

# Die Beschreibungskapazität von Finite State-Grammatiken

## Eine Relativierung der Inadäquatheitsbehauptungen von Chomsky und die Perspektive eines Paradigmenwechsels in der Grammatiktheorie\*

Walther Kindt

### 1 Überblick

Die Entwicklung der modernen Grammatikforschung und der heute üblichen Grammatiktheorien wurde in den 50er Jahren ganz entscheidend durch die Arbeiten von N. Chomsky beeinflusst. Die Notwendigkeit einer Abkehr von dem besonders einfachen Systemtyp, der Finite State-Grammatik, begründete Chomsky (1957) mit der Aussage, bestimmte in natürlichen Sprachen vorkommende grammatische Konstruktionen seien nicht in Finite State-Grammatiken erzeugbar. Außerdem wurde von Chomsky auch der Grammatiktyp mit der nächst höheren Komplexitätsstufe, die kontextfreie Phrasenstrukturgrammatik, als unzureichend verworfen. Beide Einschätzungen sind in verschiedener Hinsicht zu relativieren. Sie betreffen nämlich unmittelbar nur den Leistungsvergleich eindimensionaler Grammatiksysteme, also solcher Systeme, die nur eine grammatische Beziehung (z.B. die der Konstituenz) zugrunde legen. Eine ausschließliche Betrachtung solcher Systeme ist jedoch selbst inadäquat. Schon seit langer Zeit liegen empirische Ergebnisse vor, die zeigen, daß die für Sprache und Kommunikation einschlägigen Verarbeitungssysteme mehrdimensional sind. Im Rahmen meiner aktuellen Forschungen habe ich jetzt eine Theorie mehrdimensionaler Grammatiken entwickelt (vgl. Kindt 1998) und bewiesen, daß die von Chomsky und anderen Autoren als Beleg für die prinzipielle Inadäquatheit von Finite State- und kontextfreien Systemen genannten grammatischen Konstruktionen in lokalen Erweiterungen von strukturell korrespondierenden Finite State-Grammatiken erzeugbar sind. Zugleich reicht aber aus empirischen Gründen die Betrachtung von finiten Systemen aus.

---

\* Kurzfassung eines Berichts aus dem DFG-Sonderforschungsbereich 360 „Situierete Künstliche Kommunikatoren“ vom 24.3.1998.

## 2 Detaildarstellung

In Bartsch et al. (1977) werden beispielsweise Sätze des Typs

- (1) *Franz, Frieda, Berta und Paul ... sind 1,57 m, 1,73 m, 1,98 m bzw. 1,45 m ... groß.*

als einschlägige, natürlichsprachige Konstruktion gewertet, die nicht in Finite State-Grammatiken erfaßbar sei. Formal präzisiert handelt es sich um die Frage, im Rahmen welcher Grammatiktypen die Sprache  $L = \{a^n b^n \mid 1 \leq n < \omega\}$  erzeugt werden kann. Die Grundidee zweidimensionaler Grammatiksysteme basiert darauf, daß das Regelsystem von Phrasenstrukturgrammatiken in zwei Teilsysteme zerlegt wird, die die grammatischen Beziehungen in zwei partiell voneinander unabhängigen Dimensionen abbilden. Genauer gesagt handelt es sich um die Teil-Ganze-Beziehung (ein Äußerungssegment ist Teil eines anderen Segments) und die Kookkurenzbeziehung (ein Äußerungssegment kommt zusammen mit einem anderen Segment vor).

Eine einfache zweidimensionale Grammatik, in der die Sprache  $L$  erzeugbar ist, besteht aus folgenden Regeln, bei denen nur Übergänge von einem Symbol zu einem anderen vollzogen werden.

Teil-Ganze-Regeln	Kookkurenzregeln
$S^0 \rightarrow a$	$a \Rightarrow b$
$S \rightarrow a$	$a \Rightarrow S$
	$S \Rightarrow b$

Aus der Regelformulierung geht insbesondere hervor, daß das Startsymbol  $S^0$  in einer und  $S$  in beiden Dimensionen nichtterminal ist. Weiterhin ist  $a$  bzgl.  $\rightarrow$  terminal und bzgl.  $\Rightarrow$  nichtterminal. Und schließlich ist  $b$  in beiden Dimensionen terminales Symbol.

In dieser Grammatik läßt sich beispielsweise  $a^3 b^3$  folgendermaßen ableiten.

$S^0$					
a	S				b
	a	S		b	
		a	b		

Der Ableitungsprozeß erfolgt spaltenweise und zwar zunächst in der Richtung von oben nach unten. Erst wenn in der letzten Zeile nur noch terminale Symbole hinsichtlich der Teil-Ganze-Relation stehen, können noch in der Richtung von unten nach oben solche Zeilen expandiert werden, in denen am rechten Ende ein (hinsichtlich der Kookkurenzbeziehung) nichtterminales Symbol steht.

Der Zustandsraum der Grammatik besteht primär aus den vier Zuständen  $S^0$ ,  $S$ ,  $a$  und  $b$ . Allerdings gibt es zwei kontextuelle Restriktionen der Regeln. Die

Reihenfolgebestimmung für die Anwendung der Regeln  $S \rightarrow a$  und  $S \Rightarrow b$  ist von einem zweiwertigen Kontextparameter abhängig zu machen und bleibt somit im Rahmen finiter Systeme. Demgegenüber muß zur vollständigen Darstellung der Grammatik noch eine 'Fokusregel'  $b \rightsquigarrow S$  eingeführt werden, deren Anwendbarkeit davon abhängt, ob und wie oft das Symbol  $S$  mit der Regel  $a \Rightarrow S$  produziert wurde. Der Wertebereich eines entsprechenden Kontextparameters ist – wenn ganz  $L$  erzeugt werden soll – nicht beschränkbar und insofern führt die Regel  $b \rightsquigarrow S$  zu einer punktuellen Überschreitung des Rahmens finiter Systeme. Empirisch kann man aber davon ausgehen, daß Sätze des Typs (1) mit mehr als zehn Koordinationsgliedern im Subjekt nicht oder allenfalls ganz selten vorkommen. Bei einer geeigneten Wertebeschränkung des Kontextparameters bleibt die Grammatik finit. Die Nichterzeugung von ganz  $L$  als prinzipielle Inadäquatheit von Finite State-Grammatiken zu werten, ist daher nicht gerechtfertigt.

Die Sprache  $L = \{a^n b^n c^n \mid 1 \leq n < \omega\}$  bildet ein einfaches Beispiel, das nur in einem kontextsensitiven, nicht aber in einem kontextfreien System erzeugt werden kann. Wieder läßt sich eine einfache zweidimensionale Grammatik angeben, die  $L$  erzeugt.

Teil-Ganze-Regeln	Kookkurenzregeln
$S^0 \rightarrow a$	$a \Rightarrow b$
$S \rightarrow a$	$b \Rightarrow c$
	$a \Rightarrow S$
	$S \Rightarrow b$

### 3 Fazit

Generell gesehen zeigt die mit den angegebenen Regelsystemen illustrierte Entwicklung zweidimensionaler Grammatiken, daß die Strategie, relevante Phänomenbereiche systematisch in systemtheoretischem Rahmen zu erforschen, auch in der Linguistik zu neuen und überraschenden Erkenntnissen führt. Und speziell wird deutlich, daß eine Neubewertung der Beschreibungskapazität von Finite State-Systemen und naheliegenden Erweiterungen erforderlich ist. Sie wird über kurz oder lang zu einem Paradigmenwechsel in der Grammatiktheorie führen und die von Chomsky geprägte Epoche der Untersuchung eindimensionaler Systeme beenden.

#### Literatur

- Bartsch, R., J. Lenerz & V. Ullmer-Ehrich (1977): Einführung in die Syntax. Kronberg/Ts.  
 Chomsky, N. (1957): Syntactic Structures. The Hague.  
 Kindt, W. (1998): Integrative Phrasenstrukturgrammatik. Theorie und Anwendungen der mehrdimensionalen Syntax. Teil I. Report 98/3, SFB 360.