113. Varietätengrammatik

1. Sprachliche Variabilität und ihre Beschreibung

Natürliche Sprachen wie das Deutsche, Französische, Lateinische oder Chinesische sind nicht einheitlich, sondern sie setzen sich aus einer Vielzahl verwandter Varietäten — Dialekten, Soziolekten, Registern, Fachsprachen usw. — zusammen, die in manchen der sie kennzeichnenden Regelhaftigkeiten übereinstimmen, in andern hingegen nicht. Diese Variabilität stellt den Forscher, sofern er es nicht vorzieht, seine Bemühungen auf eine einzige — vielleicht besonders hoch angesehene — Varietät zu beschränken, vor eine dreifache Aufgabe:

(1) Er muß die einzelnen Varietäten in ihren verschiedenen sprachlichen Eigenschaften charakterisieren. Zu diesen Eigenschaften zählen nicht nur jene, die man gewöhnlich der Grammatik im engeren Sinne (Phonologie, Morphologie, Syntax) zurechnet, sondern auch lexikalische ebenso wie bestimmte Besonderheiten im kommunikativen Verhalten (z. B. die Wahl der Anredeform oder die Regeln des Rederechts).

(2) Er muß die einzelnen Varietäten in ihrem Verhältnis zueinander beschreiben, also inwieweit sie in bestimmten Eigenschaften übereinstimmen oder nicht übereinstimmen. Dabei ergeben sich zumindest drei Probleme. Zum ersten gibt es angesichts der eben erwähnten Bandbreite dieser Eigenschaften keine durchgängige Methode, um dies zu leisten. Zum zweiten hat die moderne Linguistik ihre Beschreibungsinstrumentarien größtenteils für homogene, idealisierte Sprachformen entwickelt; es bedarf daher zusätzlicher Mittel und Wege, um diese Instrumentarien für den Vergleich von Varietäten nutzbar zu machen. Und zum dritten sind die Unterschiede zwischen den einzelnen Varietäten oft kontinuierlich, d. h. zwei Varietäten unterscheiden sich oft nicht im Vorhandensein und Fehlen einer bestimmten Regel, sondern im Mehr oder Minder ihrer Anwendung.

(3) Schließlich muß er beschreiben, wie die innersprachliche Variabilität mit irgendwelchen außersprachlichen Faktoren zusammenhängt — Faktoren wie beispielsweise soziale Schicht, geographische Verteilung oder Besonderheiten der Redesituation, um nur drei der bekanntesten zu nennen (vgl. Nabrings 1981, Kapitel 3). Eine wichtige Dimension der Variabilität ist nicht zuletzt die Entwicklung in der Zeit — sei es eines Individuums (im Erst- oder Zweitspracherwerb) oder einer Sprachgemeinschaft. Welche Rolle diese und viele weitere denkbare außersprachliche Faktoren für die Variation tatsächlich spielen, läßt sich nicht a priori sagen, sondern es ist eine Frage der empirischen Analyse.

In der traditionellen Sprachwissenschaft spielt die Variation in der einen oder anderen Ebene eine wichtige Rolle. Dies gilt auch heute noch für einen großen Teil der empirischen Sprachforschung, wobei man natürlich, je nach Zeit, Geld und Interesse, nicht die gesamte Variation in allen möglichen Eigenschaften einbezieht, sondern sich auf einige Varietäten und einige der sie kennzeichnenden Eigenschaften beschränkt. Umgekehrt ist die Entwicklung der theoretischen Linguistik in diesem Jahrhundert durchweg durch sehr starke Homogenitätsannahmen gekennzeichnet. Leitende Begriffe wie Saussures „langue" oder Chomskys „competence" abstrahieren von der sprachlichen Variation. Dies führt entweder, wie bei Saussure selbst und bei vielen Strukturalisten, zu einem kaum vermittelten Nebeneinander von Sprachtheorie und empirischer Sprachforschung, oder dazu. Probleme der Variation für uninteressant und belanglos zu erklären, wie in der orthodoxen Chomsky-Schule (Chomsky 19X5, Abschnitt 2.2). In beiden Fällen geht der Zugewinn an Genauigkeit, Explizitheit und an theoretischen Einsichten, den wir der theoretischen Linguistik verdanken, der empirischen Sprachforschung verloren.

In jüngerer Zeit wurden verschiedene Vorschläge gemacht, um diese Kluft zu überbrücken (Überblicke finden sich in Klein 1976, Nabrings 1981). Einer davon ist die im folgenden beschriebene Varietätengrammatik. Sie ist vergleichsweise leicht empirisch anwendbar und erlaubt es, sprachliche Variation, gleichwie bedingt, für den Bereich der Grammatik (Phonologie. Morphologie, Syntax) mit praktisch beliebiger Genauigkeit und Explizitheit zu beschreiben, sofern die entsprechenden Daten zur Verfügung stehen. Wir werden uns in der Darstellung auf die Syntax beschränken, weil dafür die meisten Anwendungen vorliegen. Es ist aber leicht zu sehen, daß das Verfahren auf alle Teile der Grammatik anwendbar ist. für die sich präzise Regeln angeben lassen.

Die moderne Linguistik hat uns zur Beschreibung der Syntax eine ganze Reihe von Grammatiktypen beschert, etwa Phrasenstrukturgrammatik (kontextfrei oder kontextsensitiv), Transformationsgrammatik. Dependenzgrammatik. Kategorialgrammatik. Unabhängig von ihren jeweiligen Vor- und Nachteilen, die wir hier nicht abwägen können, ist diesen Grammatiken zweierlei gemeinsam: es handelt sich um präzise definierte, formale (oder zumindest formalisierbare) Grammatiken mit expliziten Regeln, und sie wurden, wie schon erwähnt, allesamt für homogene Sprachen entwickelt. Die Varietätengrammatik versucht nun, diese Grammatiken unter Wahrung von Genauigkeit und Explizitheit für die Aufgaben (1) — (2), wie sie oben erläutert wurden, nutzbar zu machen. Die Grundidee ist es dabei, eine solche Grammatik — z. B. eine Transformationsgrammatik — als feste, gleichbleibende Bezugsgröße zu nehmen und die Anwendung ihrer Regeln dann so einzuschränken, daß jeweils eine bestimmte Varietät beschrieben wird; dies geschieht dadurch, daß den einzelnen Regeln auf noch zu erläuternde Weise Wahrscheinlichkeiten für ihre Anwendung zugeordnet werden. Alle Varietäten, die man beschreiben möchte, haben also dieselbe Grammatik und dieselben Regeln; sie unterscheiden sich lediglich durch die Wahrscheinlichkeit, mit der diese Regeln angewandt werden; im Grenzfall ist diese Wahrscheinlichkeit Null, d. h. diese Regel existiert praktisch für die betreffende Varietät nicht. Dies erlaubt es, auch graduelle Unterschiede beliebiger Feinheit zwischen einzelnen Varietäten genau zu erfassen. Bildlich kann man sich die Bezugsgrammatik als auf ein Gummituch geschrieben vorstellen, das sich beliebig dehnen und damit den einzelnen abzudeckenden Varietäten genau anpassen läßt. In Abschnitt 2 wird dieser Gedanke erläutert und präzisiert. In Abschnitt 3 — 5 wird gezeigt, wie man verschiedene Grammatiktypen probabilistisch gewichten kann. Diese Teile sind relativ allgemein und theoretisch. Die Varietätengrammatik ist jedoch in erster Linie für praktische Zwecke, zur empirischen Analyse und Beschreibung der Variation, entwickelt worden. In Abschnitt 6 wird daher noch kurz auf ihre praktische Brauchbarkeit eingegangen.

2. Varietätenraum, Bezugsgrammatik und probabilistische Gewichtung

Ein Varietätenraum ist eine geordnete Menge von Varietäten, die man untersuchen möchte. Nehmen wir an, jemand möchte die Syntax

— von Liebesbriefen gegenüber Beileidsbriefen (Faktor Register mit den Ausprägungen r_1 und r_2)
— um 1850, 1900 und 1950 (Faktor Zeit mit den Ausprägungen t_1, t_2 und t_3)
— in gehobener Bürgerschaft und bei Arbeitern (Faktor soziale Schicht mit den Ausprägungen s_1 und s_2)

untersuchen. Dies ergibt einen dreidimensionalen Varietätenraum mit 2 x 3 x 2 = 12 einzelnen Varietäten; so ist (r_1, t_3, s_2) die Varietät der Liebesbriefe von Arbeitern um 1950. Ob ein solcher Varietätenraum wissenschaftlich sinnvoll ist, ist eine empirische Frage; man läßt sich, wenn man einen Varietätenraum für irgendeinen Zweck aufstellt, natürlich von allerlei Vorüberlegungen leiten; aber wie bei allen heuristischen Hypothesen kann man mehr oder minder viel Glück damit haben.

Eine Bezugsgrammatik ist eine (formale) Grammatik gleich welchen Typs, die alle in den einzelnen Varietäten vorkommenden syntaktischen Erscheinungen erfaßt. Sie trennt also noch nicht zwischen den einzelnen Varietäten, sondern muß erst durch Beschränkungen in der Regelanwendung auf die jeweilige Varietät zugeschnitten werden. Dies leisten die probabilistischen Gewichtungen. Am leichtesten läßt sich die Idee der probabilistischen Gewichtung an (kontextfreien) Phrasenstrukturgrammatiken erläutern. Bei solchen Grammatiken können bestimmte Regeln alternativ angewandt werden, und zwar solche, die auf der linken Seite dasselbe Symbol haben. So kann eine kontextfreie Grammatik beispielsweise die folgenden Regeln für NP haben:

NP —> N
NP —> Det N
NP —> Det Adj N

d. h. NP kann zu einem einfachen Nomen werden *{Vater)* oder zu einem Nomen mit Determinativpartikel *{der Vater, ein Vater)* oder zu einem Nomen mit Determinativpartikel und Adjektiv *(der müde Vater);* in Wirklichkeit gibt es natürlich sehr viel mehr Möglichkeiten. Alle alternativen Regeln dieser Art kann man zu einem „Regelblock" zusammenfassen; zur Notation benutzt man gewöhnlich geschweifte Klammern, etwa

$$\mathrm{NP} \rightarrow \left\{ \begin{array}{lll} & \mathrm{N} & \\ \mathrm{Det} & \mathrm{N} & \\ \mathrm{Det} & \mathrm{Adj} & \mathrm{N} \end{array} \right\}$$

Man beachte, daß dies nicht eine Regel, sondern eine abgekürzte Schreibweise für drei Regeln ist. Regelblöcke können zwei, drei, ... beliebig (aber endlich) viele Regeln umfassen; es ist zweckmäßig, auch den Grenzfall von Regelblöcken mit nur einer Regel zuzulassen, solche also, die gar keine Alternative zulassen (dies ist nur eine terminologische Konvention).

Wenn nun in einer Ableitung eines Satzes das links stehende Symbol eines solchen Regelblocks auftaucht, dann *muß* eine der Regeln des Regelblocks angewandt werden (im Grenzfall halt die einzige vorhandene). Die Anwendung der Regeln ist aber nicht (oder selten) gleich wahrscheinlich; es gibt Regeln, die in einer Varietät sehr wahrscheinlich sind, während andere sehr selten auftauchen, und dies kann von Varietät zu Varietät sehr verschieden sein. Um diesem Umstand Rechnung zu tragen, ordnet man den einzelnen Regeln eines Blockes Wahrscheinlichkeiten ihrer Anwendung zu. Wahrscheinlichkeiten werden gewöhnlich durch (reelle) Zahlen zwischen 0 und 1 ausgedrückt; dabei steht 1 für „tritt immer ein" und 0 für „tritt nie ein"; die Werte dazwischen bezeichnen alle Zwischenstufen zwischen diesen beiden Extremen. Man ordnet nun dem Block insgesamt die Wahrscheinlichkeit 1 zu, denn irgendeine seiner Regeln muß ja angewandt werden, und verteilt diese Gesamtwahrscheinlichkeit auf die einzelnen Regeln. So könnte in unserem Beispiel die Regel NP —> N die Wahrscheinlichkeit 0.3 haben, NP —> Det N die Wahrscheinlichkeit 0.5 und NP —> Det Adj N die Wahrscheinlichkeit 0.2; wie dies in der Tat ist. muß natürlich empirisch ermittelt werden. Es kann, wie schon gesagt, für die einzelnen Varietäten sehr verschieden aussehen; in einer anderen Varietät könnte z. B. NP —> N den Wert 0.7, NP —> Det N den Wert 0.2 und NP —> Det Adj N den Wert 0.1 haben, d. h. die NP's sind hier insgesamt gesehen einfacher strukturiert: sie bestehen meist nur aus einem einfachen Nomen. Es kann sogar sein, daß eine bestimmte Regel in einer Varietät überhaupt den Wert 0 hat (d. h. sie tritt nie auf), während sie in anderen und ansonsten ähnlichen Varietäten eine große Rolle spielt.

Wir können nun die Anwendung der Varietätengrammatik an einem Beispiel erläutern. Dazu betrachten wir einen sehr einfachen, eindimensionalen Varietätenraum, nämlich sechs Lernstadien V_t— V_6 im Zweitspracherwerb eines Lerners: Bezugsgrammatik ist ein Ausschnitt aus einer kontextfreien Grammatik, nämlich jene Regeln, die die Entwicklung der NP beschreiben. Dazu nehmen wir zu den obigen drei Regeln noch zwei weitere hinzu: NP —> Det N Adj und NP —> Det N Adv; die erste erzeugt Strukturen wie *der Mann alte,* wie sie in einer Lernervarietät durchaus auftauchen können: die zweite erzeugt Strukturen wie *der Mann dort.* Der gesamte Regelblock für NP besteht also aus fünf alternativen Regeln. Eine Varietätengrammatik — bzw. ein Ausschnitt daraus, denn es geht hier nur um nominale Einheiten — könnte dann im Ergebnis so aussehen:

	V_1	V_2	V_3	V_4	V_5	V_6
NP —> N	0.9	0.6	0.3	0.2	0.2	0.2
NP —> Det N	0.1	0.3	0.3	0.3	0.3	0.3
NP —> Det Adj N	0	0	0	0	0.4	0.4
NP —> Det N Adj	0	0.1	0.3	0.4	0	0
NP —> Det N Adv	0	0	0.1	0.1	0.1	0.1

Dies ist eine genaue Darstellung einer ganzen Reihe von Entwicklungen. Zunächst (d. h. in V_1) tauchen nur ganz einfache NP's auf: meist — genau gesagt, mit der Wahrscheinlichkeit *0.9,* d. h. in 90% aller Fälle - reines Nomen, daneben auch noch Nomina, die durch eine Determinativpartikel erweitert sind; komplexere Nominalkonstruktionen kommen nicht vor. In V_2 verschiebt sich das Gewicht nun zugunsten komplexerer NP's; es treten die ersten mit Adjektiv auf; allerdings steht das Adjektiv an „falscher Stelle". Diese Entwicklung hält dann in V, an: die einfachen NP's werden noch seltener, die (falschen) Attributivkonstruktionen häufiger; zudem taucht eine neue, diesmal gleich richtige Regel auf. die allerdings keine allzu große Rolle spielt. In V_4 hat sich nur sehr wenig verschoben, d. h. es ist keine nennenswerte Entwicklung eingetreten. In V_5 hat sich auch nur eine einzige Veränderung ergeben, allerdings eine gravierende: der Lerner hat die „richtige" Attributivregel kapiert, er plaziert das Adjektiv nun vor das Nomen statt dahinter; der Übergang ist hier abrupt. Es könnte natürlich auch sein, daß ein solcher Übergang allmählich ist. d. h. daß zunächst einige wenige richtige Konstruktionen auftauchen, und sich die höhere Wahrscheinlichkeit erst langsam von NP —> Det N Adj auf NP —> Det Adj N verlagert: solche Übergänge können durch die Varietätengrammatik mit beliebiger Feinheit erfaßt werden. Betrachten wir nun noch V_6: hier hat sich gegenüber V, nichts mehr verändert, d. h. die Endvarietät scheint erreicht zu sein, jedenfalls soweit man dies aus dieser Untersuchung ermitteln kann; V_6 (bzw. V_5) mag dabei von der Zielvarietät durchaus noch um einiges entfernt sein: dort sind die Werte vielleicht (von oben) 0.1, 0.4, 0.4, 0, 0.2: sie werden jedoch vom Lerner nicht erreicht.

Im folgenden wollen wir. was hier am Beispiel illustriert wurde, für einige Grammatiktypen präzisieren.

3. Probabilistische Gewichtungen I: Kontextfreie Grammatiken

3.1. Wahrscheinlichkeiten

Wir können hier natürlich keine noch so elementare Einführung in die Wahrscheinlichkeitstheorie geben, sondern deuten nur einige Grundbegriffe an (gute Einführungen sind Feller 1968 und Stegmüller 1973). Wahrscheinlichkeiten deutet man als die Ausgänge eines im Prinzip beliebig wiederholbaren Geschehens. So hat der Wurf eines Würfels sechs mögliche Ausgänge, die alternativ sind und von denen einer eintreten *muß:* es fällt die Eins oder die Zwei ... oder die Sechs. Gewöhnlich ist man nicht nur an der Wahrscheinlichkeit dieser Elementarereignisse interessiert, sondern auch an der zusammengesetzter, z. B. an der, eine Eins *oder* eine Zwei zu erhalten oder eine gerade Augenzahl. Die Menge der Elementarereignisse bezeichnet man gewöhnlich als Stichprobenraum O; beim Würfelspiel ist O = $\{o_1, o_2, o_3, o_4, o_5, o_6\}$, wobei o_1 das Ereignis ist, daß die Eins fällt, usw. Die Gesamtmenge aller Ereignisse nennt man den Ereignisraum F. Man kann F — in den hier interessierenden Fällen — als die Potenzmenge von O auffassen; dabei ist beispielsweise $\{o_2, o_4, o_6\}$ das Ereignis, daß eine gerade Zahl fällt.

Die Wahrscheinlichkeit eines Ereignisses drückt man durch eine reelle Zahl p zwischen 0 und 1 aus, wobei 1 das „sichere Ereignis" ist. Diesen Wert hat auf jeden Fall das Ereignis $\{o_1, o_2, o_3, o_4, o_5, o_6\}$, denn dies besagt ja „Eins oder Zwei oder ... oder Sechs". Man ordnet also dem Stichprobenraum die Gesamtwahrscheinlichkeit 1 zu. Sie kann sich nun in unterschiedlicher Weise auf die Einzelereignisse o_1, o_2 usw. verteilen. Wenn der Würfel in Ordnung ist, gibt es keinen Grund für die Annahme, daß eine Zahl wahrscheinlicher kommt als eine andere. Man erhält einfach aufgrund von a-priori-Überlegungen die folgende Wahrscheinlichkeitsverteilung für O:

Ereignis o_i	o_1	o_2	o_3	o_4	o_5	o_6	O
Wahrscheinlichkeit p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\sum_{i=1}^{6} p_i = 1$

Was ist nun die Wahrscheinlichkeit eines komplexen Ereignisses $\{o_2, o_4, o_6\}$, d. h. die, eine gerade Zahl zu würfeln? Offenbar dieselbe wie die. eine ungerade Zahl zu würfeln,

d. h. ½, und dies ergibt sich einfach durch

Addition der Einzelwahrscheinlichkeiten $p(o_2)$ r $p(o_4)$ r $p(o_6)$. Die Wahrscheinlichkeit, eine 5 zu erhalten ist $\frac{1}{6}$, die Wahrscheinlichkeit, *keine* 5 zu erhalten — d. h. die Wahrscheinlichkeit von $\{o_1, o_2, o_3, o_4, o_6\}$ ist offenbar $1 - \frac{1}{6} = \frac{5}{6}$, und dies erhält man andersherum durch Addition der Wahrscheinlichkeiten von o_1, o_2, o_3, o_4 und o_6.

Im Beispiel waren die Elementarereignisse gleich wahrscheinlich. Dies braucht nicht der Fall zu sein. Wenn der Würfel gebleit ist, helfen uns a-priori-Überlegungen wenig zur Bestimmung der einzelnen Wahrscheinlichkeiten. Wir können aber durch wiederholtes Werfen empirisch ermitteln, wie oft die einzelnen Fälle eintreten. Wenn z. B. o_6 bei m-Versuchen n-mal eintritt, bezeichnet man $\frac{\quad}{m}$ als die relative Häufigkeit f von o_6, kurz $f(o_6)$. Bei einer großen Zahl von Versuchen stabilisiert sich dieser Wert, und man kann ihn als die Wahrscheinlichkeit p von o_6 *deuten.* Bei dieser statistischen Interpretation wird die Wahrscheinlichkeit eines Ereignisses also als Grenzwert der relativen Häufigkeit aufgefaßt, die man in irgendeinem — notwendigerweise endlichen — Experiment ermittelt. So könnten sich aufgrund vieler Würfe mit dem gebleiten Würfel die folgenden relativen Häufigkeiten ergeben:

Ereignis o_i	o_1	o_2	o_3	o_4	o_5	o_6	O	
relative Häufigkeit	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{2}$	$\sum_{i=1}^{6} f_i = 1$

Es ist wichtig zu betonen, daß diese Werte keine Wahrscheinlichkeiten sind, sondern als solche interpretiert werden. Diese statistische Auffassung der Wahrscheinlichkeit liegt praktisch allen empirischen Untersuchungen zugrunde, so auch allen Anwendungen der Varietätengrammatik. Mathematisch ist sie übrigens nicht ganz unproblematisch; aber das braucht uns hier nicht weiter zu stören. Bei Varietätengrammatiken haben wir es nun nicht mit dem Fall von Würfeln zu tun, sondern mit der Anwendung von Regeln. Will man dafür Wahrscheinlichkeiten angeben, dann sind zwei Probleme zu klären: (1) Was muß man als Stichprobenraum ansehen? (2) Wie kann man aufgrund empirischer Untersuchungen die Wahrscheinlichkeiten für einen gegebenen Stichprobenraum ermitteln? Auf die zweite Frage gehen wir in Abschnitt 6 ein. Die Antwort auf die erste hängt von der Art der Grammatik ab. Wir betrachten zunächst kontextfreie Grammatiken.

3.2. Gewichtungen vom Typ Suppes

Kontextfreie Phrasenstrukturgrammatiken (KF — PSG) wurden von Chomsky als Formalisierung der traditionellen strukturalistischen Konstituentenstrukturanalyse entwikkelt. Eine KF — PSG beschreibt alle und nur die Sätze einer Sprache, indem sie sie „generiert", d. h. dadurch, daß ein bestimmtes Symbol (das „Axiom") durch eine Kette von Symbolen ersetzt wird, diese wiederum einzeln durch andere Symbolketten, bis der Prozeß sich nicht mehr fortsetzen läßt. Welche Möglichkeiten der Ersetzung es gibt, wird durch eine Menge von Regeln festgelegt; diese Regeln müssen so formuliert werden, daß die aus dem Axiom letztlich ableitbaren Symbolketten genau die Sätze der zu beschreibenden Sprache sind. Zur vollen Bestimmung einer KF — PSG müssen alle vorkommenden Symbole (unter Einschluß des Axioms) und die einzelnen Regeln angegeben werden. Wir können dies alles in der folgenden Definition präziser fassen:

(1) Eine kontextfreie Grammatik G ist ein Quadrupel $\langle V_T, V_N, S, R\rangle$, wobei
- V_T eine endliche Menge von Symbolen, das *Endvokabular* (d. h. praktisch die Wörter)
- V_N eine endliche Menge von Symbolen, das *Hilfsvokabular* (d. h. praktisch die syntaktischen Kategorien)
- S ein ausgezeichnetes Element aus V_N, das *Axiom*
- R eine endliche Menge von Folgen der Form $A \rightarrow x$, wobei $A \in V_N$, $x \in (V_T \cup V_N)^+$, wobei $(V_T \cup V_N)^+ = \{(V_T \cup V_N)^* - \{0\}\}$, die *Regeln* (d. h. A ist eine Kategorie und x eine beliebige, aber nicht leere Folge von Symbolen).

Betrachten wir dazu als Beispiel eine einfache KF — PSG, die korrekte Sätze des Deutschen — natürlich nicht alle, aber immerhin unendlich viele — erzeugt:

(2) V_T = {die, einige, und, oder, lieben, sehen, toben, Kinder, Schweine, Löwen, kleinen, großen, dort, hier}

V_N = {S, NP, VP, KON, DET, N, ADJ, ADV, VI, VT}

R = {r_1, ... r_{22}, wobei
r_1: S — S KON S
r_2: S —> NP VP
r_3: S —> NP VP ADV
r_4: VP —> VT NP
r_5: VP —> VI
r_6: NP —> DET N
r_7: NP —> DET ADJ N
r_8: NP —> DET N ADV
r_9: DET —> die
r_{10}: DET —> einige
r_{11}: KON —> und
r_{12}: KON —> oder
r_{13}: N —> Kinder
r_{14}: N —> Schweine
r_{15}: N —> Löwen
r_{16}: VT —> lieben
r_{17}: VT —> sehen
r_{18}: VI —> toben
r_{19}: ADJ —> großen
r_{20}: ADJ —> kleinen
r_{21}: ADV —> dort
r_{22}: ADV — hier

Die Ableitung eines Satzes sieht dann beispielsweise so aus

(3)	S	(Axiom)
	NP VP ADV	(nach r_3)
	NP VT NP ADV	(nach r_4)
	DET N VT NP ADV	(nach r_6)
	DET N VT DET ADJ N ADV	(nach r_7)
	einige N VT DET ADJ N ADV	(nach r_{10})
	einige Kinder VT DET ADJ N ADV	(nach r_{13})
	einige Kinder lieben DET ADJ N ADV	(nach r_{16})
	einige Kinder lieben die ADJ N ADV	(nach r_9)
	einige Kinder lieben die großen N ADV	(nach r_{19})
	einige Kinder lieben die großen Schweine ADV	(nach r_{14})
	einige Kinder lieben die großen Schweine dort	(nach r_{21}, keine weitere Regel anwendbar)

Bei jedem Übergang von einer Zeile zur nächsten muß eine Regel angewandt werden. Dabei gibt es oft Alternativen: so hätte man zu Beginn statt Regel 3 auch Regel 1 oder Regel 2 anwenden können. Es liegt daher nahe, für die probabilistische Gewichtung alle Regeln mit dem gleichen Symbol auf der linken Seite zu Regelblöcken zusammenzufassen und diese als die Stichprobenräume zu nehmen; im Grenzfall hat ein solcher Regelblock nur eine einzige Regel, die dann — sofern das betreffende Symbol in der Ableitung auftaucht — obligatorisch angewandt werden muß. Unsere Beispielgrammatik hat also unter anderem die folgenden Regelblöcke:

(4)	1. Block	2. Block
	S —> S KON S	VP —> VT NP
	S —> VP NP	VP —> VI
	S —> VP NP ADV	
	3. Block	4. Block usw.
	NP —> DET N	VI —> toben
	NP —> DET ADJ	
	NP —> DET N ADV	

In vielen Darstellungen faßt man solche Regelblöcke mit geschweiften Klammern zu einer Gesamtregel zusammen, z. B.

(5) $$S \rightarrow \left\{ \begin{array}{l} \text{S KON S} \\ \text{VP NP} \\ \text{VP NP ADV} \end{array} \right\}$$

Diesem Gebrauch werden wir hier, um Mißverständnisse zu vermeiden, nicht folgen.

Jedem Block der Grammatik wird also die Gesamtwahrscheinlichkeit 1 zugeordnet, die sich unterschiedlich auf die einzelnen Regeln verteilen kann; wie dies geschieht, ist empirisch zu ermitteln, grob gesagt dadurch, daß man in bestimmten Datenkorpora die relativen Anwendungshäufigkeiten feststellt. Dabei sind zwei Fälle von besonderer Bedeutung. Zum einen kann ein Block aus einer einzigen Regel bestehen; die erhält dann den Wert 1. Zum andern kann eine Regel auch die Wahrscheinlichkeit 0 erhalten, d. h. sie wird nie angewandt — jedenfalls nicht in der betreffenden Varietät, während in einer andern Varietät ihre Anwendungswahrscheinlichkeit sehr wohl von 0 verschieden sein kann.

Diese einleuchtende Art der probabilistischen Gewichtung wurde unabhängig von verschiedenen Mathematikern entwickelt. Zur Beschreibung natürlicher Sprachen wurde sie erstmals von Patrick Suppes benutzt, nach dem sie deshalb hier benannt wird. Wir geben eine genaue Definition.

(6) Eine probabilistische Grammatik G vom Typ Suppes ist ein Quintupel $\langle V_T, V_N, S, R, p\rangle$. wobei $\langle V_T, V_N, S, R\rangle$ eine kontextfreie Grammatik, und p eine reell-wertige Funktion

auf R ist, so daß
(a) für alle $r \in R$, $p(r) \geq 0$
(b) für alle $A_i \in V_N$: wenn es m viele Folgen $x_j \in (V_T \cup V_N)^+$ gibt, so daß $A_i \rightarrow x_j \in R$, dann

$$\sum_{j=1}^{m} p\,(a_i \rightarrow x_j) = 1$$

Die Funktion p weist also jeder Regel einen Wert zu, der größer oder gleich 0 ist; daß er nicht größer als 1 sein kann, ergibt sich aus Bedingung (b), derzufolge die Wahrscheinlichkeiten für Regeln mit dem gleichen Symbol zur Linken sich zu 1 aufaddieren.

Eine probabilistische Grammatik dieses Typs besitzt also eine ganze Anzahl von Stichprobenräumen, deren jeder die Gesamtwahrscheinlichkeit 1 hat. Wie sich diese Gesamtwahrscheinlichkeit über die Regeln eines Blocks verteilt, ist unabhängig davon, wie es in einem andern Block aussieht: die Anwendung von Regel i in Block j hat immer dieselbe Wahrscheinlichkeit, unabhängig davon, ob irgendeine andere Regel außerhalb von Block j angewandt wird. Wie wir gleich sehen werden, wird diese „Unabhängigkeitsannahme" oft den empirischen Fakten nicht gerecht. Zuvor soll aber noch kurz auf zwei allgemeine Einwände gegen probabilistische Grammatiken eingegangen werden.

Der erste bezieht sich auf die Wahrscheinlichkeit von Sätzen bzw. von Ableitungen, die sie erzeugen (die Grammatik erlaubt manchmal, einen Satz auf verschiedene Weise abzuleiten; dies können wir aber für das Argument einmal vernachlässigen). Die Wahrscheinlichkeit einer ganzen Ableitung ergibt sich — jedenfalls unter der Unabhängigkeitsannahme — multiplikativ aus den Wahrscheinlichkeiten der einzelnen zu ihrer Bildung angewandten Regeln. Man kann sich nun leicht ausrechnen, daß bei einer einigermaßen realistischen Grammatik, die viele Regeln umfaßt, die Wahrscheinlichkeit einer ganzen Ableitung — und damit des von ihr erzeugten Satzes — sehr schnell verschwindend gering wird: sie ist für alle Sätze einer natürlichen Sprache nahezu Null. Das ist aber selbstverständlich kein wirklicher Einwand gegen probabilistische Grammatiken, denn dort ist man nicht an der Wahrscheinlichkeit von Sätzen, sondern an der von alternativen Regeln interessiert und daran, wie sie von Varietät zu Varietät schwanken.

Der zweite, etwas kuriose, aber dennoch wiederholt vorgebrachte Einwand ist, daß der Sprecher kein Zählwerk im Kopf hat, das ihm die Wahrscheinlichkeit, mit der er eine Regel anzuwenden hat, zu berechnen oder abzuschätzen erlaubt. Das ist eine eigentümliche Fehldeutung der Idee der Wahrscheinlichkeit; auch der Würfel zählt nicht und fällt mit bestimmten Wahrscheinlichkeiten. Daß eine bestimmte Regel von bestimmten Sprechern mit einer bestimmten Wahrscheinlichkeit angewandt wird, ist einfach eine deskriptive Tatsache — nicht anders als die Tatsache, daß man mit der Wahrscheinlichkeit von (sagen wir) 0.31 jemand Gleichaltriges heiratet oder mit der Wahrscheinlichkeit von 0.92 vor dem siebzigsten Lebensjahr stirbt, und auch hier zählt keiner nach, ob er in der Quote liegt oder ob die Zeit gekommen ist, den Löffel wegzulegen.

Kommen wir nun zur „Unabhängigkeitsannahme" zurück. Bei einer Suppes-Bewertung haben beispielsweise die einzelnen NP-Regeln (innerhalb einer Varietät) immer denselben Wert, unabhängig davon, ob die jeweilige NP in Subjektstellung steht, d. h. durch r_2 oder durch r_3 eingeführt wurde, oder in Objektstellung d. h. durch r_4 eingeführt wurde. Ob dies tatsächlich so ist. ist eine empirische Frage, und fast jede empirische Untersuchung zeigt, daß es *nicht* so ist: in Subjektstellung werden bevorzugt andere Regeln angewandt als in Objektstellung. Man kann diesem Problem dadurch Rechnung tragen, daß man nicht mehr von NP spricht, sondern von NPS und NPO, die Regeln entsprechend umformuliert (r'_1: S —> NPS VP; r'_3: S —> NPS VP ADV; r'_4: VP —> VT NPO) und dann den NP-BLock in zwei einzelne aufspaltet, die sich getrennt bewerten lassen. Dies heißt aber zum einen, daß man linguistisch sinnvolle Regeln zugunsten weniger genereller aufgibt, nur um die richtigen Wahrscheinlichkeitsverhältnisse zu erfassen. Und zum andern ist dieses Verfahren nicht möglich bei rekursiven Regeln, d. h. bei solchen, in denen das linksstehende Symbol rechts erneut auftaucht, wie in r_1. r_1 kann in einer Ableitung immer wieder angewandt werden; es leuchtet aber ein, daß die Wahrscheinlichkeit dafür immer geringer wird, die der Alternativen r_2 und r_3 immer größer. Dem läßt sich nur dadurch Rechnung tragen, daß man berücksichtigt, welche Regel bereits *zuvor* angewandt wurde, und damit wird die Unabhängigkeitsannahme aufgegeben. Es gibt verschiedene Möglichkeiten, solche Unabhängigkeiten bei der probabilistischen Gewichtung zu erfassen. Eine davon, die von dem Mathematiker A. Salomaa vorgeschlagen wurde, betrachten wir im folgenden Abschnitt.

3.3. Gewichtungen vom Typ Salomaa

Wichtig für diese Art der Gewichtung ist der Begriff des „Kontrollwortes" einer Ableitung. Man hat sich dazu vorzustellen, daß alle Regeln der Grammatik eine Bezeichnung tragen, wie etwa r_1, r_2, ..., r_{22} in unserem Beispiel. Das Kontrollwort einer Ableitung ist dann eine Folge von Regelbezeichnungen, und zwar jener Regeln, die bei der betreffenden Ableitung angewandt wurden. Dabei ist zweierlei zu beachten. Für die erste Ableitungszeile gibt es keine Regel; der Einheitlichkeit halber führt man hier eine „leere Regel" mit der Bezeichnung r_0 ein. Weiterhin muß man festlegen, in welcher Reihenfolge die Symbole in einer Ableitungszeile ersetzt werden, weil sonst mehrere Kontrollwörter zum selben Ergebnis führen können; hier soll die Vereinbarung „von links nach rechts" gelten, die wir auch schon bei (3) stillschweigend befolgt haben. Das Kontrollwort der Ableitung (3) ist dann: r_0 r_3 r_4 r_6 r_7 r_{10} r_{13} r_{16} r_9 r_{19} r_{14} r_{2t}. Die Menge aller Kontrollwörter einer gegebenen Grammatik nennt man „Kontrollsprache" und man kann nun probabilistische Beschränkungen (und übrigens auch andere) für die Kontrollsprache, statt für einzelne Regeln, formulieren.

Bei einer Salomaa-Grammatik betrachtet man dazu alle Digramme. d. h. alle Zweierfolgen von Regelbezeichnungen, die in der Kontrollsprache vorkommen, und gibt dann für jede Regel an, wie ihre Anwendungswahrscheinlichkeit ist. wenn ihr eine bestimmte andere Regel vorausgeht. Viele mögliche Digramme sind bereits durch die Art, wie die Regeln formuliert sind, ausgeschlossen; so kann z. B. in unserer obigen Beispielgrammatik r_1 nicht auf r_4 folgen. Das stört aber nicht weiter: das Digramm r_4r_1 erhält einfach den Wert 0. Für die tatsächlich möglichen Übergänge müssen die Wahrscheinlichkeiten wiederum empirisch ermittelt werden.

Es ist leicht zu sehen, inwieweit eine Salomaa-Gewichtung über eine Suppes-Gewichtung hinausgeht: letztere ist einfach ein Sonderfall einer Salomaa-Gewichtung, und zwar jener, bei der p für eine gegebene Regel r_1 immer denselben Wert hat, unabhängig davon, welche Regel r_1 ihr in der Anwendung vorausgeht. Die Salomaa-Grammatik erlaubt es hingegen, hier zu differenzieren — allerdings eben nur im Hinblick auf die unmittelbar vorausgehende Regel. Man kann nun den Gedanken der Salomaa-Grammatik so verallgemeinern, daß beliebig weit zurückliegende Regelanwendungen berücksichtigt werden, und kommt so zu dem Konzept der allgemeinen „Ableitungsbewertenden Grammatik". Die Gewichtungen werden aber dann so aufwendig und unhandlich, daß sie für die empirische Arbeit kaum zu gebrauchen sind. Wir gehen deshalb hier nicht weiter darauf ein. (Eine ausführliche Darstellung mit genauen Definitionen der Salomaa-Gewichtung und der Ableitungsbewertenden Grammatik findet sich in Klein und Dittmar 1979, 48-56.)

4. Probabilistische Gewichtungen II: Kontextsensitive Grammatiken

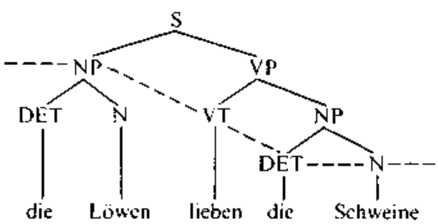
In kontextfreien Grammatiken haben alle Regeln die Form A —> x; in einer Ableitungszeile darf A durch x ersetzt werden, unabhängig davon, was sonst in der Ableitungszeile steht. Bei kontextsensitiven Regeln ist eine Ersetzung nur zulässig, wenn der „Kontext" in der jeweiligen Ableitungszeile paßt, d. h. wenn rechts und links von A bestimmte Symbole stehen. Kontextsensitive Regeln haben daher die Form uAv —> u x v (auch: A—> x/u — v, d. h. wird zu x im Kontext u — v), wobei wie üblich A ein Hilfssymbol und x eine nicht leere Kette von Symbolen ist; u und v sind gleichfalls Symbolketten, die auch leer sein können. Offenbar ist eine kontextfreie Grammatik nur ein Sonderfall einer kontextsensitiven, nämlich jener, bei der u und v immer leer sind.

Kontextsensitive Grammatiken lassen sich ebenso probabilistisch gewichten wie kontextfreie. Der einzige Unterschied liegt darin, daß man nicht alle Regeln mit dem gleichen Symbol zur Linken als Stichprobenraum auffaßt, sondern alle mit gleichem Symbol *im gleichen Kontext.* Im übrigen gilt alles in Abschnitt 3 Ausgeführte ganz analog. Kontextsensitive Varietätengrammatiken haben einige Vorzüge gegenüber kontextfreien, insofern sie gewisse Abhängigkeiten schon bei Suppes-Gewichtung erfassen. So läßt sich das in Abschnitt 3.2 erwähnte Problem des unterschiedlichen Verhaltens von Subjekt-NP und Objekt-NP mit kontextsensitiven Regeln leicht lösen, weil die Objekt-NP nur im Kontext VT _, die Subjekt-NP hingegen nur im Kontext _ VP ersetzt wird; man kann daher leicht in der Zuweisung von Wahrscheinlichkeiten zwischen beiden Fällen unterscheiden. Dies geht aber nur in manchen Fällen von Abhängigkeit; beispielsweise läßt sich das gleichfalls in Abschnitt 3.2 erwähnte Problem der rekursiven Regeln so nicht lösen.

5. Probabilistische Gewichtungen III: Transformationsgrammatiken

Eine Transformationsgrammatik besteht aus (zumindest) zwei Komponenten: einer ersten, die Strukturen erzeugt, und einer zweiten, die nach festen Regeln vorhandene Strukturen umformt. Ersteres kann beispielsweise durch eine (kontextfreie oder kontextsensitive) Phrasenstrukturgrammatik geschehen, die sich wie üblich probabilistisch gewichten läßt. Letzteres erfolgt nach speziellen Transformationsregeln. Eine solche Regel besteht aus zwei Teilen: einer Strukturbeschreibung (SB) und einer Strukturveränderung (SV). Die Strukturbeschreibung gibt an, auf welche Strukturen sich die Regel anwenden läßt. Man kann sie sich im einfachsten Fall als einen „Schnitt" durch einen Strukturbaum vorstellen. So „paßt" der folgende einfache Strukturbaum zu der Strukturbeschreibung SB_1: NP-VT-DET-N, wie die gestrichelte Linie anzeigt, hingegen nicht zu der Strukturbeschreibung SB_2: NP-VT-DET-ADJ-N oder SB_3: NP-VI.

(7)



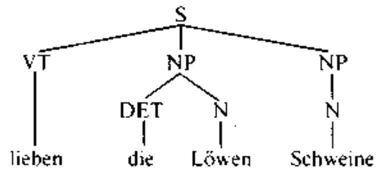
Eine SB kann auch komplizierter sein; z. B. kann sie angeben, daß ein „Teilbaum" mit einem andern identisch sein muß, daß er ein bestimmtes lexikalisches Element zu enthalten hat u. a.

Die SV gibt nun an, was mit den einzelnen in der SB angegebenen Teilbäumen geschieht; sie können z. B. an eine andere Stelle geschoben, durch etwas anderes ersetzt, eventuell auch ganz weggelassen (d. h. durch 0 ersetzt) werden. Natürlich geschieht dies nicht nach Belieben, sondern man hat gewisse Vorstellungen darüber, was linguistisch sinnvoll ist. Hier geht es aber nur um das Prinzip. Eine ganze Transformationsregel, die sich unter anderem auf Struktur (7) anwenden läßt, ist beispielsweise die folgende:

(8) SB: NP - VT - DET - N
SV: 1 2 3 4 —> 2 1 Ø 4

Dies besagt: der erste Teilbaum, also NP, rückt an die Stelle des zweiten, der zweite (VT) an die Stelle des ersten, der dritte (DET) fällt weg, der vierte bleibt unverändert. Auf (7) angewandt, ergibt dies die folgende Struktur:

(9)



Damit haben wird natürlich nur den Grundgedanken der Transformationsregel umrissen; eine präzise formale Definition ist nicht ganz einfach, so daß wir hier darauf verzichten. Wichtig ist noch, daß eine solche Regel optional oder obligatorisch sein kann: im letzteren Fall muß sie immer angewandt werden, wenn SB paßt; dies muß für jede Regel vermerkt werden. Die probabilistische Gewichtung von Transformationsregeln ist im Prinzip sehr einfach: Man faßt all jene Regeln zu einem „T-Block" zusammen, die die gleiche SB haben, und sieht diese T-Blöcke als Stichprobenräume an; die Gesamtwahrscheinlichkeit 1 wird dann über die verschiedenen SV verteilt. Dies führt dann zu einem Problem, wenn es für eine gegebene SB nur eine SV gibt: dann erhielte diese SV automatisch den Wert 1 und wäre somit auf jeden Fall obligatorisch, obwohl sie vielleicht aus linguistischen Gründen als optional anzusehen ist. Um dies zu vermeiden, fügt man jedem Regelblock noch die „identische Transformation" zu, d. h. eine SV. die nichts verändert: ob eine Regel dann obligatorisch ist oder optional, ergibt sich aus der Gewichtung. Eine präzise Definition probabilistischer Transformationsgrammatiken ist, was die Gewichtung selbst angeht, trivial, so daß wir hier darauf verzichten (vgl. dazu Klein/Dittmar 1979. 57-63).

6. Varietätengrammatiken in der empirischen Forschung

Probabilistische Gewichtungen von Grammatiken sind keine Erfindung von Linguisten, sondern von Mathematikern, die an bestimmten formalen Eigenschaften solcher „bewertender" Grammatiken interessiert waren (vgl. dazu und zur bewertenden Grammatik in der Linguistik die ausgezeichnete Darstellung in Habel 1979). Empirisch angewandt wurden sie in Suppes (1972) — allerdings in noch sehr beschränkter Form. Das allgemeine Konzept der Varietätengrammatik wurde, z. T. inspiriert durch Labovs Variablenregel, in Klein (1974) entwickelt. Seither wurde es in verschiedenen Untersuchungen angewandt (Heidelberger Forschungsprojekt 1977; Klein; Dittmar 1979; Senft 1982; Tropf 1983; Carroll 1984).

In der empirischen Anwendung ergeben sich für den Linguisten verschiedene Teilaufgaben, die in einer gewissen Abfolge zu lösen sind:

(1) Es muß festgelegt werden, auf welche (außersprachlich definierte) Varietäten sich die Untersuchung erstrecken soll: Wahl des Varietätenraums (d. h. der unabhängigen Variablen);
(2) Ebenso muß festgelegt werden, welche linguistischen Variablen zu untersuchen sind irgendwelche morphologischen Kategorien, die Entwicklung phonologischer Regeln, die NP-Struktur. usw.: Wahl des Strukturtyps (d. h. der abhängigen Variablen);
(3) Der Linguist muß für seine Varietäten Korpora zusammenstellen, die für diese Varietäten und für die gewählten Äußerungstypen repräsentativ sind: Datengewinnung.
(4) Kr muß die Häufigkeitsverteilung seiner Äußerungstypen in den einzelnen Korpora feststellen: Korpusanalyse.
(5) Kr muß einen geeigneten Grammatiktyp wählen und entsprechend eine Grammatik bzw. einen Teil einer Grammatik schreiben: Ausarbeitung der Bezugsgrammatik.
(6) Kr muß auf die Grundlage der in den Korpora ermittelen Häufigkeiten die Bewertungen vornehmen: Regelbewertung.

Damit ist die Varietätengrammatik für den betreffenden Teilbereich ausgearbeitet. Man kann auf dieser Grundlage dann Untersuchungen über den Zusammenhang zwischen den Varietäten anschließen. So könnte man z. B. feststellen, daß eine Regelveränderung, die sich in unterschiedlichen Bewertungen ausdrückt, ihren Ausgang in einer bestimmten sozialen Schicht nimmt und sich von dort auf die anderen Varietäten ausbreitet usw.

Die Vorzüge der Varietätengrammatik als Beschreibungsinstrument liegen vor allem in dreierlei:

(I) Sie läßt sich auf beliebige Bereiche der Grammatik anwenden, und sie ist nicht an einen bestimmten Grammatiktyp gebunden. Dies läßt dem beschreibenden Linguisten alle Freiheiten.
(2) Sie erlaubt es, den Einfluß beliebig vieler außersprachlicher Faktoren der Variation präzise zu erfassen — sofern diese Faktoren selbst sich präzisieren lassen.
(3) Sie kann auch graduelle Unterschiede zwischen einzelnen Varietäten mit beliebiger Feinheit darstellen.

In diesen drei Punkten wird sie nur durch die verfügbaren Daten beschränkt. Auf der andern Seite ist sie bloß ein Instrument der *Beschreibung:* sie erklärt nichts, sondern sie stellt Zusammenhänge präzise dar und läßt ihre Deutung offen. Nun sind Beschreibungen, so genau sie sein mögen, sicher nicht das Endziel aller linguistischen Forschung; aber besser eine gute Beschreibung als eine schlechte Erklärung.

7. Literatur (in Auswahl)

Caroll, Mary (1984) *Cyclic learning processes in second language production,* Frankfurt a. M.

Chomsky, Noam (1985) *Knowledge of language: Its origins, scope, and use.* Cambridge. Mass.

Feller. William (1968) *An introduction to probability theory and its applications.* New York.

Habel, Christopher (1979) *Aspekte bewertender Grammatiken.* Berlin.

Heidelberger Forschungsprojekt „Pidgin Deutsch" (1977) "Die ungesteuerte Erlernung des Deutschen durch spanische und italienische Arbeiter. Eine soziolinguistische Untersuchung", in: *Osnabrücker Beiträge zur Sprachtheorie.* Beihefte 2, Osnabrück.

Klein. Wolfgang (1974) *Variation in der Sprache,* Kronberg/Ts.

Klein. Wolfgang (1976) „Sprachliche Variation", in: *Studium Linguistik* 1, 19 — 46.

Klein, Wolfgang/Dittmar, Norbert (1979): *Developing grammars.* Berlin Heidelberg New York.

Nabrings. Kirsten (1981) *Sprachliche Varietäten,* Tübingen.

Senft, Gunter (1982) *Sprachliche Varietät und Variation im Sprachverhalten Kaiserslauterer Metallarbeiter.* Bern Frankfurt a.M.

Stegmüller. Wolfgang (1973) *Personelle und statistische Wahrscheinlichkeit.* Berlin Heidelberg/New York.

Suppes, Patrick (1972) „Probabilistic Grammar for Natural Languages", in: *Semantics of natural language.* Davidson. D./Harman. G.. eds.. Dordrecht.

Tropf. Herbert (1983) *Variation in der Phonologie des ungesteuerten Zweitspracherwerbs,* Phil. Diss.. Heidelberg.

Wolfgang Klein, Nijmegen (Niederlande)