich sowie an Jens Kreutzer und Manu Braun aus der Abteilung Erickson dafür, daß sie mich während meiner Zeit in Chemnitz mit Kopien von dringend benötigter Literatur versorgt haben.

Schließlich möchte ich mich ganz besonders bei Sebastian, Elke und meinen Eltern für ihre Geduld und Unterstützung während der Promotionszeit bedanken.

Köln, im September 2000 Sabine Reich

1 Einleitung

1.1 Thema, Motivation und Zielsetzung der Arbeit

Die Fragen, ob das Sprachvermögen angeboren ist und welche Komponenten des sprachlichen Wissens genetisch "vorprogrammiert" sind, beschäftigen die Linguistik schon seit langem. Die generative Sprachforschung hat solche Fragen zu den Schwerpunkten ihrer Aktivitäten erklärt, wie beispielsweise im Eingangskapitel zu Chomskys *Knowledge of Language* (1986) deutlich wird:

(i) What constitutes knowledge of language?
(ii) How is knowledge of language acquired?
(iii) How is knowledge of language put to use?
(Chomsky 1986a: 3)

Kurz gefasst spricht Chomsky (1965, 1986a, u.a.) von der Natur, dem Ursprung und dem Gebrauch der menschlichen Sprache als Kernfragen der generativen Theoriebildung. Ziel der generativen Sprachforschung ist es, ein explanatorisch adäquates Modell der Sprache zu finden – also ein Modell, das sowohl die vorhandenen Sprachdaten explizit beschreiben als auch die kognitive Repräsentation und den Erwerb der Sprache angemessen erklären kann.

Zentral in diesem Zusammenhang ist der Terminus der Universalgrammatik. Die Universalgrammatik nach Chomsky (1986a) ist ein angeborenes System von Prinzipien, das alle menschlichen Sprachen gemeinsam haben. Diese Prinzipien können noch innerhalb einer Menge von Optionen (Parameter) variieren – für jede Sprache wird eine Option festgelegt (parametrische Variation zwischen Sprachen). Darüber hinaus ist die Universalgrammatik ein Modell des Anfangszustandes sprachlichen Wissens des neugeborenen Kindes, bevor es sprachlichen Input aus seiner Umgebung erhält. Das Kind verfügt bereits über alle Prinzipien der Universalgrammatik, aber die Parameter sind noch nicht in der Grammatik des Kindes festgelegt.[1] Die Konfrontation mit sprachlichem Input ermöglicht dem Kind, die Parameter in eine Richtung zu 'setzen' und damit eine sprachspezifische Grammatik auf der Grundlage der Universalgrammatik aufzubauen (siehe die ausführliche Darstellung in Kapitel 5, vgl. auch Chomsky 1986a: 3, 1993: 1).

In der Auseinandersetzung mit der Frage, was Bestandteile der Universalgrammatik sind, sind funktionale Elemente wie Determinierer, Tempus/Aspekt oder Konjunktionen seit etwa zehn bis fünfzehn Jahren verstärkt in den Blickpunkt gerückt (vgl. Chomsky 1986b, Fukui/Speas 1986, Abney 1987, Pollock 1989, Olsen/Fanselow 1991, Ouhalla 1991, Chomsky 1995b, Clahsen 1996, Bobaljik/Thráinsson 1998, Hudson 1998, Parodi 1998, u.a.). In der generativen Sprachforschung wird von einem Gegensatz zwischen sogenannten funktionalen Kategorien und lexikalischen Kategorien ausgegangen. Während lexikalische Kategorien beispielsweise die Kategorien der Verben, Nomina und Adjektive sind, werden funk-

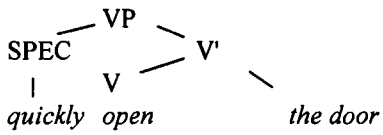
---

[1] Dies sind Annahmen, wie sie z.B. in Chomsky (1986a) geäußert werden. Der Status der Prinzipien und Parameter in der Grammatik des Kindes wird in der Erstspracherwerbsforschung jedoch kontrovers diskutiert, vgl. dazu Kapitel 5.

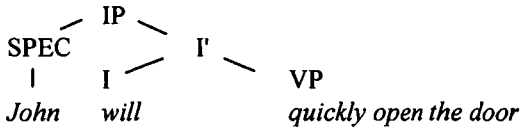
tionale Kategorien durch grammatische Elemente, beispielsweise Determinierer oder Konjunktionen, realisiert.[2] Für die generative Sprachforschung sind nur die lexikalischen Kategorien Teil der Universalgrammatik, während funktionale Kategorien sprachspezifisch unterschiedlich realisiert werden ("parametrische Variation", s.o.). Genauer gesagt sind laut Chomsky alle Parameter und damit auch die Parametersetzung ausschließlich mit dem Erwerb funktionaler Kategorien verknüpft (vgl. Chomsky 1989: 44).

Das verstärkte Interesse der generativen Linguistik an funktionalen Kategorien nimmt seinen Anfangspunkt unter anderem mit dem Vorschlag von Stowell (1981), *clauses* als Projektionen ihrer Flexion (*inflection*, I) zu betrachten; ferner sollen Sätze als Complementizer-Phrasen (CPs) interpretiert werden. Während vorher nur von Verbalphrasen, Nominalphrasen oder Adjektivphrasen ausgegangen wurde, wird nun von englischen Modalverben wie *will* und Complementizern wie *that* angenommen, dass sie auch Phrasen projizieren:

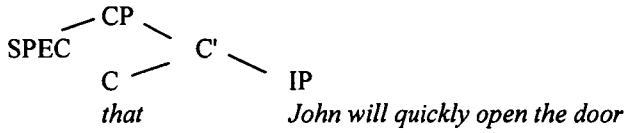
(1)



(2)



(3)



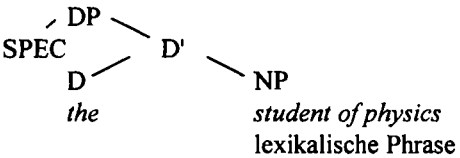
Abney (1987) schlägt ferner vor, die traditionelle Nominalphrase nicht mehr als NP zu betrachten. Vielmehr sieht Abney die gesamte Nominalphrase[3] nun als DP, d.h. als maximale Projektion der funktionalen Kategorie D:

---

[2] Der Terminus 'funktionale Kategorie' ist mehrdeutig. Grundsätzlich müsste zwischen 'funktionaler Kategorie' als Synonym für (grammatische) Wortklasse, 'funktionaler Kategorie' als Synonym für ein Bündel von Eigenschaften (wie etwa INFL) und 'funktionaler Kategorie' als Position im Phrasenstrukturbaum (C, D, I) unterschieden werden. In der Literatur wird diese Unterscheidung häufig nicht gemacht, beispielsweise setzt Haspelmath (1994) funktionale Kategorien mit Funktionswörtern gleich. In der Regel wird aber aus dem Zusammenhang deutlich, was gemeint ist. Eine ausführlichere Diskussion dieser Problematik findet man in Hudson (1998).

[3] Im folgenden bezeichne ich nominale Strukturen als "Nominalphrase", wenn auf keine spezielle Theorie Bezug genommen wird, d.h. "Nominalphrase" bezieht sich auf die DP in der DP-Analyse bzw. auf die NP in der Standardtheorie (Chomsky 1965). Ich benutze NP im folgenden nicht als Abkürzung für "Nominalphrase". NP ist für mich die maximale Projektion von N. Mit dieser Terminologiewahl folge ich Abney (1987: 59-60).

(4)



Auch Chomsky/Lasnik (1995: 58) schließen sich dieser Analyse an: "We have so far considered two functional categories: I and C. A natural extension is that just as propositions are projections of functional categories, so are the traditional noun phrases. The functional head in this case is D, a position filled by a determiner[4], a possessive agreement element, or a pronoun [...]." Die Interpretation aller Nominalphrasen als Determiniererphrasen nach Abney wird in der Literatur auch als die "DP-Analyse" der Nominalphrase bezeichnet. Im Anschluss an die DP-Analyse sind eine Reihe neuer funktionaler Kategorien vorgeschlagen worden. Ferner sind innerhalb des Minimalistischen Programms (Chomsky 1989, 1993, 1995c), einer Weiterentwicklung der generativen Government/Binding-Theorie der achtziger Jahre (Chomsky 1981), die funktionalen Kategorien in den Mittelpunkt des Interesses gerückt (vgl. auch Kapitel 2).

Vor allem für Sprachvergleiche wurden zunehmend funktionale Kategorien herangezogen, um Unterschiede zu erläutern. So behaupten beispielsweise Fukui (1986) und Fukui/Speas (1986), dass im Japanischen eine overte Wh-Bewegung fehlt, weil das Japanische keine funktionale Kategorie COMP besitzt. Nur Sprachen wie das Englische, die eine Position C haben, verfügen dann mit [Spec, C] über einen Landeplatz für Wh-Elemente:

(5) $[_{CP}[_{Spec}\ When_i]\ [_C\ will_j]\ [_{IP}\ we\ three\ t_j\ meet\ again\ t_i]]$

Das große Interesse an komparativen Untersuchungen dieser Art führte letztlich zu einer Flut von Vorschlägen zu funktionalen Kategorien. Viele Wortklassen, Morpheme und abstrakte morphosyntaktische Merkmale wurden als Realisierungen funktionaler Kategorien interpretiert, so wurden beispielsweise im verbalen Bereich die folgenden Kategorien postuliert:

- AgrS (*subject agreement*)
- AgrO (*object agreement*)
- T (*tense*)
- Asp (*aspect*)
- Prog (*progressive*)
- Neg (*negation*) (vgl. Chomsky 1989, Pollock 1989, Ouhalla 1991).

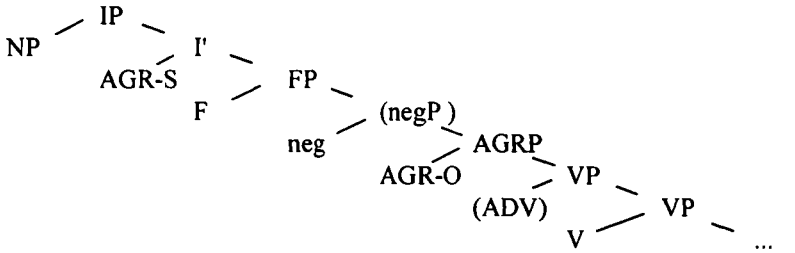
Zusätzlich wird noch angenommen, daß alle genannten funktionalen Kategorien Phrasen projizieren, was letztlich zu sehr komplexen Strukturen führt. Chomsky (1989) z.B. schlägt

---

[4] Die deutsche Terminologie ist hier nicht einheitlich. Deutsche Entsprechungen für den englischen Ausdruck *determiner* sind Determinierer, Determinans, Determinante oder Determinator. Ich folge der Terminologie von Löbel (1990).

folgenden Phrasenstrukturbaum für die IP vor, was eine erhebliche Erweiterung gegenüber den Annahmen in Chomsky (1986b) darstellt:

(6) a. S = I'' = [NP [$_{I'}$ I [$_{VP}$ V ...]]] (vgl. Chomsky 1986b: 3)
b.



(Chomsky 1989: 58)

Diese komplexen Phrasenstrukturbäume sind problematisch, weil oftmals die postulierten funktionalen Kategorien und ihre Projektionen kaum rechtfertigt werden. Viele Ansätze gehen einfach davon aus, dass jedes Flexionsaffix eine zugehörige funktionale Phrase besitzt. Deshalb ist in den letzten Jahren Kritik an dem Übermaß funktionaler Köpfe geübt worden. So widerlegen Iatridou (1990) und Chomsky (1995c) die Argumente, die ursprünglich die Position Agr im Englischen motiviert haben. Van Gelderen (1993, 1997) argumentiert ferner gegen die Postulierung der funktionalen Projektionen AgrSP, AgrOP und AspP, und sie zweifelt ebenso wie Joseph/Smirniotopoulos (1993), Thráinsson (1996) und Bobaljik/Thráinsson (1998) die Universalität sämtlicher vorgeschlagener funktionaler Kategorien an. Auch Haider (1997) plädiert für eine minimale funktionale Struktur oberhalb der VP, also möglichst wenige funktionale Projektionen. Grundsätzlich kann man also die Fragen stellen, welche funktionalen Kategorien existieren und welche funktionalen Kategorien Phrasen projizieren.

Trotz dieser vielerorts geäußerten Kritik fehlen bisher verlässliche Kriterien für die Auswahl und Legitimation funktionaler Köpfe. Da die Einstufung eines funktionalen Elements als Kopf einer Phrase in der Regel von der Konzeption des Begriffes 'Kopf' abhängt, ist es dringend erforderlich, die Definition des Begriffes 'Kopf einer Phrase' zu klären und damit unter Umständen auch den Status einiger funktionaler Kategorien als Köpfe zu verifizieren oder zu falsifizieren. Dass dies immer noch eine notwendige Aufgabe ist, zeigt die einleitende Bemerkung Frasers, Corbetts und McGlashans in *Heads in Grammatical Theory*: "The majority of grammatical theories refer explicitly to the head of a phrasal constituent. Yet while the term 'head' has entered the common currency of theoretical linguistics, this does not provide evidence of agreement on what it means." (Fraser/Corbett/McGlashan 1993: 1)

Eng verknüpft mit dieser Aufgabenstellung ist die Frage nach den Eigenschaften funktionaler Köpfe, und ob diese sich von lexikalischen in fundamentaler Weise unterscheiden. So ist unklar, ob alle Kategorien entweder als lexikalische oder funktionale Köpfe zur Verfügung stehen – eines von vielen Problemen, die bereits Leffel/Bouchard (1991: 5-6) als noch ungelöst betrachten:

> In general, a number of issues became apparent as proposals about functional elements were developed. It is unclear whether functional heads (e.g. D, C, I) and lexical heads (e.g. N, V, A) project in the same way, or in different ways, as proposed by Fukui and others. [...] There are categories that do not easily fit into the functional-or-lexical mold, such as prepositions and various kinds of conjunctions.

Als besondere Beispiele für Kategorien, die weder eindeutig lexikalisch noch funktional sind, werden häufig Präpositionen, Quantoren und Konjunktionen genannt.

Die vorliegende Arbeit soll einen Beitrag zu der Diskussion um lexikalische und funktionale Kategorien leisten. Die Arbeit konzentriert dabei sich auf funktionale Kategorien in der Nominalphrase, vor allem weil zum verbalen Bereich schon eine Vielzahl neuerer und auch kritischer Vorschläge zu funktionalen Kategorien vorliegen (vgl. Iatridou 1990, Joseph/ Smirniotopoulos 1993, van Gelderen 1993, 1997, Thráinsson 1996, Haider 1997, Bobaljik/ Thráinsson 1998). Auch für die Nominalphrase sind eine Reihe funktionaler Kategorien vorgeschlagen worden, beispielsweise D (*determiner*), Q (*quantifier*) oder Num (*number*), aber diese wurden im Gegensatz zu funktionalen Kategorien im verbalen Bereich bislang noch nicht kritisch auf ihre Berechtigung überprüft.

Aufgrund der oben aufgeworfenen Probleme wird die vorliegende Arbeit vor allem folgende Fragen versuchen zu beantworten:

a) Welchen Status haben funktionale Kategorien in der Universalgrammatik?
b) Welche funktionalen Kategorien existieren in der Nominalphrase?
c) Artikel und Demonstrativpronomina werden zu der funktionalen Kategorie "Determinierer" (D) gezählt. Welchen Status haben Quantoren wie *some*, *any*, *many* oder *all*? Gibt es eine Kategorie "Quantor" (Q)? Wenn ja, ist sie lexikalisch oder funktional? Projizieren alle Quantoren eigene Phrasen (QPs)?
d) Macht es Sinn, von einer Dichotomie lexikalisch - funktional auszugehen? Sind Quantoren nicht eher ein Beispiel für eine "Zwischenkategorie", die weder eindeutig lexikalisch noch funktional ist?
e) Welche Eigenschaften machen eine Kategorie zum "Kopf einer Phrase"? Gibt es auch Formen, die weder eindeutig Köpfe noch eindeutig Adjunkte in einer Phrase sind? Gibt es Kriterien für Köpfe und wie zuverlässig sind sie?
f) Welche Schlüsse kann man aus Erstspracherwerbsdaten zum Status von Determinierern und Quantoren ziehen?

Weil hier Fragen nach dem Status der funktionalen Kategorien in der Universalgrammatik gestellt werden, schien es sinnvoll, die in dieser Arbeit gewonnenen Erkenntnisse zu funktionalen Kategorien durch einen Blick auf den Spracherwerb zu überprüfen und zu ergänzen. Deshalb wird die Diskussion um funktionale Kategorien in der Nominalphrase um eine Analyse des Gebrauchs der Determinierer und Quantoren in der englischen Kindersprache vervollständigt. Diese Analyse ist eine wichtige Ergänzung, weil Quantoren Kandidaten für Formen sind, die weder eindeutig lexikalisch noch funktional sind und weil der Status der Quantoren als 'funktionale Köpfe' weit weniger eindeutig ist als beispielweise der der Artikel oder Demonstrativa. Deshalb wird der Erwerb der Quantoren *some*, *any*, *no*, *many*, *all* und *more* genauer untersucht und mit dem Erwerb der Demonstrativa verglichen. Mit dieser Analyse englischer Kindersprache werden gleichzeitig einige Fragen aus den ersten Kapiteln wieder aufgegriffen und um neue Fragen aus der Perspektive der Erstspracherwerbsforschung ergänzt:

- Sind Quantoren funktionale Köpfe in einer QP, vergleichbar mit Artikeln und Demonstrativa, die Köpfe einer DP sein können?
- Wenn Quantoren weder eindeutig funktional noch eindeutig lexikalisch sind, wie vollzieht sich dann ihr Erwerb? Bestehen Ähnlichkeiten zum Erwerb der Determinierer?
- Gibt es universelle Erwerbssequenzen für den Erwerb von Determinierern und Quantoren?

Die vorliegende zweigeteilte Arbeit hat also folgende Struktur: Im ersten Teil, der Kapitel 2, 3 und 4 umfasst, werden vor allem theoretische Fragen geklärt und die Rolle der funktionalen Kategorien im Minimalistischen Programm untersucht. In Kapitel 2 werden exemplarisch anhand der Nominalphrase Gründe präsentiert, warum funktionale Kategorien Phrasen projizieren. Dies wird ergänzt um eine Diskussion der Rolle funktionaler Kategorien im Minimalistischen Programm (Kapitel 2.3). In Kapitel 3 werden Kriterien für "lexikalisch", "funktional" und "Kopf einer Phrase" genauer untersucht und einer Kritik unterworfen. In Kapitel 4 werden die bisher für den nominalen Bereich vorschlagenen funktionalen Köpfe vorgestellt. Es wird untersucht, ob diese funktionalen Köpfe ihre Berechtigung haben, beispielsweise D (*determiner*), Q (*quantifier*), Num (*number*), Agr (*agreement*), Poss (*possessor*). Dabei fließen sowohl Ergebnisse aus Kapitel 3 als auch Prinzipien des Minimalistischen Programms in die Bewertung ein. Speziell für die englische Nominalphrase werden verschiedene Determinierer und Quantoren analysiert. Dabei wird versucht zu klären, ob diese Formen lexikalisch, funktional oder eher "semi-lexikalisch" sind und ob englische Quantoren Köpfe eigener Phrasen (QPs) sind.

Der zweite Teil der Arbeit befasst sich mit der Rolle funktionaler Kategorien im Erstspracherwerb. Auch liegt der Schwerpunkt wieder auf der englischen Nominalphrase. Es wird zunächst nach Entwicklungsmodellen zum Erstspracherwerb gefragt und der Erwerb funktionaler Kategorien gemäß diesen Modellen vorgestellt und diskutiert (vgl. Kapitel 5). Im Anschluss daran wird der Erwerb der "Zwischenkategorie" Quantor genauer untersucht und mit dem Erwerb der Demonstrativa verglichen (vgl. Kapitel 6.2 bis 6.4). Hier interessieren vor allem Fragen wie: Werden Determinierer und Quantoren gleichzeitig erworben? Viele Determinierer werden ab dem Alter von etwa zwei Jahren als Köpfe von DPs verwendet (vgl. Radford 1990, 1995), einzelne Formen und deren Vorläufer eventuell auch schon drei Monate früher (vgl. Stenzel 1997). Ab welchem Alter werden Quantoren als Köpfe von QPs verwendet? Werden Quantoren konsequent als Köpfe interpretiert oder gibt es Übergangsphasen? Gibt es universelle Erwerbssequenzen für Quantoren? Ferner wird versucht, die Einordnung von Q als "semi-lexikalische Kategorie" mit den Ergebnissen aus dem Erstspracherwerb in Beziehung zu setzen (vgl. Kapitel 6.5). Abschließend wird in Kapitel 6.5 nach Konsequenzen aus den Erstspracherwerbsdaten für das Minimalistische Programm und das Modell der Nominalphrase gefragt, womit sich der Bogen schließt: Die theoretischen Erkenntnisse und die empirischen Daten (hier: Kindersprachdaten) sollen dazu beitragen, einen Überblick über den Status funktionaler Kategorien in der Universalgrammatik zu gewinnen.

1.2 Bemerkungen zur methodischen Vorgehensweise

Wie bereits im ersten Abschnitt erläutert wurde, soll diese Arbeit u.a. einen Beitrag zur Diskussion um funktionale Kategorien innerhalb des Minimalistischen Programms bieten. Vor allem die Kapitel 2, 4, 5 und 6 zur Nominalphrase und zum Erstspracherwerb werden sich innerhalb dieses theoretischen Rahmens bewegen. In Kapitel 3 allerdings, in dem die theoretischen Grundlagen für weitere Analysen gelegt werden sollen, wird von diesem Grundsatz abgewichen, und zwar aus folgendem Grund: Während der Erarbeitungsphase, in der die Analyse des Konzeptes 'Kopf einer Phrase' erstellt wurde, stellte es sich sehr schnell als nützlich heraus, auch Ergebnisse aus nicht-generativen Theorien heranzuziehen (Grammatikalisierungs-Ansätze, kognitive Linguistik). Zum einen ließen sich so wichtige Erkenntnisse aus anderen Theorien für die generative Syntax gewinnbringend nutzen, zum anderen machte es wenig Sinn, sich ausschließlich auf die Government/Binding-Theorie bzw. das Minimalistische Programm zu beschränken, wenn ein Konzept wie der 'Kopf der Phrase' untersucht wird, das in (fast) jeder aktuellen syntaktischen Theorie seinen Stellenwert hat. Die Übernahme einiger Ideen aus der kognitiven Linguistik (Lakoff 1987) sowie aus verschiedenen Ansätzen zur Grammatikalisierung (*grammaticalization*, Hopper/Traugott 1993) waren sinnvoll, um das Problem aus nicht-generativer bzw. diachroner Sicht zu beleuchten.

Bei der Auswahl der Daten wurde ebenfalls auf methodologische Vielfalt geachtet. Während sich viele generative Linguistinnen und Linguisten im allgemeinen auf ihre sprachliche Intuition verlassen und ihre Thesen mit selbsterdachten Beispielen belegen, sind in dieser Arbeit auch Korpusdaten herangezogen worden.

Dies ist in der generativen Theoriebildung häufig unüblich, weil Chomsky auf den Unterschied zwischen Kompetenz und Performanz verweist, also den Unterschied zwischen sprachlichem Wissen und dem Gebrauch der Sprache in realen Alltagssituationen, und das Interesse der generativen Forschung gilt der Kompetenz eines Individuums. Korpusdaten können eventuell problematisch sein, weil sie als Widerspiegelungen der u.U. fehlerhaften Performanz keine genaue Auskunft über die Kompetenz des Sprechers/der Sprecherin geben. In vielen Arbeiten wird daraus der Schluss gezogen, dass die sprachliche Intuition über die Wohlgeformtheit von Sätzen (*sentence well-formedness*) als Evidenz für die Formulierung linguistischer Theorien adäquater ist, da sprachliche Intuitionen Kompetenz eher reflektieren als Korpusdaten.

Auch Chomsky (1965: 18-27) diskutiert die Frage, ob Introspektion ein ausreichendes Mittel ist, um linguistische Theorien zu rechtfertigen. Die Beschränkung auf Introspektion scheint – so Chomsky – kein entscheidendes Problem für linguistische Theorien, da allein durch die Introspektion reichlich Evidenz zur Verfügung steht. Es ist die Aufgabe linguistischer Theorien, diese Evidenz angemessen zu beschreiben (Chomsky 1965: 19-20). Eine Weiterentwicklung von Methoden zur objektiven Erhebung sprachlicher Daten ist offensichtlich kein vorrangiges Ziel für Chomsky:

> In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. One who disagrees with this estimate of the present situation in linguistics can justify his belief in the current importance of more objective operational tests by showing how they can lead to new and deeper understanding of linguistic structure. (Chomsky 1965: 20-21)

An dieser Einstellung hat sich auch über zwanzig Jahre später zumindest für Chomsky wenig geändert: "You don't take a corpus, you ask questions. You do exactly what they do in the natural sciences. You do experimentation. [...] You want an answer to a non-trivial question; you got to go beyond corpus data." (Chomsky, p.c., in: Aarts 1999)[5]

Eine direkte Antwort auf diese Einwände gegen den Gebrauch von Korpora bieten Fillmores (1992: 35) Beobachtungen:

> I have two main observations to make. The first is that I don't think there can be any corpora, however large, that contain information about all of the areas of English lexicon and grammar that I want to explore; all that I have seen are inadequate. The second observation is that every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way. My conclusion is that the two kinds of linguists [d.h. der Korpuslinguist und der theoretische Linguist, SR] need each other.

Die Motivation für den Einsatz von Korpusdaten aus dem British National Corpus und der CHILDES-Datenbank in der vorliegenden Arbeit lässt sich kaum besser formulieren. Dennoch möchte ich noch einige Punkte ergänzen, die für den Einsatz eines Computerkorpus sprechen:

- Ausgangspunkt für diesen methodischen Zugang ist eine Überlegung von Stig Johansson (1991: 4): "[...] computer corpora provide data which can serve as input to descriptive linguistic work and as a testing ground for linguistic theories." Vor allen als Testmaterial für Hypothesen zu funktionalen Kategorien in der Nominalphrase scheinen m.E. Korpusdaten sehr wertvoll zu sein. In Kapitel 4 sind aus diesem Grund gezielt Korpusdaten zu einzelnen Fragestellungen herangezogen worden.
- Korpora werden in der Kindersprachforschung den experimentellen Daten für das frühe Kindesalter vorgezogen – dies hängt u.a. damit zusammen, dass Kinder im entscheidenden Alter nicht geeignet für experimentelle Studien sind.[6] Trotz der "Beschränkung" auf Korpusdaten sind in letzter Zeit eine Fülle von neuen Ergebnissen gerade für die generative Theoriebildung aus der Kindersprachforschung geliefert worden (vgl. z.B. Radford 1990, Meisel 1992, Hoekstra/Schwartz 1994, Clahsen 1996, Dittmar/Penner 1998, Meibauer/Rothweiler 1999). Warum sollte ein ähnliches methodisches Vorgehen für die Erwachsenensprache nicht sinnvolle Ergebnisse liefern? Heutige Korpora von einem Umfang von 100 Millionen Wörtern sind weniger beschränkt und stellen mehr einen repräsentativen Querschnitt durch die Sprache dar als die Korpora der fünziger und sechziger Jahre. Leistungsstärkere Computer erleichtern darüber hinaus die automatische Recherche in solchen Computerkorpora.

---

[5] In Chomsky (1995a) wird das Problem der Datenerhebung nicht angeschnitten. Die verwendeten Daten beruhen ausschließlich auf Sprecherintuitionen, so dass wenig Zweifel über eine unveränderte Bevorzugung der Introspektion übrig bleiben.

[6] Vgl. auch Wagner (1996) zur Rolle des Korpus in der Erstspracherwerbsforschung.

- Es ist eine Tatsache, dass die Kompetenz (das sprachliche Wissen des Individuums) stets nur über die Performanz zugänglich ist, denn es existieren keine technischen Möglichkeiten, die Kompetenz eines Individuums direkt zu überprüfen.[7] Auch selbst erdachte Beispiele bzw. ein Urteil über deren Grammatikalität reflektieren Kompetenz nicht ungebrochen und sind keineswegs absolut zuverlässig. In vielen Fällen kann es sogar sein, dass eine Sprecherin bzw. ein Sprecher zu verschiedenen Zeiten unterschiedliche Urteile zur Wohlgeformtheit von Ausdrücken abgibt. Deshalb macht es Sinn, Grammatikalitätsurteile durch Korpusdaten zu ergänzen. Korpusdaten haben zudem den Vorteil, dass sie im Gegensatz zu den subjektiven Grammatikalitätsurteilen einzelner Sprecher/innen eher "objektiven" Analysemethoden unterworfen werden können: Die verwendeten Korpusmaterialien sind in der Regel frei zugänglich und damit nachprüfbar; ferner können Strukturen quantitativ analysiert werden. Quantitative Analysen wurden vor allem in Kapitel 6 genutzt, um Aussagen zur Produktivität einzelner Formen und Konstruktionen in der Kindersprache machen zu können.
- Der Versuch, in Korpora sprachliche Daten aus möglichst vielen Bereichen des alltäglichen Lebens zu erfassen, ist eine gute Ergänzung zur Introspektion des Linguisten, der zwar prinzipiell alle Daten "in seinem Kopf" zur Verfügung hat, aber auf diese nicht systematisch zugreifen kann, wie McEnery/Wilson (1993: 12-13) feststellen:

  > We must also consider that the process of introspection may not be at all systematic and in fairness look at the extremes reached by the non-corpus linguist. If the corpus linguist can often seem the slave of the available data, so the non-corpus linguist can be seen to be at the whim of his or her imagination.

- Eine Einbeziehung von Korpusdaten für eine generative Arbeit scheint durch Aussagen Chomskys wie die folgende zumindest ansatzweise legitimiert zu sein:

  > In principle, evidence concerning the character of the I-language and the initial state [of the language faculty] could come from many different sources apart from judgements concerning the form und meaning of expressions: perceptual experiments, the study of acquisition and deficit or of partially invented languages such as creoles [...], or of literary usage or language change, neurology, biochemistry, and so on. [...] As in the case of any inquiry into some aspect of the physical world, *there is no way of delimiting the kinds of evidence that might, in principle, prove relevant.* (Chomsky 1986a: 37; meine Hervorhebung)

Abschließend muss noch betont werden, dass trotz der Verwendung von Korpusdaten die erste Hälfte der vorliegenden Arbeit keine korpuslinguistische Studie im engeren Sinne ist. Für die Erwachsenensprachdaten wurde darauf verzichtet, detaillierte Studien zu Nominalphrasenstrukturen und Häufigkeiten einzelner Formen durchzuführen, weil zu diesem Thema bereits umfangreiche Literatur aus der Korpuslinguistik existiert bzw. in Arbeit ist.[8] Statt dessen wurden in der vorliegenden Arbeit theoretische Grundlagen für die Beschreibung der DP ausgelotet und Korpusdaten als Zusatzinformationen genutzt. Für die Analyse der Kindersprache im zweiten Teil der Arbeit sind die Korpusdaten dagegen zentral, denn es wurden detaillierte quantitative und qualitative Analysen zum Gebrauch der Determinierer und Quantoren sowie der Struktur der Nominalphrase durchgeführt, weil vor allem im

---

[7] Darauf weist auch Chomsky (1965: 18-19) hin.

[8] Vgl. beispielsweise Abberton (1977), Estling (1999) und Nilsson (1999) sowie die detaillierten korpuslinguistischen Studien von Wischer (1997) und Meyer (1999).

Bereich der Quantoren bisher wenige Studien für Kinder unter vier Jahren vorliegen (vgl. Kapitel 5.1).

Insgesamt setzt sich also die vorliegende Arbeit über einige für Jahrzehnte fundamentale Grundannahmen generativer Ansätze hinweg:

- Sprache wird nicht ausschließlich synchron als Gebilde von Kontrasten betrachtet (strukturalistisches Erbe der generativen Syntax), sondern es werden auch einige diachrone Daten herangezogen.
- Es wird nicht ausschließlich mit Sprecherintuitionen gearbeitet.

Damit werden keine neuen Wege beschritten, sondern die Arbeit befindet sich durchaus im Rahmen einer neueren Entwicklung innerhalb der generativen Tradition: Neben synchronen Betrachtungen gewinnen zunehmend diachrone Untersuchungen für die generative Syntax an Bedeutung (vgl. beispielsweise van Gelderen 1993, Brandner/Ferraresi 1996). Ebenso hat die generative Erstspracherwerbsforschung in den letzten Jahren einen Boom erlebt und damit haben Korpusdaten (bzw. generell nicht-introspektive Daten) neben Sprecherintuitionen einen entscheidenden Stellenwert gewonnen. In den folgenden beiden Abschnitten (1.3 und 1.4) werden die verwendeten Korpora genauer vorgestellt.

1.3 Die Erwachsenensprachdaten: das British National Corpus

Das für diese Arbeit verwendete British National Corpus ist ein etwa 100 Millionen Wörter umfassendes Korpus der modernen englischen Sprache. Etwa zehn Prozent dieses Korpus beruhen auf (orthographischen) Transkriptionen gesprochener Sprache; die übrigen neunzig Prozent des Korpus bestehen aus Texten unterschiedlicher Gattungen. Die Kompilierung des Korpus erfolgte für ein Konsortium mehrerer Institutionen, darunter mehrere britische Verlage sowie die British Library und Forschungszentren in Lancaster und Oxford. Die Kompilierung wurde mit dem erklärten Ziel verfolgt, das heutige britische Englisch in einem möglichst umfassenden repräsentativen Querschnitt einzufangen. "The BNC was designed to characterize the state of contemporary British English in its various social and generic uses." (Aston/Burnard 1998: 28) Die BNC-Forschungsgruppen in Lancaster und Oxford versuchten dieses ehrgeizige Ziel unter anderem dadurch zu erreichen, dass sie auf eine möglichst breite Streuung in Bezug auf Texttypen im geschriebenen und gesprochenen Teil und in Bezug auf die regionale Herkunft der Sprecher/innen im gesprochenen Teil geachtet haben. Die Texte im geschriebenen Teil des Korpus gehören beispielsweise folgenden Kategorien an:

Tabelle 1: Kategorien des British National Corpus (vgl. Aston/Burnard 1998: 29)

	texts	percentage	words	percentage
Imaginative	625	19.47	19664309	21.91
Arts	259	8.07	7253846	8.08
Belief and thought	146	4.54	3053672	3.40
Commerce and finance	284	8.85	7118321	7.93
Leisure	374	11.65	9990080	11.13
Natural and pure science	144	4.48	3752659	4.18
Applied science	364	11.34	7369290	8.21
Social science	510	15.89	13290441	14.80
World affairs	453	14.11	16507399	18.39
Unclassified	50	1.55	1740527	1.93

Für die gesprochene Teilkomponente des Korpus wurden im wesentlichen zwei Kompilierungsstrategien angewandt: Einerseits wurde Material gesammelt, indem man spontane Äußerungen von Sprechern in informellen Situationen sammelte. Die Auswahl der Sprecher erfolgte systematisch nach ihrer sozialen Herkunft, ihrer geographischen Heimat, ihres Geschlechtes und ihres Alters, um einen möglichst genauen Querschnitt zu erhalten. Andererseits wurde ebensoviel Material nach dem Kriterium des Kontextes gesammelt; hier handelt es sich um eher formale Anlässe wie Reden, Debatten, Seminare oder Radioprogramme, die nach Thema und Interaktionstyp klassifiziert wurden. Auch hier wurde versucht, ein möglichst ausgewogenes Verhältnis zwischen den einzelnen Kategorien zu erreichen (vgl. Aston/Burnard 1998: 31).

Die Kodierung und Annotation des Korpus erfolgte nach den Vorgaben der *Text Encoding Initiative* (TEI) unter Verwendung von SGML[9]. Dies sollte vor allem die vielseitige Verwendbarkeit des Korpus unter möglichst vielen Hardware- und Software-Umgebungen gewährleisten. Das BNC ist mit Wortklassen-Tags versehen; darüber hinaus stehen durch die Kodierung mit SGML-Tags dem Benutzer des British National Corpus eine Reihe von abrufbaren Informationen zum Text bzw. zu den Sprechern gesprochener Texte zur Verfügung. Der folgende Satz ist ein typisches Beispiel eines getaggten Satzes aus dem British National Corpus:

[9] SGML, Standard Generalized Markup Language, wird zur Kodierung von Eigenschaften elektronischer Texte verwendet. Die Grundidee besteht darin, physikalische Eigenschaften des Textes (z.B. Fettschrift oder Absätze) sowie jegliche Zusatzinformationen (z.B. Autor, Publikationsjahr) in sogenannten *tags* (Kürzel in spitzen Klammern) zu kodieren. Die Texte sind damit – ähnlich wie Html-Dokumente im World Wide Web – unter verschiedenen Softwareprogrammen und unterschiedlichsten Hardwarevoraussetzungen lesbar und verwendbar.

```
<hit text=A14 n=812><p> <s n=0812> <c PUQ>&bquo<w CJC>And <w
PNP>they <w VM0>will <w VVI>forget <w AT0>the <w NN1>bread <w
CJC>and <w NN1-VVB><hi>butter</hi> <w NN1>customer <w PNQ>who <w
VVZ>drinks <w CRD>three <w PRP>to <w CRD>four <w NN2>pints <w
AT0>a <w NN1>day<c PUN>, <w CRD>seven <w NN2>days <w AT0>a <w
NN1>week<c PUN>.<c PUQ>&equo </p></hit>
```

In diesem Rahmen soll nicht weiter auf die weiteren Eigenschaften des British National Corpus eingegangen werden; für weitere Informationen sei der/die Leser/in auf Burnard (1995) und Aston/Burnard (1998) verwiesen. Zu SGML und zur *Text Encoding Initiative* gibt es eine Reihe von Publikationen, vergleiche beispielsweise Ide/Véronis (1995).

Für die konkrete Verwendung des British National Corpus hat die Kompilierung und Annotation einige Konsequenzen. Durch das automatische Tagging ist ein kleiner Teil des Korpus falsch oder nicht eindeutig mit Wortklassen versehen. Negativ hat sich dies vor allem auf die Suche nach den Nominalphrasenstrukturen mit dem Demonstrativpronomen *that* ausgewirkt, da sich die Suche nach *that* in seiner ausschließlichen Funktion als Demonstrativpronomen als unmöglich erwies: Die Suche nach diesem *that* ergab stets, dass auch alle Token des Complementizers *that* gelistet wurden.

Die Recherche nach Nominalphrasenstrukturen und Determinierern erfolgte mit dem Konkordanzprogramm SARA, das mit dem British National Corpus mitgeliefert wurde (vgl. auch Burnard 1996, Aston/Burnard 1998). SARA macht es möglich, sowohl nach Worten, Wortkombinationen (Phrasen) als auch nach Wörtern unter spezifischer Angabe der Wortklasse zu suchen. Darüber hinaus verfügt SARA über weitere Funktionen wie beispielsweise die gezielte Suche nach SGML-Tags oder die Verbindung mehrerer Suchtypen in einem *Query builder*. Die in den folgenden Kapiteln verwendeten Zahlenangaben und Beispiele basieren auf den Ergebnissen meiner Recherchen mit SARA. Die Quellenangaben für Beispielsätze haben dabei stets dasselbe Format: eine Zahlen/Buchstabenkombination, die auf einen Text im BNC verweist, gefolgt von der Satznummer in dem entsprechenden Text. Wo es erforderlich ist, wird an entsprechender Stelle auf offensichtliche Mängel des Korpus, der Korpusannotation oder des Konkordanzprogrammes SARA hingewiesen, die zu ungenauen oder fehlerhaften Ergebnissen geführt haben oder geführt haben könnten.

1.4 Die Kindersprachdaten: das CHILDES-System

Im folgenden soll ein kurzer Überblick über die Art der verwendeten Kindersprachdaten gegeben werden. Ergänzend dazu werden noch einige Prinzipien der Datenerhebung in der Spracherwerbsforschung vorgestellt, denen im wesentlichen im anschließenden gefolgt wird. Darüber hinaus wird die CHILDES-Datenbank beschrieben, aus der der Großteil der in dieser Arbeit verwendeten Kindersprachdaten stammt.

1.4.1 Prinzipien der Datensammlung und Dateninterpretation

Grundsätzlich lassen sich zwei Arten von Datensammlungen in Spracherwerbsstudien unterscheiden: sogenannte naturalistische[10] Datensammlungen spontaner Äußerungen des Kindes in seiner gewohnten häuslichen/schulischen Umgebung und Datenerhebungen durch experimentelle Studien in einem Sprachforschungslabor unter kontrollierten Bedingungen. Die erstgenannten naturalistischen Studien blicken auf eine lange Geschichte zurück: In den Anfängen der Erstspracherwerbsforschung sind Ergebnisse bzw. Beobachtungen eigener Kinder von Linguist/inn/en in der Form von Tagebüchern festgehalten worden. Heutzutage werden im allgemeinen Sitzungen mit Kindern bevorzugt, in denen die Forscher/innen die Äußerungen und Verhaltensweisen der Kinder mit einem Rekorder oder einer Videokamera festhalten und dabei zusätzliche Notizen machen. Die Äußerungen werden anschließend transkribiert und zusammen mit Bemerkungen über den kommunikativen Kontext in einem Korpus[11] gesammelt. Die Häufigkeit der Sitzungen und die Zahl der Kinder kann dabei stark variieren, beispielsweise von Langzeitstudien von wenigen, häufig nur ein bis drei Kindern (z.B. Bloom 1970, Brown 1973) bis zu einer Reihe von Sitzungen mit einer großen Zahl von Kindern verschiedener Alters- und Entwicklungstufen (Querschnittsstudien, z.B. Wells 1985).

Für naturalistische Studien sind viele Vor- und Nachteilen genannt worden (vgl. beispielsweise Mißler 1993: 10-12, Demuth 1996: 20-22, Stromswold 1996: 24-26). Als Vorteil wird unter anderem die Möglichkeit genannt, in Langzeitstudien Entwicklungslinien zu verfolgen und den sprachlichen Input aus der Umgebung des Kindes mitzustudieren. Ferner können die sprachlichen Äußerungen des Kindes, das sich in seiner gewohnten Umgebung (zu Hause, in der Schule/Kindergarten) befindet, in ihrer spontanen "natürlichen" Form aufgezeichnet werden, während unter Laborbedingungen immer die Gefahr herrscht, dass Kinder nicht ihre "normalen" sprachlichen Fähigkeiten zeigen. Ein großer Nachteil naturalistischer Studien sind mögliche zufällige Lücken im Korpus zu einzelnen grammatischen Phänomenen, die durch die Beschränkung auf die reine Beobachtung spontaner Äußerungen entstehen. Studien, die sich auf die Entwicklung von ein bis zwei Kindern konzentrieren, laufen außerdem Gefahr, Phänomene überzubewerten, die lediglich für einzelne Kinder idiosynkratisch sind. Schließlich nennt insbesondere Stromswold (1996: 26) noch die Schwierigkeit, die Bedeutung naturalistischer Äußerungen richtig zu interpretieren, vor allem wenn Forscher/innen die Daten anderer übernehmen.

Trotz der genannten Nachteile werden in der vorliegenden Arbeit vor allem Daten aus naturalistischen Studien berücksichtigt; dies geschieht aus mehreren Gründen. Es wird im

---

[10] Der Terminus "naturalistisch" (*naturalistic*) wird hier in dem Sinne gebraucht, wie ihn Radford (1990: 9) einführt: "A conventional alternative to experimental studies are *naturalistic studies*: typically, these involve collecting and analysing a corpus (i.e. a sample) of children's 'free-speech' [...]. The corpus is collected by making videotape or audiotape recordings of a child's spontaneous speech behaviour in a natural setting (e.g. at home) where the child is in conversation with someone he is familiar with and talks freely to (e.g. his mother)."

[11] Der Terminus "Korpus" wird hier informell für eine Datensammlung gebraucht, insbesondere die Sammlung sämtlicher Äußerungen eines Kindes – abweichend von McEnery/Wilsons Definition eines Korpus als elektronische Datenbank mit einem repräsentativem Querschnitt der Strukturen einer Sprache (vgl. McEnery/Wilson 1996).

folgenden die Entwicklung der Quantoren im frühen Kindesalter (1;2 bis 5;0) untersucht, für die bislang vor allem für das Alter vor 3;0 m.E. keine detaillierten Studien vorliegen. Der naturalistische Ansatz wird hier gewählt, weil er sich für explorative Studien wie die vorliegende Arbeit besonders eignet, in denen erste Hypothesen zu einer Entwicklung erst gebildet werden (vgl. auch Mißler 1993: 10-12). Ferner wird der naturalistische Ansatz bevorzugt, weil ein experimenteller Ansatz vor allem in den frühen kindlichen Altersstufen von ein bis drei Jahren kaum zuverlässige Ergebnisse liefert.[12] Dies wird von Forschern, die experimentelle Daten verwenden, auch bemerkt. Chiat (1986: 344) beispielsweise, die sich in ihrer Untersuchung der Personalpronomina auf experimentelle Studien stützt, gibt zu, dass deren Zuverlässigkeit eher fraglich ist:

> It is difficult to obtain reliable responses which are not experimental artifacts from children at the crucial stages of development (in the case of pronouns, around 2 years for many children). By the time the child responds reliably, the acquisition process may be well advanced, and the experiment may fail to tap it.

Die folgenden Analysen basieren zum größten Teil auf naturalistischen Daten aus der Computerdatenbank von CHILDES, wie sie beispielsweise in MacWhinney (1995b) beschrieben wird. Der Einsatz der CHILDES-Datenbank ermöglichte es, weit mehr Daten zu berücksichtigen, als es bei der bloßen Verwendung selbst erhobener Daten möglich gewesen wäre: Insgesamt wurden die Daten aus den Langzeitstudien von fünf Kindern berücksichtigt, deren Beobachtung sich z.T. über mehrere Jahre erstreckte und den Einsatz eines größeren Forscherteams erforderten. Die Verwendung der CHILDES-Daten hatte ferner den großen Vorteil, dass gerade die in spontanen Äußerungen eher seltenen Quantoren mit einem Suchprogramm automatisch in der Computerdatenbank gesucht werden konnten. Auf diese Weise konnte zumindest für einige Quantoren ein größerer Satz an Beispielen zusammengestellt werden. Die Daten aus CHILDES und die methodische Vorgehensweise für die Analyse der CHILDES-Daten werden in den folgenden Abschnitten noch ausführlicher erläutert. Über die CHILDES-Daten hinaus sind Beobachtungen und Beispiele aus der Literatur zum Spracherwerb berücksichtigt worden (z.B. Brown 1973, Bowerman 1973, Bloom 1973, Halliday 1975, Ede/Williamson 1980, Hill 1983, Radford 1990). Auf einige Ergebnisse experimenteller Ansätze wird zur Ergänzung verwiesen.

Bei der Sichtung der Korpora und Beispiele hat sich herausgestellt, dass zum Erwerb der ersten Determinierer in der Kindersprache (Demonstrativpronomina, Personalpronomina,

---

[12] Einen Kompromiss zwischen dem naturalistischen und dem experimentellen Ansatz versprechen Studien, in denen im naturalistischen Setting versucht wird, durch geplante Aktivitäten und Gespräche dem Kind die Konstruktionen von Interesse zu „entlocken" (elizitieren). Die Elizitation verbindet das (fast) einflussfreie familiäre Setting naturalistischer Studien mit der gezielten Untersuchung einzelner Phänomene, wie sie in experimentellen Studien üblich ist. Elizitation verkleinert damit die zufälligen Lücken im Korpus unter gleichzeitiger Beibehaltung der natürlichen Umgebung für das Kind. Dieser Ansatz ist allerdings fast ebenso aufwendig wie eine experimentelle Studie vorzubereiten und daher für eine große Zahl von Kindern nur mit außerordentlich großen Ressourcen und Personal durchführbar. Ferner ist Elizitation laut Thornton (1996: 81) ebenso wie experimentelle Studien für jüngere Kinder unter drei Jahren kaum geeignet, wenn man zuverlässige Daten erheben möchte. Weitere Details zur Elizitation und ihrer Durchführung bieten Eisenbeiß et al. (1994) und Thornton (1996); eine allgemeine Diskussion zur Methodik in der Erstspracherwerbsforschung findet man in Mißler (1993) und Crain/Thornton (1998).

Artikel) relativ viele Daten und Untersuchungen vorliegen, aber für Quantoren ist die Datenlage wesentlich schwieriger. Zu einigen Quantoren können fast gar keine Aussagen gemacht werden, weil sie selbst in CHILDES extrem selten vorkommen. Deshalb ist versucht worden, sich auf die Quantoren *some, any, no, many, (a) few, much, more* und *all* im wesentlichen zu konzentrieren. Diese Tatsachen müssen natürlich bei der Einschätzung der folgenden Analysen berücksichtigt werden. In Kapitel 6 kann es sich nur um die Darstellung von Tendenzen handeln, wie sie aufgrund der aktuellen Datenlage plausibel erscheinen.

Da die im folgenden untersuchten Kinderäußerungen ausschließlich aus Datensammlungen stammen, die nicht von mir erhoben wurden, mussten einige Vorsichtsmaßnahmen bei der Dateninterpretation besonders beachtet werden. Wichtig bei der anschließenden Analyse der Daten ist vor allem, zwischen formelhaften oder semi-formelhaften Äußerungen und spontaner Sprache zu unterscheiden. Unter "formelhaft" wird hier ein Ausdruck verstanden, der vom Kind geäußert wird, ohne dass Hinweise darauf bestehen, dass das Kind den Ausdruck segmentieren kann und Wissen über die morphologischen, syntaktischen und semantischen Eigenschaften der Konstituenten des Ausdrucks hat. Als Beispiel für eine formelhafte Äußerung nennt Radford (1990: 16) einen Ausruf von Kate (22 oder 23 Monate), den sie als Antwort auf die Türklingel oder das Telefon gibt: [*agedit*] (= 'I'll get it'). Da der Ausruf wie eine "Formel" nur in diesem Kontext verwendet wird und Kate ansonsten nie Pronomina oder *will* benutzt, ist *agedit* für Kate wohl ein unanalysiertes Ganzes. Es kann aus dem Gebrauch von *agedit* nicht geschlossen werden, dass Kate schon Pronomina (wie *I*) oder Auxiliarverben (wie *will*) beherrscht.

Mit ähnlicher Vorsicht sind Imitationen der Erwachsenensprache durch Kinder zu betrachten, da auch diese keine Aussagen über die tatsächliche Kompetenz der Kinder erlauben. Interessanter sind da schon nicht-perfekte Imitationen, weil sie Rückschlüsse auf die Entwicklung des Kindes gestatten: "Such examples suggest that when a child produces imperfect imitations of adult speech, he ist not simply repeating unprocessed sound-sequences, but rather is 'processing heard speech according to his own inner structures'." (Radford 1990: 18). Um formelhafte Äußerungen und Imitationen nicht mit produktiven Verwendungen einzelner Formen zu verwechseln, ist bei der Analyse eines Beispiels stets der Kontext (also mindestens zwei Sätze vor und nach der Äußerung) berücksichtigt worden. Der Kontext liefert in der Regel ausreichend Aufschlüsse darüber, ob eine Imitation vorliegt. Im Fall der formelhaften Äußerungen wurde geprüft, ob über längere Zeit eine Form lediglich in einer festgelegten Konstruktion geäußert wurde, die ansonsten nicht produktiv verwendet wurde. War dies der Fall, so wurde die Äußerung als tendenziell formelhaft bewertet. Das zu CHILDES gehörige Computerprogramm CLAN bietet bequeme Möglichkeiten, nach Äußerungen mit ihrem Kontext in einem Kindersprachkorpus systematisch zu suchen, wie im folgenden genauer geschildert wird.

1.4.2 Das CHILDES-System

1.4.2.1 Datenbank, Kodierungssystem und Software

CHILDES (= *Child Language Data Exchange System*) besteht als Projekt mit dem Ziel des Austausches von Kindersprachdaten und der Entwicklung und Standardisierung von Kodie-

rungssystemen und Software seit Beginn der achtziger Jahre (vgl. MacWhinney 1995a, 1995b, 1996). Im Vordergrund dieses Projektes stehen drei Komponenten des Systems: eine Datenbank, ein Kodierungssystem für Kindersprachdaten (CHAT) und ein Softwarepaket (CLAN). Die Datenbank von CHILDES umfasst hauptsächlich Kindersprachdaten aus dem englischen Sprachraum, stellt aber auch Daten zu 19 weiteren Sprachen für Forschungszwecke frei zur Verfügung. Im Gegenzug dazu ermutigt das Projekt seine Nutzer/innen, eigene Kindersprachdaten an das CHILDES-Projekt weiterzuleiten, damit die Datenbank auch in der Zukunft weiter wächst.[13] Eine neuere Entwicklung für CHILDES ist die wachsende Zahl von Beiträgen zum Zweitspracherwerb, bilingualem Erstpracherwerb sowie Studien zu Sprachstörungen und Aphasie bei Kindern und Erwachsenen (vgl. MacWhinney 1996).

Wenn man CHILDES lediglich als umfangreiche Datenbank bezeichnete, würde man dem Projekt kaum gerecht werden. Eine weitere wichtige Errungenschaft von CHILDES ist die Schaffung eines Standardtranskriptionsformates und einer Diskursnotation für Kindersprachdaten namens CHAT. Alle in der CHILDES-Datenbank enthaltenen Dateien sind in diesem Kodierungsformat CHAT erstellt bzw. wurden nachträglich in diesem Format erfasst. Dieses Vorgehen ist vor allem vorteilhaft, weil die einheitliche Kodierung intersubjektiv nachvollziehbar ist und die Daten leichter vergleichbar sind. Das Transkriptionsformat lässt sich relativ flexibel an detaillierte oder weniger detaillierte Transkriptionen anpassen, aber auf jeden Fall werden für jede Datenerfassung in CHAT einige Grundprinzipien beachtet (vgl. MacWhinney 1996):

- Jede Äußerung wird als einzelner Eintrag in der Datei abgespeichert.
- Zusätzliche Kodierungen werden separat von der Transkription der Äußerung (der *'main line'*) abgelegt. CHILDES ermöglicht die Kodierung von phonologischen, morphologischen und syntaktischen Informationen ebenso wie die systematische Kodierung von Sprechakten und Sprechfehlern in separaten Zeilen (*lines*).
- In der Transkription der 'main line' werden möglichst nur Wortformen in ihrer Standardorthographie verwendet – dies erleichtert die Wortsuche mit dem Computer. Da Kinder in ihrer Verwendung und Aussprache von diesen Standardformen oft abweichen, bietet CHAT eine Reihe von Optionen, diese Abweichungen zusätzlich festzuhalten.

Das folgende Beispiel illustriert, wie eine Interaktion aussieht, wenn sie im CHAT-Format transkribiert wurde:

*CHI: train # train # penny .
*MOT: no more pennies # you don't have any in your pocket # all gone .
*CHI: all gone .
*PAT: is your pocket empty ?
*CHI: xxx # xxx .

Hier ist lediglich orthographisch transkribiert worden. Die Äußerungen von Mutter und Kind sind jeweils eindeutig durch CHI (=child) und MOT (=mother) gekennzeichnet, ebenso die Äußerung einer dritten anwesenden Person (PAT). Die Einteilung in Sprecherebenen (*speaker tiers*) erlaubt die gezielte Suche nach Äußerungen eines vorher festgelegten Sprechers, beispielsweise des Kindes oder eines Elternteils. Über die orthographische

[13] Diese Idee des Datenaustausches unter Linguist/innen geht auf Roger Brown zurück, der seine Kindersprachdaten weiteren Forscher/innen zur Analyse frei zur Verfügung stellte (vgl. MacWhinney 1996).

Transkription hinaus werden einige Symbole verwendet, um schwer oder nicht identifizierbares Material zu kodieren. Zu den häufigsten Symbolen für nicht-identifizierbares Material gehören die folgenden:

Tabelle 2: Symbole und ihre Bedeutung in der CHILDES-Datenbank
(vgl. MacWhinney 1995a: 39-75)

Symbol	**Bedeutung**
xxx	unverständliche Äußerung/unverständliches Wort
&	es folgt die phonologische Form einer unvollständigen oder unverständlichen phonologischen Sequenz
#	Pause
##	längere Pause
@c	Ad-hoc-Wortbildung
@	Wort ist keine Standardform in der Erwachsenensprache
[ ]	eckige Klammern enthalten Kommentare
< >	spitze Klammern enthalten das Material, auf das sich der Kommentar bezieht
[/]	Wiederholung, ohne die vorherige Äußerung zu ändern
[//]	Wiederholung einer Äußerung mit veränderter Syntax aber gleichgebliebener Bedeutung
[?]	vorangegangenes Wort wurde bei der Transkription geraten
[!]	Betonung
["]	Sprecher/in macht metalinguistische Referenz
- .	fallende Intonationskontur
+...	Sprecher/in verstummt

Weil die Transkription der Daten auf dem einheitlichen Kodierungssystem CHAT beruht, genügt ein einziges, auf CHAT abgestimmtes Softwareprogramm, um Dateien aus der immerhin inzwischen über 160 Megabyte großen Datenbank zu durchsuchen. Dies ist das Softwarepaket CLAN, das Kindersprachdaten und ihre spezifischen Eigenarten analysieren hilft. Insbesondere können mit CLAN Häufigkeit und Kontext einzelner Morpheme und Wortformen abhängig vom Sprecher (Kind, Erwachsene) abgefragt werden, also Frequenzlisten und Konkordanzlisten erstellt werden. Im folgenden Abschnitt werden einige Komponenten von CLAN, die für diese Arbeit relevant sind, kurz vorgestellt. Einen Überblick über die gesamte Funktionspalette und die Nutzungsmöglichkeiten der CLAN-Programme bietet MacWhinney (1996); weitere ausführliche Informationen zu CLAN und CHILDES findet man in MacWhinney (1991, 1995a). Einführende Informationen zur Verwendung und zur Entwicklung des CHILDES-Systems liefern MacWhinney (1995b) und MacWhinney/Snow (1990).

1.4.2.2 Bemerkungen zur Datenerhebung und -analyse, das Programm CLAN

Wie in der Einleitung bereits erläutert wurde, sollen in der vorliegenden Arbeit die Rolle der Determinierer und Quantoren in der Nominalphrase analysiert werden, unter anderem auch in der frühen Kindersprache. Einige Fragen lauteten unter anderem:

- In welcher Reihenfolge erscheinen Determinierer und Quantoren in der kindlichen Nominalphrase?
- Zu welchem Zeitpunkt tauchen die ersten Formen auf? Wie häufig sind sie in den anschließenden Phasen der Sprachentwicklung?
- Wie früh oder spät werden Quantoren im Vergleich zu Determinierern oder Nomen verwendet?
- Welche Charakteristika zeichnet die Übergangsphase vom Auftauchen der ersten Formen bis zur vollen Beherrschung aus?
- Welche Rolle spielt die Häufigkeit der einzelnen Formen in der Erwachsenensprache?

Wie man sieht, spielen für die meisten dieser Fragen die Begriffe 'Häufigkeit' und 'Kontext' eine Rolle. Diese und andere Fragen lassen sich demnach leichter beantworten, wenn für eine Auswahl von Kindern[14] Frequenzlisten und KWIC-Konkordanzen für alle Determinierer und Quantoren vorliegen. Deshalb wurden für die vorliegende Arbeit zwei Komponenten der CLAN-Programme genutzt: FREQ und KWAL, die sich für lexikalische Analysen besonders eignen. FREQ erstellt Frequenzlisten für Wortformen, während KWAL für vom Nutzer spezifizierte Wortformen KWIC-Konkordanzen ausgibt (KWIC = *key word in context*). Beide Programmkomponenten besitzen eine Vielzahl von Optionen, so lassen sich beispielsweise die Suchanfragen für FREQ und KWAL auf bestimmte Wortformen, Sprecher/innen oder Dateien beschränken. Das folgende Beispiel zeigt einen FREQ-Output für die Wortform *many*; hier wurde die Suche auf eine Datei (nina02.cha) beschränkt, und nur die Äußerungen des Kindes Nina (CHI) wurden berücksichtigt:

```
Sun Feb 14 14:23:44 1999
FREQ (25-JUN-98) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file <nina01.cha>
  3 many
------------------------------
    1 Total number of different word types used
    3 Total number of words (tokens)
0.333 Type/Token ratio
```

KWAL bietet über die genannten Optionen hinaus auch noch die Möglichkeit, die Größe des Kontextes für das Schlüsselwort festzulegen. Das folgende Beispiel ist das Ergebnis einer KWAL-Suchanfrage nach dem Wort *this* in den Äußerungen des Kindes Naomi (NAO); für den Output wurde als Kontext zwei Zeilen vor und zwei Zeilen nach dem Schlüsselwort festgelegt:

---

[14] Zu den Kriterien dieser Auswahl wird weiter unten noch einiges gesagt.

```
----------------------------------------
*** File "h:\privat\childes\sachs\n16.cha": line 183. Keyword: this
*NAO:       cups it ?
*FAT:       I-'ll see if there is any juice left .
*NAO:       this bowl too .
*NAO:       water .
*NAO:       I-'m finished .
----------------------------------------
```

FREQ und KWAL sollten für die Analyse eines Teiles der CHILDES-Datenbank Hilfestellung leisten. Die ausgewählten Daten werden im folgenden Kapitel 1.4.3 kurz vorgestellt.

1.4.3 Die ausgewählten Kindersprachdaten

Die Daten aus CHILDES sind nach den folgenden Kriterien ausgewählt worden: Alter des Kindes, Sprache und Art der Studie. Insgesamt waren Studien mit ein- bis fünfjährigen Kindern am ergiebigsten, weil hier erst die naturalistischen Daten eine höhere Anzahl von Quantoren aufweisen. Um einen Überblick über die allmähliche Entwicklung der Quantifikation und Determination zu erhalten, wurden Daten aus naturalistischen Langzeitstudien ausgewählt. Auf eine systematische Berücksichtigung nicht-englischer Daten wurde verzichtet, weil dies den Rahmen der Arbeit sprengen würde.

Bei der Auswahl der Daten aus Langzeitstudien wurde versucht, eine Gruppe von Kindern zu finden, die vom Alter und/oder Entwicklungsstand her relativ homogen ist. Diesem Wunsch war durch die Struktur der CHILDES-Datenbank Grenzen gesetzt, aber durch die Einbeziehung von fünf Langzeitstudien wurde versucht, ein möglichst umfassendes Bild von der Entwicklung der Quantoren vom Alter 1;2 bis 5;0 zu erhalten. Alle Dateien enthielten vollständige Transkriptionen der Äußerungen des Kindes und der Personen in seiner Umgebung, so dass beispielsweise Imitationen relativ leicht identifiziert werden konnten. Zusätzlich enthielten alle Dateien Informationen zum Ort der Aufzeichnung und zu den beteiligten Personen.

Die Daten der folgenden Langzeitstudien wurden berücksichtigt: Bloom (1970) bzw. Bloom/Lightbown/Hood (1975), Brown (1973), Suppes (1973), Kuczaj (1976) und Sachs (1983). Sachs (1983) hat das sprachliche Verhalten ihrer Tochter Naomi vom Alter 1;2 bis 4;9 studiert und aufgezeichnet.[15] Es handelt sich um die transkribierten Aufzeichnungen von 93 Sitzungen, die teilweise im Abstand von mehreren Monaten, im Alter von 1;8 bis 3;5 in der Regel im Abstand von wenigen Tagen stattfanden. Kuczaj hat vom Alter von 2;4 bis 4;1 jede Woche zwei halbe Stunden die spontanen Äußerungen seines Sohnes Abe aufgezeichnet. Die Aufzeichnungen wurden ergänzt um wöchentliche Aufnahmen von je einer halben Stunde ab Abes Alter von 4;1 bis zum Alter von 5;1 (vgl. Kuczaj 1979: 3). Insgesamt handelt es sich hier um 210 Dateien transkribierter Äußerungen. Die Daten eines dritten Kindes, Nina, stammen aus der Datenbank, die Patrick Suppes zu CHILDES beigetragen hat (vgl. Suppes 1973). Nina war zu Beginn der Studie 1;11 Jahre alt und am Ende 3;3 Jahre. Die Aufzeichnungen erfolgten in der Regel im Abstand von wenigen Tagen; es gab maximal zwei Wochen Unterbrechung zwischen zwei Sitzungen. Die Daten des vierten Kindes, Eve,

[15] Im folgenden werden die Altersangaben für ein Kind immer in Jahren und Monaten angegeben.

stammen aus der bekannten Langzeitstudie von Brown (1973) und seinen Studenten, die das sprachliche Verhalten der drei Kinder Adam, Eve und Sarah über längere Zeit beobachtet haben. Eve ist das jüngste Kind der Studie: Sie war 1;6, als die Studie begann, und 2;3, als sie aus der Studie ausschied. Ungefähr jede zweite Woche fand eine zweistündige Aufnahme statt, so dass insgesamt 20 Dateien mit Daten für Eve existieren. Eve wird als sprachlich frühreifes Kind eingestuft (vgl. Brown 1973: 51-53). Dies wird sich auch unten in der Diskussion rasch zeigen: Eve ist in ihrer Entwicklung auch im nominalen Bereich ihren Altersgenossen voraus. Als letztes Kind wurde Peter aus einem Forschungsprojekt von Bloom zu "Language Development: Form and Function in Emerging Grammars" berücksichtigt (vgl. Bloom 1970, Bloom/Lightbown/ Hood 1975). Blooms Forschungsgruppe studierte das sprachliche Verhalten von vier Kindern – Peter, Eric, Gia und Kathryn. Von Peter lag das umfangreichste Material in computerlesbarer Form vor, weshalb lediglich seine Daten hier berücksichtigt wurden. Peter war 1;9 Jahre alt, als die Studie einsetzte, und 3;1 am Ende der Studie; er wurde etwa alle drei Wochen besucht und seine Äußerungen für ca. 3 bis 4,5 Stunden aufgezeichnet (vgl Bloom/Lightbown/Hood 1975: 5, 7). Die folgende Tabelle gibt noch einmal einen Überblick über die verwendeten Daten:

Tabelle 3: Die Kindersprachdaten

Kind	**Quelle**	**Alter**	**Zahl der Dateien**	**Größe des Korpus**
Naomi	CHILDES (Sachs 1983)	1;2 - 4;9	93	854 Kilobyte
Eve	CHILDES (Brown 1973)	1;6 - 2;3	20	903 Kilobyte
Peter	CHILDES (Bloom 1970, Bloom/Lightbown/Hood 1975)	1;9 - 3;1	20	2,62 Megabyte
Nina	CHILDES (Suppes 1973)	1;11 - 3;3	52	2,39 Megabyte
Abe	CHILDES (Kuczaj 1976)	2;4 - 5;1	210	1,78 Megabyte

Von allen verwendeten Studien wurden lediglich die transkribierten Daten ausgewertet. Alle Daten standen in computerlesbarer Form zur Verfügung und konnten als Dateien sowohl direkt eingesehen werden als auch mit dem Konkordanzprogramm CLAN automatisch durchsucht werden. Tonband- oder Videoaufnahmen standen nicht zur Verfügung, weshalb an einigen Stellen nur vorsichtige Einschätzungen zur Genauigkeit der Transkription oder zur Bedeutung der Kinderäußerungen gemacht werden konnten. Die Auswahl und Analyse der Daten wurden ausschließlich von mir vorgenommen. Etwaige Fehleinschätzungen sind also meine Fehler, nicht die der oben genannten Autoren.

Die in der vorliegenden Arbeit aus der CHILDES-Datenbank erhobenen Daten stammen aus Langzeitstudien weniger Subjekte (fünf Kinder). Deshalb wurden auch lediglich Methoden der deskriptiven Statistik eingesetzt, die sich für diese Arten von Daten eignen (vgl. dazu Mißler 1993: 26-34). Weil es sich um einen kleinen Satz von Probanden (fünf Kinder) handelt, ist auf Methoden der Interferenzstatistik verzichtet worden. Die vorliegende Pilotstudie müsste in der Zukunft durch eine Querschnittsstudie zu Quantoren ergänzt werden, um die hier aufgestellten Hypothesen an einer größeren Zahl von Probanden zu verifizieren.