

Published in the omnibus volume: *Δωρημα Ευχαριστημιου* for Prof. Dr. P. A. Gemtou (2012), N. Sakkoula, Athens.

## **Eine strukturalistische Rekonstruktion einer linguistischen Theorie**

Wolfgang Balzer<sup>1</sup> (Ludwig-Maximilians-Universität, München)

und

Adriana Gonzalo (Universidad Nacional del Litoral, Santa Fé)

Chomsky hat verschiedene linguistische Theorienansätze vorgelegt, die in einem gewissen Sinn aufeinander aufbauen. Wir haben hier die erste, von Chomsky im Jahr 1957 veröffentlichte Theorie rekonstruiert.<sup>2</sup> Diese Theorie ist einerseits die Grundlage für weitere, spätere Theorien, andererseits wurde sie in der linguistischen Gemeinschaft sehr schnell rezipiert.<sup>3</sup> Wir kürzen im Folgenden das Werk *Syntactic Structures* (Chomsky, 1957) durch *SynStr* ab.

Unsere Arbeit gliedert sich in vier Teile. Im ersten Teil befassen wir uns kurz mit einigen allgemeineren Begriffen und Hypothesen, die für unsere Rekonstruktion verwendet werden und die sich relativ zur hier erörterten Theorie neutral verhalten. Im zweiten Teil beschreiben und erklären wir diejenigen Grundbegriffe und Hypothesen, die für die Rekonstruktion speziell von *SynStr* benötigt werden. Wir formulieren im dritten Teil die Theorie in wissenschaftstheoretischer, strukturalistischer<sup>4</sup> Notation und im letzten Teil diskutieren wir einige metatheoretische Punkte und Fragen der hier erörterten Theorie.

### **1) Linguistische Hintergrundbegriffe**

Die Linguistik untersucht natürliche Sprachen. Eine natürliche Sprache besteht aus Äußerungen, mit denen die Personen einer Sprachgemeinschaft kommunizieren, und aus einem Produktionssystem, das diese Äußerungen systematisch erzeugen kann. Eine Äußerung kann mehr oder weniger komplex sein und sie wird durch eine Person hervorgebracht.<sup>5</sup> In einer natürlichen Sprache läßt sich ein Produktionssystem

---

<sup>1</sup>Einige hilfreiche Anregungen bei Diskussionen mit Stefanie Seith waren nützlich.

<sup>2</sup>Drei Ansätze von Rekonstruktionen im Umfeld von Chomsky finden sich in (Gonzalo, 2001), (Peris Viñé, 1990) und (Quesada, 1993).

<sup>3</sup>Zum Beispiel (Newmeyer, 1996).

<sup>4</sup>Der hier benutzte Strukturalismus ist natürlich nicht zu verwechseln mit dem strukturellen oder strukturalistischen Ansatz in der Linguistik. Die strukturalistische Wissenschaftstheorie ist beschrieben in (Balzer et al., 1987), (Balzer und Moulines, 1996), (Balzer et al., 2000), (Diederich et al., 1989), (Diederich et al., 1994).

<sup>5</sup>Den Begriff der virtuellen Sprachen können wir hier nicht diskutieren.

als eine Menge von natürlichen Prozessen ansehen, die von verschiedenen Menschen generiert werden. In anderen Bereichen wird ein Produktionssystem in sehr idealisierter Weise als eine Maschine verstanden, die zum Beispiel aus einem Computer und der dazugehörigen Peripherie besteht. Der Begriff eines Produktionssystems ist alt; in der Logik wurde er in (Thue, 1914) beschrieben und analysiert.

Ein Produktionssystem besteht aus einem Alphabet  $A$  von *Symbolen*, aus einer Menge  $\Sigma$  von *Startsymbolen*, die aus dem Alphabet stammen, und aus einer Liste  $R$  von *Produktionsregeln*. Eine Produktionsregel  $r$  hat die Form  $(v, h)$ .  $r$  ist ein Paar von Tupeln (oder Listen), so dass beide Listen  $v$  und  $h$  aus Symbolen des Alphabets bestehen und die Liste  $v$  der Regel in einer Produktion durch die Liste  $h$  ersetzt wird. In einer Produktionsregel  $r$  der Form  $(v, h)$ , bezeichnen wir die Liste  $v$  als *Vorderglied* und die Liste  $h$  als *Hinterglied* von  $r$ . Eine solche Produktionsregel  $r$  wird wie folgt angewandt. Wenn eine Liste  $(a_1, \dots, a_s, x_1, \dots, x_m, b_1, \dots, b_q)$  von Symbolen gegeben ist, und  $(x_1, \dots, x_m)$  die Form des Vordergliedes  $v$  der Regel  $r$  hat und  $(y_1, \dots, y_n)$  das Hinterglied  $h$  von  $r$  ist, wird die Liste  $(a_1, \dots, a_s, x_1, \dots, x_m, b_1, \dots, b_q)$  durch die Liste  $(a_1, \dots, a_s, y_1, \dots, y_n, b_1, \dots, b_q)$  ersetzt. Mit anderen Worten wird in der gegebenen Liste ein Teil, der zugleich Vorderglied der Regel  $r$  ist, durch das Hinterglied der Regel ersetzt, d.h.

die Liste  $(a_1, \dots, a_s, x_1, \dots, x_m, b_1, \dots, b_q)$   
wird mit der Regel  $((x_1, \dots, x_m), (y_1, \dots, y_n))$   
umgewandelt in  $(a_1, \dots, a_s, y_1, \dots, y_n, b_1, \dots, b_q)$ .

Um den Prozess zu starten, wird eine Liste der Form  $(\sigma)$  erzeugt, wobei  $\sigma$  ein Startsymbol des Produktionssystems ist ( $\sigma \in \Sigma$ ).

Mit anderen Worten wird ein Produktionsprozess induktiv erzeugt. Die Induktionsbasis besteht aus den Listen  $(\sigma)$ , mit  $\sigma \in \Sigma$ , und aus dem Induktionsschritt, mit dem eine vorher gewonnene Liste  $(a_1, \dots, a_s, x_1, \dots, x_m, b_1, \dots, b_q)$  durch eine Produktionsregel eine neue Liste  $(a_1, \dots, a_s, y_1, \dots, y_n, b_1, \dots, b_q)$  erzeugt. Ein solcher induktiver Prozess kann entweder zu einem Ende kommen, oder ständig weiter laufen. Im ersten Fall paßt die letzte, erzeugte Liste zu keinem Vorderglied einer Regel. In diesem Fall gibt es keine Produktionsregel, die auf die letzte Liste angewandt werden kann. Im zweiten Fall gibt es immer eine Produktionsregel, die auf die „letzte“, gerade erzeugte Liste paßt. Es wird also unterschieden zwischen *beendeten* und *offenen* Prozessen. Die Menge aller beendeten Prozesse, die in einem bestimmten Produktionssystem erzeugt werden können, nennt man in der Linguistik *Grammatik*.<sup>6</sup> Ein beendeter Produktionsprozess besteht aus endlich vielen Listen der Art  $(x_1^1, \dots, x_{n_1}^1), \dots, (x_1^m, \dots, x_{n_m}^m)$ . Die in einem solchen Prozess  $p$  eindeutig bestimmte Liste  $(x_1^m, \dots, x_{n_m}^m)$  nennen wir *die letzte Liste des Prozesses  $p$*  und kürzen sie durch  $end(p)$  ab.

In der wirklichen Welt gibt es Sprachen, deren Äußerungen nicht aufgeschrieben oder in anderer, materieller Form haltbar gemacht werden. Und es gibt Sprachen, bei denen ein kleiner Teil der Äußerungen aufgeschrieben wird. Dies führt zu dem alten

---

<sup>6</sup> Siehe (Aristoteles, 1949).

Unterschied zwischen *Satz* und *Äußerung* eines Satzes. In der Linguistik wird die Dimension der Sätze (*Syntax*) und die Dimension der Lautäußerungen (*Phonologie*) unterschieden. Die Sätze bestehen aus elementaren Bestandteilen, nämlich aus Wörtern, und die Lautäußerungen aus Lauten. Zusätzlich wird in der Linguistik eine theoretische Ebene benutzt, auf der Sätze, Wörter, Lautäußerungen und Laute systematisch untersucht werden, und zwar sowohl in speziellen Sprachen als auch in allen Sprachen. Im Prinzip, läßt sich wahrscheinlich die Satzdimension auf die Äußerungsdimension „reduzieren“, aber wie dies genau geschehen soll, ist unter den Gelehrten umstritten.

Die beiden Dimensionen hängen auf jeden Fall stark zusammen. In der Satzebene läßt sich einerseits jeder Satz analysieren, andererseits läßt er sich aus Wörtern erzeugen. Diese beiden Richtungen der Analyse und der Erzeugung stehen in einem Reflexionsverhältnis und jeder Satz läßt sich mit einer Liste von Wörtern identifizieren. Die Frage, welche dieser beiden Entitäten (Satz oder Wort) primär ist, hat dieselbe Antwort, wie die Frage ob das Huhn oder das Ei wichtiger ist. In der Dimension der Laute finden wir dieselbe Situation. Jede Lautäußerung kann in elementare Laute zerlegt werden, die in einem international festgelegten Lautalphabet dokumentiert sind. In der anderen Richtung werden durch elementare Laute Lautäußerungen erzeugt. Jede Lautäußerung läßt sich also mit einer Liste von elementaren Lauten identifizieren. In beiden Dimensionen wird in der Linguistik die Konkatenationsrelation verwendet, mit der zwei Wörter oder zwei Laute zu einer neuen Einheit zusammengefügt werden. Diese Relation läßt sich rekursiv anwenden. Eine komplexe, schon erzeugte Einheit von Wörtern  $x$  können wir zu einem noch größeren Komplex zusammenfügen, indem wir ein weiteres Wort  $w$  hinzufügen:  $x \circ w$ . Genauso wird in der phonetischen Ebene verfahren. Wir sehen, dass die Wörter, Sätze und die entsprechende Konkatenationsrelation homomorph auf Wortäußerungen, Satzäußerungen und die entsprechende Konkatenationsrelation abgebildet werden können. In der Computersprache können wir die beiden Richtungen -Analyse und Erzeugung- auch als Ströme (*streams*) auffassen.

Sowohl in der Dimension der Laute als auch in der Wortdimension werden in der Linguistik Alphabete benutzt. Eine Äußerung setzt sich aus den Zeichen des Lautalphabets und ein Wort aus dem Schriftalphabet der jeweiligen Sprache zusammen. Genauso läßt sich aus den Wörtern ein Satz zusammenfügen. Hier wird allerdings die Menge der Wörter nicht als ein Alphabet bezeichnet. Die Schriftalphabete in verschiedenen Sprachen sind oft sehr weit von der Lautschrift entfernt. Die Schriftzeichen in der chinesischen Sprache<sup>7</sup> sind völlig anders aufgebaut, als zum Beispiel die in den indogermanischen Sprachen. Der Term „Alphabet“ hat in der Linguistik daher einen etwas einseitigen, indogermanischen Unterton. Auch in den indogermanischen Sprachen sind die Schriftalphabete für die Lautalphabete als Repräsentation wenig effizient. Zum Beispiel wird im Französischen der Laut „o“ durch „eau“ geschrieben, oder im Amerikanischen läßt sich das Wort „world“ in der Lautschrift nur annähernd

---

<sup>7</sup>Die Sprache des Han-Chinesisch wird wahrscheinlich im Moment am häufigsten auf der Erde gesprochen.

beschreiben.

Mit diesen sehr einfachen Strukturen können wir wissenschaftstheoretisch eine Sprache wie folgt beschreiben. Eine Sprache besteht aus<sup>8</sup> der Menge  $U$  der tatsächlich produzierten Äußerungen, dem Produktionssystem  $(A, \Sigma, R)$ , aus der durch  $(A, \Sigma, R)$  definierten Menge  $G$  (der Grammatik von  $(A, \Sigma, R)$ ) und aus der dazugehörigen empirischen Behauptung. Die empirische Behauptung besagt, dass jede tatsächliche Äußerung  $u$  ( $u \in U$ ) auch in der theoretisch definierten Grammatik  $G$  liegt ( $u \in G$ ). Diese einfachste Art von wissenschaftstheoretischer Behauptung sagt mit anderen Worten, dass die Menge der tatsächlichen Äußerungen zu der theoretisch definierten Grammatik *paßt*. Natürlich wird es in jeder natürlichen Sprache einige „Ausrutscher“ geben, so dass einige tatsächliche Äußerungen nicht unter die Grammatik subsumierbar sind und so dass viele grammatisch mögliche Äußerungen nicht produziert werden und wurden.

## 2) Die Grundbegriffe und Hypothesen der Theorie

In *SynStr* werden (mindestens) drei Ebenen unterschieden: die phonetische und die morphologische Ebene und die Ebene der Phrasenstruktur.<sup>9</sup> In jeder Ebene werden die Entitäten durch die entsprechenden Konkatenationsrelationen in elementare Einheiten zerlegt: die Elementarlaute werden zu Äußerungen, die Wörter zu Sätzen und die Phrasen zu konkatenierten Listen von Phrasen.

Wir definieren im allgemeinen den mengentheoretischen Begriff des *n-Tupels* wie folgt. Sei  $n$  eine natürliche Zahl größer Null und eine Menge  $Q$  gegeben. Wir definieren die Menge  $Q^n$  der *n-Tupel* durch  $\{y/\exists x_1 \in Q \dots \exists x_n \in Q(y = (x_1, \dots, x_n))\}$ . Weiter definieren wir die *Menge aller Listen*  $Q^*$  über einer gegebenen Menge  $Q$  wie folgt.  $Q^* = \{\diamond\} \cup \cup_{n>0} Q^n$ , wobei  $\diamond$  die *leere Liste* bezeichnet.

Für die Ebene der Lautäußerungen wird ein internationales Alphabet  $A_{ph}$  benutzt. Über dem Alphabet  $A_{ph}$  ist die Menge  $A_{ph}^*$  aller Listen definiert. Eine solche Liste ist also eine Liste von Zeichen des Lautalphabets. Eine Teilmenge  $U$  von  $A_{ph}^*$  ( $U \subseteq A_{ph}^*$ ) beschreibt die *Äußerungen*, die den Wörtern der Sprache entsprechen. Wir nennen diese auch *Wortäußerungen*. Die Menge  $U$  der Wortäußerungen läßt sich natürlich nicht definieren; sie ist eine Grundmenge der linguistischen Theorie. In *SynStr* bleiben Äußerungen oft unanalysiert, d.h. in bestimmten theoretischen Kontexten können Wortäußerungen als Grundobjekte aufgefaßt werden. In *SynStr* werden Zeichen für Wortäußerungen mit + konkateniert und oft zusätzlich mit speziellen Klammern „/“ umgeben, z.B. umschreibt */teyk/* das normal geschriebene, englische Wort *take*.

In der Ebene der Wörter und Sätze wird ein Alphabet  $A_{bu}$  eingeführt.<sup>10</sup> Ein Wort

<sup>8</sup>Die Gruppe der Menschen bleibt wie üblich in der Linguistik implizit.

<sup>9</sup>Der Term *Phrase* hat leider in der deutschen Sprache eine negative Nebenbedeutung. In *SynStr* wird der Term *phrase* nicht verwendet, er wird nur in dem Term *phrase structure* benutzt. Wir könnten ohne Weiteres hier die *Phrasen* als Äquivalenzklassen von Wortlisten behandeln.

<sup>10</sup>Interessanterweise leitet sich der Term „Buchstabe“ in der deutschen Sprache von dem Baum

besteht aus Buchstaben und ein Satz aus Wörtern. Während das Zusammenschreiben von Buchstaben in *SynStr* normalerweise implizit bleibt, wird die Konkatenation von Wörtern mit demselben Zeichen + geschrieben, die auch für Äußerungen verwendet wird. In der deutschen Sprache ließe sich zum Beispiel das Wort *Buch* weiter in Buchstaben zerlegen:  $B + u + c + h$ . Dies wird aber in *SynStr* nicht weiter erörtert. Dagegen werden dort Wörter in speziellen Kontexten mit speziellen Klammern umgeben, wie etwa # *gegeben* #.<sup>11</sup> In *SynStr* brauchen wir sowohl das „normale“ Alphabet  $A_{bu}$  als auch die Menge  $L$  der Wörter. Wir definieren über dem Alphabet  $A_{bu}$  die Menge  $A_{bu}^*$  aller Listen von Buchstaben. Aus dieser Menge  $A_{bu}^*$  wird die Teilmenge  $L$  nicht definiert, sondern als Grundmenge eingeführt ( $L \subseteq A_{bu}^*$ ). In *SynStr* werden die Wörter, Phrasen und Wortäußerungen mit demselben Zeichen + konkateniert, etwa *das + Buch + ist + Adjektiv* oder *John # S + n't # arrive*. Zusammenfassend sind Wörter und Sätze -mengentheoretisch betrachtet- Listen von Buchstaben und Listen von Listen von Buchstaben. In *SynStr* wird -vermutlich aus Lesbarkeitsgründen- diese Analyse abgekürzt.

Auf der dritten Ebene werden in *SynStr* verallgemeinerte Terme benutzt, die wir hier als *Phrasenarten* bezeichnen. Eine Phrasenart ist mengentheoretisch ausgedrückt eine Menge von endlichen Listen von Wörtern. Damit gelangen wir zu der Menge aller Listen  $L^*$  über der Menge  $L$  von Wörtern und zu der Menge der Mengen aller Listen  $\wp(L^*)$ .<sup>12</sup> Jede Phrasenart ist also eine Menge, wobei natürlich nicht jedes Element aus  $\wp(L^*)$  eine Phrasenart ist. Zum Beispiel ist in der deutschen Sprache die Liste (*der, große, Mann*) ein Element einer Phrasenart, während die Liste (*ist, ist, ist, ist*) kein Element einer Phrasenart ist. Inhaltlich gesprochen ist auch die Menge aller Sätze eine Phrasenart.

Wir kürzen die *Menge aller Phrasenarten einer Sprache* durch  $PA$  und die *Menge aller Sätze einer Sprache* durch  $S$  ab. Die Phrasenarten werden oft in *lexikalische* Kategorien und Phrasenkategorien im engeren Sinne eingeteilt. Im Prinzip ist ein Satz nicht nur eine Liste von Wörtern, sondern auch eine Liste von Listen von Buchstaben. Eine Phrasenart ist nicht nur eine Menge von Listen von Wörtern, sondern auch eine Mengen von Listen von Listen von Buchstaben. Mengentheoretisch ist also  $S \subset L^* \subset (A_{bu}^*)^*$  und  $PA \subset \wp(L^*) \subset \wp((A_{bu}^*)^*)$ . Wir vermeiden aber die Buchstabenebene und begnügen uns mit der Rekonstruktion, die auch in *SynStr* in praktischen Belangen verwendet wird.

In *SynStr* wird noch eine vierte Menge von Hilfszeichen benutzt. Zum Beispiel gibt es einige Silben, in Englisch: *ing, en*, oder Symbole in der Phrasenebene, wie  $S, \emptyset$ , oder spezielle Klammern wie #, die in dem englischen Standardbeispiel von Chomsky eine Rolle spielen. Wir kürzen die *Menge der Hilfssymbole für eine Sprache* durch  $H$  ab.

---

*Buche* und dem Verb *stellen, aufstellen* ab. Die ersten Buchstaben waren in der jetzt als Deutschland bekannten Region die Runen.

<sup>11</sup> Als ein Standardbeispiel wird in *SynStr* etwa das Wort # *be + ing* # benutzt, in der deutschen Sprache wäre ein ähnliches Beispiel # *ge + geben* #.

<sup>12</sup>  $\wp(X)$  ist die *Potenzmenge* einer Menge  $X$ .

All diese verschiedenen Arten von Zeichen nennen wir im Weiteren systematisch *linguistische Elemente*. Die Menge all dieser linguistischen Elemente ist also identisch mit der Vereinigung folgender Mengen: der Menge  $U$  der Äußerungen, der Menge  $L$  der für Wörter, der Menge  $PA$  der Phrasenarten und der Menge  $H$  der Hilfssymbole. Diese Gesamtmenge kürzen wir durch  $B$  ab:

$$B = U \cup L \cup PA \cup H.$$

Diese linguistischen Elemente sind die Bausteine, aus denen letzten Endes die Sätze zusammengesetzt werden. Jeder Satz läßt sich von der theoretischen Seite her analysieren, indem der Satz in Phrasenarten und in weitere, kleinere Phrasenarten zerlegt wird. Es können auch Phrasenarten in andere Phrasenarten transformiert werden. Sobald diese Regeln abgearbeitet sind, enthalten die einfachsten Phrasenarten (von einigen Ausnahmen abgesehen) nur noch lexikalische Kategorien (Äquivalenzklassen von Wörter). Mit weiteren Regeln werden die Wörter in verschiedene grammatische Varianten der Wörter umgewandelt. Dieser hier sehr vage umschriebene Prozess wird in *SynStr* formal geklärt.

Dazu wurde der Begriff des Produktionssystems<sup>13</sup> verwendet. Als Alphabet eines Produktionssystems wird hier die Menge  $B$  benutzt. Weiter wird eine Menge von Startsymbolen<sup>14</sup> festgelegt. Schließlich werden drei Arten von Produktionsregeln eingeführt. In der allgemeinen, abstrakten Ebene lassen sich diese Regeln natürlich nicht so konkretisieren, dass diese direkt auf eine bestimmte Sprache angewendet werden können. Die allgemeine Theorie in *SynStr* kann nur die drei *Arten* von Regeln unterscheiden. In einem gegebenen Beispiel, lassen sich dann diese Regeln genau explizieren. In *SynStr* werden im dortigen Appendix II diese Arten in 26 Regeln für die englische Sprache aufgegliedert.<sup>15</sup> Im allgemeinen sind die Startsymbole Elemente von  $B$  (genauer: von  $PA$ ). Die Regeln  $r$  haben die Form  $((x_1, \dots, x_m), (y_1, \dots, y_n))$ , wobei alle  $x_1, \dots, x_m, y_1, \dots, y_n$  linguistische Elemente aus  $B$  sind.

Der zentrale Punkt von *SynStr* kreist um die drei verschiedenen Arten von Produktionsregeln. Die erste Art von Regeln betrifft die Phrasenstruktur, die zweite wandelt Phrasenarten in transformierte Phrasenarten um, und die dritte Art von Regeln formt, etwas vereinfacht gesagt, ein Wort in eine Wortäußerung um. Die Regeln der zweiten Art grenzen den Bereich ab, den Chomsky in *SynStr* als Transformationsstruktur (einer Sprache) bezeichnet. Solche Produktionsregeln transformieren Listen von linguistischen Elementen in andere linguistische Elemente. Die dritte Art von Regeln grenzt den Bereich von morphophonetischen Regeln ab, den Chomsky *morphophonemics* nennt.

Schließlich werden die drei Arten von Produktionsregeln in der „richtigen“ Reihenfolge angewendet. Bestimmte Regeln können erst benutzt werden, wenn vorher

<sup>13</sup>Dieser Term wird allerdings in *SynStr* mit anderen Worten umschrieben.

<sup>14</sup>In *SynStr* wird erwähnt, daß neben dem Startsymbol *Sent*, auch andere Startsymbole sinnvoll sind, wie *Quest* oder *Passiv*. Ein Satz kann eine Frage sein, er beginnt dann mit dem Startsymbol *Quest*. Ein Satz kann auch im Passiv mit dem Startsymbol *Passiv* beginnen.

<sup>15</sup>In *SynStr* wird nicht behauptet, dass diese Regeln vollständig sind.

andere Regeln schon abgearbeitet wurden. Diese Reihenfolge ist einerseits für die Linguistik zentral. Chomsky unterscheidet in *SynStr* zum Beispiel *obligatorische* und andere Regeln. Andererseits wird dort in verschiedenen Formulierungen gesagt, dass dieses Thema ziemlich neu und in vielen Bereichen noch nicht völlig untersucht ist. In *SynStr* wird ein erster Schritt unternommen, der die Reihenfolge eines Produktionsprozesses betrifft. „Normalerweise“ werden in einem Produktionsprozess zuerst Phrasenregeln benutzt, dann in einer zweiten Phase die Transformationsregeln und erst im dritten Schritt die morphophonetischen Regeln. Die zweite Phase kann in einfachen Sätzen auch fehlen. Diese so unterteilten, speziellen Prozesse bezeichnen wir auch als *linguistische Ableitungen*. Ein etwas allgemeinerer Begriff, der in der Linguistik auch schon vor *SynStr* benutzt wurde, beschreibt eine Äquivalenzklasse von linguistischen Ableitungen. Solche Äquivalenzklassen werden in der linguistischen Literatur, auch in *SynStr*, als *Bäume (trees)* bezeichnet und gezeichnet.<sup>16</sup>

Der Unterschied der drei Arten von Produktionsregeln wird in *SynStr* auf der allgemeinen Ebene nicht systematisch dargestellt, er wird dort nur in einer bestimmten Sprache, nämlich der englischen Sprache, genau erörtert.

### 3) Strukturalistische Rekonstruktion

Nach dem strukturalistischen Ansatz besteht eine empirische Theorie  $\mathbf{T}$  im Allgemeinen aus drei Teilen,<sup>17</sup> dem *Kern*  $\mathbf{K}$ , der Menge der *intendierten Systeme*<sup>18</sup> und dem *Approximationsapparat*  $\mathbf{A}$ :  $\mathbf{T} = \langle \mathbf{K}, \mathbf{A}, \mathbf{I} \rangle$ . Der *Kern*  $\mathbf{K}$  enthält die Klasse  $\mathbf{M}_p$  von *potentiellen Modellen*, die Klasse  $\mathbf{M}$  von *echten Modellen*, die Klasse der *Constraints* und die Klasse der *partiellen potentiellen Modelle*:

$$\mathbf{K} = \langle \mathbf{M}_p, \mathbf{M}, \mathbf{C}, \mathbf{M}_{pp} \rangle.$$

Die potentiellen und die echten Modelle sind mengentheoretische Strukturen der Form  $\langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_n \rangle$ . Die Strukturen beschreiben nicht nur die möglichen, sondern auch wirkliche Systeme. Einige der Strukturen enthalten sogar reale Teile von wirklichen Systemen. Nach einer bestimmten Sichtweise „ist“ eine solche Struktur ein wirkliches System.

In einem (potentiellen) Modell finden wir die *Grundmengen*  $D_1, \dots, D_k$ , die *Hilfsbasisismengen*  $A_1, \dots, A_m$  und die *Relationen* und *Funktionen*  $R_1, \dots, R_n$ , die zu einem mengentheoretischen Gesamtsystem integriert sind. Die Klasse der potentiellen Modelle beschreibt inhaltlich die Typisierung der Relationen und Funktionen, die über den Grundmengen und Hilfsbasisismengen konstituiert sind. Die Klasse  $\mathbf{M}$  ist eine Teilklasse von  $\mathbf{M}_p$ , sie stellt die empirischen Hypothesen dar, die für die Modelle charakteristisch und gültig sind. Die partiellen potentiellen Modelle entstehen aus den potentiellen

<sup>16</sup>In der nächsten Theorie von Chomsky werden solche linguistischen Ableitungen verallgemeinert und als *markers* bezeichnet, siehe (Chomsky, 1965).

<sup>17</sup>Siehe zum Beispiel (Balzer et al., 1987).

<sup>18</sup>Sie werden auch oft als *intendierte Anwendungen* bezeichnet.

Modellen, indem bestimmte Realisierungen  $R_{nt+1}, \dots, R_n$  der *theoretischen Terme* weggelassen werden. Mit anderen Worten wird eine Funktion  $\mathbf{r}$  definiert, mit der in jeder Struktur  $\langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_n \rangle$  genau die Komponenten  $R_{nt+1}, \dots, R_n$  entfernt werden. Eine Struktur  $\langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_n \rangle$  wird durch  $\mathbf{r}$  auf eine partielle potentielle Struktur  $\langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_{nt} \rangle$  ( $nt \leq n$ ) abgebildet ( $\mathbf{r}(\langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_n \rangle) = \langle D_1, \dots, D_k, A_1, \dots, A_m, R_1, \dots, R_{nt} \rangle$ ). Die Funktion  $\mathbf{r}$  bildet insbesondere auch jedes echte Modell auf ein partielles potentielles Modell ab. Die Menge der Funktionswerte von  $\mathbf{r}$ , die aus echten Modellen entstanden sind, d.h. die Menge  $\{\mathbf{r}(x)/x \in \mathbf{M}\}$ , nennen wir *die durch  $\mathbf{M}$  induzierte Klasse*.

Jedes intendierte System  $z$  ist ein partielles potentielles Modell der Theorie. Inhaltlich betrachtet, beschreibt oder „ist“  $z$  ein intendiertes System, das wirklich untersucht wurde oder das für eine Untersuchung bereitsteht. Schließlich wird der *Approximationsapparat  $\mathbf{A}$*  verwendet, mit dem die Menge  $\mathbf{I}$  der intendierten Systeme in eine durch  $\mathbf{M}$  induzierte Klasse  $\mathbf{r}(\mathbf{M})$  approximativ „eingepaßt“ wird ( $\mathbf{I} \subseteq^{\approx} \mathbf{r}(\mathbf{M})$ ).<sup>19</sup> Anders gesagt wird jedes intendierte System der Theorie in eine Restriktion eines echten Modells eingebettet, wobei die Einbettung einige Ausnahmen zulässt. Einige intendierte Systeme sind statistische „Ausrutscher“; sie liegen nur in der Nähe von eingeschränkten Modellen.

Wir rekonstruieren nun die in *SynStr* beschriebene Theorie in strukturalistischem Format. Die rekonstruierte Theorie bezeichnen wir als *Theorie der linguistischen Syntax*, kurz  $\mathbf{T}(LS)$ . Die Theorie  $\mathbf{T}(LS)$  hat vier Grundmengen: die Menge  $U$  der Wortäußerungen, die Menge  $L$  der Wörter, die Menge  $PA$  der Phrasenarten und die Menge  $H$  der Hilfssymbole, und fünf Relationen, die aus diesen Grundmengen konstituiert werden. Wir typisieren durch die Klasse  $\mathbf{M}_p$  die mengentheoretischen Strukturen der Theorie und formulieren durch  $\mathbf{M}$  die empirischen Hypothesen dieser Theorie.

**D1**  $\mathbf{M}_p(LS)$  die Klasse der *potentiellen Modelle der Theorie der linguistischen Syntax* wird wie folgt definiert. Für alle  $x$  gilt:  $x \in \mathbf{M}_p(LS)$  gdw es  $U, gS, L, PA, H, \Sigma, PhrR, TrR$  und  $mpR$  gibt, so dass

$$x = \langle U, gS, L, PA, H, \Sigma, PhrR, TrR, mpR \rangle \text{ und}$$

- 1)  $U$  ist eine nicht leere Menge (von *Wortäußerungen*)
- 2)  $L$  ist eine Menge (von *geschriebenen Wörtern*)<sup>20</sup>
- 3)  $PA$  ist eine endliche, nicht leere Menge (von *Phrasenarten*)
- 4)  $H$  ist eine endliche Mengen (von *Hilfssymbolen*)<sup>21</sup>
- 5)  $U, PA, H$  und  $L, PA, H$  sind paarweise disjunkt
- 6)  $gS \subseteq U^*$  und<sup>22</sup>  $gS \neq \emptyset$  ( $gS$  ist eine Menge von *gesprochenen Sätze*)

<sup>19</sup>Die approximative Komponente ist z.B. in (Balzer et al., 1987), Kap. 7, genau beschrieben.

<sup>20</sup>Diese Menge kann leer sein.

<sup>21</sup>Auch diese Menge kann in einfachen Fällen leer sein.

<sup>22</sup>Das Symbol  $\subseteq$  kürzt die Relation „ist eine (echte oder unechte) Teilmenge von“ ab und  $X \neq \emptyset$  bedeutet, dass  $X$  eine nicht leere Menge ist.



- 7)  $\Sigma \subseteq PA$  und  $\Sigma \neq \emptyset$  ( $\Sigma$  ist eine Menge von *Startsymbolen*)
- 8)  $PhrR \subseteq (L \cup PA \cup H)^* \times (L \cup PA \cup H)^*$  und  $PhrR \neq \emptyset$  ( $PhrR$  ist eine Menge von *Phrasenregeln*)
- 9)  $TrR \subseteq (L \cup PA \cup H)^* \times (L \cup PA \cup H)^*$  und  $TrR \neq \emptyset$  ( $TrR$  ist eine Menge von *Transformationsregeln*)
- 10)  $mpR \subseteq (L \cup U \cup H)^* \times (L \cup U \cup H)^*$  und  $mpR \neq \emptyset$  ( $mpR$  ist eine Menge von *morphophonetischen Regeln*)
- 11)  $(U \cup L \cup PA \cup H, \Sigma, PhrR \cup TrR \cup mpR)$  ist ein Produktionssystem.

Die Menge  $PhrR \cup TrR \cup mpR$  der *Phrasenregeln* kürzen wir durch  $PR$  ab. Durch die Vereinigungsmengen  $B$  und  $PR$  wird es möglich, den Gesamtproduktionsprozess eines Satzes durch eine einfache Notation durchsichtiger zu machen.

**D2**  $\mathbf{M}(LS)$  die Klasse der *Modelle der Theorie der linguistischen Syntax* wird wie folgt definiert. Für alle  $x$  gilt:  $x \in \mathbf{M}(LS)$  gdw es  $U, gS, L, PA, H, \Sigma, PhrR, TrR, mpR$  und *Sent, Quest, Passiv* gibt, so dass

$$x = \langle U, gS, L, PA, H, \Sigma, PhrR, TrR, mpR \rangle \text{ und}$$

- 1)  $x \in \mathbf{M}_p(LS)$
- 2) für jede Äußerung  $u \in U$  gibt es eine Phrasenkategorie  $pa \in PA$  so dass  $((pa), (u))$  eine Phrasenregel in  $PhrR$  ist
- 3) für jeden gesprochenen Satz  $s \in gS$  gibt es Äußerungen  $u_1, \dots, u_n \in U$  so dass  $s$  eine Liste aus den Äußerungen  $u_1, \dots, u_n$  ist ( $s = (u_1, \dots, u_n)$ )
- 4) es gibt Äußerungen  $u_1, \dots, u_m \in U$  so dass  $(u_1, \dots, u_m)$  kein gesprochener Satz ist ( $(u_1, \dots, u_m) \notin gS$ )
- 5)  $\Sigma \subseteq \{Sent, Quest, Passiv\}$  und  $Sent \in \Sigma$
- 6) für alle Startsymbole  $\sigma \in \Sigma$  gibt es Phrasenkategorien  $x_1, \dots, x_m \in PA$ , so dass  $((\sigma), (x_1, \dots, x_m))$  eine Phrasenregel in  $PhrR$  ist
- 7) für alle Phrasenregeln  $(v, h)$  in  $PhrR$  gibt es eine Phrasenkategorie  $x$  in  $PA$ , so dass  $v = (x)$
- 8) für alle Phrasenregeln  $((x_1, \dots, x_m), (y_1, \dots, y_n))$  in  $PhrR$  und für alle  $i, j \leq 1, \dots, n$  gilt, dass  $y_i \neq y_j$
- 9) für alle Transformationsregeln  $((x_1, \dots, x_m), (y_1, \dots, y_n))$  in  $TrR$  gilt  $\{y_1, \dots, y_n\} = \{x_1, \dots, x_m\}$ <sup>23</sup>
- 10) für alle Transformationsregeln  $((x_1, \dots, x_m), (y_1, \dots, y_n))$  in  $TrR$  ist  $m \geq 2$  und für alle  $i, j \leq 1, \dots, m$  gilt, dass  $x_i \neq x_j$
- 11) für alle Äußerungen  $u$  in  $U$  gibt es ein Vorderglied  $v$ , so dass  $(v, (u))$  eine morphophonetische Regel in  $mpR$  ist

---

<sup>23</sup> Diese etwas idealisierte Hypothese läßt sich abschwächen, indem die beiden Mengen nicht gleich (=) sondern nur ähnlich ( $\approx$ ) sind.

- 12) für alle gesprochenen Sätze  $s$  in  $gS$  gibt es einen beendeten Produktionsprozess  $p$ , der aus  $(B, \Sigma, PR)$  erzeugt ist, und Äußerungen  $u_1, \dots, u_n \in U$ , so dass die letzte Liste  $end(p)$  von  $p$  identisch mit  $s$  und mit  $(u_1, \dots, u_n)$  ist ( $end(p) = (u_1, \dots, u_n) = s$ ).

Den wissenschaftstheoretischen Begriff des Constraints gab es in der Zeit der Entstehung von *SynStr* noch nicht und er wurde auch nicht in anderer Terminologie angesprochen. Trotzdem gibt es mindestens drei Constraints von  $\mathbf{T}(LS)$ , die wir kurz erwähnen, und die an der Peripherie von  $\mathbf{T}(LS)$  liegen. Der erste Constraint besagt, dass die Mengen von Äußerungen in *allen* Sprachen mit demselben Lautalphabet dargestellt werden. Ein spezieller Constraint betrifft Gruppen von Sprachen, deren Wörter und Sätze mit dem gleichen Alphabet geschrieben werden. Verschiedene Sprachen werden mit dem lateinischen Alphabet, oder mit den chinesischen Wortzeichen oder z.B. mit dem russischen, mit einem indischen, mit dem hebräischen Alphabet geschrieben. Ein dritter Constraint beinhaltet eine spezielle Menge von Phrasenregeln, die in den indogermanischen Sprachen verwendet werden, nämlich die Subjekt-Verb Verbindung. In der deutschen Sprache haben die einfachsten Sätze die Form *Peter geht*. Die beiden ersten Constraints brauchen die Alphabete, die in der Rekonstruktion nicht explizit auftauchen, da sie in den „allgemeinen“ Hypothesen nicht gebraucht werden. Sie werden aber in speziellen Produktionsregeln sofort wichtig.<sup>24</sup> Formal lassen wir die Constraints hier einfach weg.

Die Unterscheidung zwischen  $\mathbf{T}(LS)$ -theoretischen und  $\mathbf{T}(LS)$ -nicht-theoretischen Komponenten ziehen wir wie folgt. Die Begriffe der Äußerung, der gesprochenen Sätze und der Wörter sind ohne weitere Argumentation  $\mathbf{T}(LS)$ -nicht-theoretisch. Genau so ist ohne große Mühe zu erkennen, dass die drei Begriffe von Produktionsregeln  $\mathbf{T}(LS)$ -theoretisch sind. Weiterhin ist die Menge der Hilfssymbole in *SynStr* eindeutig  $\mathbf{T}(LS)$ -theoretisch. Wir müssen nur entscheiden, ob die Mengen der Phrasenarten und der Startsymbole als  $\mathbf{T}(LS)$ -nicht-theoretisch eingestuft werden sollen. Einerseits wurden diese beiden Begriffe schon vor *SynStr* benutzt, andererseits gehören sie eindeutig in den theoretischen Bereich der linguistischen Untersuchung. Nach Sneed's Kriterium<sup>25</sup> scheint es möglich, diese beiden Mengen unabhängig von den Hypothesen von  $\mathbf{T}(LS)$  zu bestimmen. Wir können die detaillierte Begründung hier nicht liefern. Wir entschließen uns, die Phrasenarten und die Startsymbole als nicht-theoretisch einzustufen.

Somit hat ein partielles potentielles Modell von  $\mathbf{T}(LS)$  die Form  $\langle U, gS, L \rangle$ . Es hat zwei Komponenten. Die erste Komponente besteht aus der Grundmenge der Äußerungen und einer einstelligen Relation. Diese Relation ist mit anderen Worten die Menge der gesprochenen Sätze, das heißt eine Menge von Listen von Äußerungen. Darüber hinaus gibt es die zweite Komponente, nämlich die Menge der geschriebenen Wörter. Diese Menge ist in  $\mathbf{T}(LS)$  sozusagen die Doppelgängerin der Menge der Äußerungen. Wenn der Unterschied zwischen den geschriebenen und den geschprochenen

<sup>24</sup>Zum Beispiel in der Regel (29) für die englische Sprache in *SynStr*.

<sup>25</sup>Siehe (Sneed, 1971); ein zweites Kriterium findet sich in (Balzer, 1986).

enen Wörtern wegfallen würde, könnten wir eine dieser Mengen eliminieren. In der Linguistik ist aber das Produktionssystem so eingerichtet, dass die Phrasenarten erst in geschriebene Wörter und Sätze umgewandelt werden und erst dann die gesprochenen Wörter und Sätze analysiert werden. Es ist leicht, zusätzlich eine Grundmenge der geschriebenen Sätze einzuführen. Wir tun dies nur deshalb nicht, weil diese Sätze in  $\mathbf{T}(LS)$  in den Hypothesen nicht gebraucht werden. Es wäre ziemlich einfach, in der „mittleren“ Ebene eine Behauptung zu formulieren, die der hier nur dargestellten empirischen Behauptung sehr ähnlich wäre. Wir führen zusammenfassend die Klasse  $\mathbf{M}_{pp}$  ein, d.h.  $\mathbf{M}_{pp} = \{\langle U, gS, L \rangle / \exists x_4, \dots, x_9 (\langle U, gS, L, x_4, \dots, x_9 \rangle \in \mathbf{M}_p)\}$ .

Die intendierten Systeme beschreiben oder *sind*-richtig verstanden- die wirklichen, natürlichen Sprachen. Eine solche Beschreibung enthält also nur die Mengen der Äußerungen und der geschriebenen Wörter und die Menge der gesprochenen Sätze. Alle weiteren Aspekte dieser Sprachen sind für die Linguistik auf der  $\mathbf{T}(LS)$ -nicht-theoretischen Ebene nicht essentiell. Nach dem strukturalistischen Theorienbegriff sagt eine Metahypothese, dass das intendierte System ein partielles potentielles Modell ist:  $\mathbf{I} \subseteq \mathbf{M}_{pp}$ .

Aus den drei Klassen  $\mathbf{M}_{pp}, \mathbf{M}, \mathbf{I}$  wird die empirische Behauptung standardmäßig formuliert. Sie besagt, dass die Menge  $\mathbf{I}$  der intendierten Systeme approximativ in die durch  $\mathbf{M}$  induzierte Klasse  $\mathbf{r}(\mathbf{M})$  eingebettet ist; formal:  $\mathbf{I} \subseteq \mathbf{r}(\mathbf{M})$ . Inhaltlich gesprochen läßt sich jede natürliche Sprache, die durch die Äußerungen und die gesprochenen Sätze (und die geschriebenen Wörter) dieser Sprache dargestellt sind, mengentheoretisch approximativ einbetten in ein echtes Modell. Der Approximationsapparat ist nötig. Es gibt einerseits Folgen von Wortäußerungen, die durch die Ableitungsregeln hergeleitet werden können, aber nicht intendiert sind. Andererseits finden wir immer gesprochene „Sätze“, die mit den gegebenen Produktionsregeln nicht abgeleitet werden können.

Eine wichtige Frage erhebt sich in jeder empirischen Theorie, nämlich ob eine Theorie einen empirischen Gehalt hat. In (Sneed, 1971) wurde dies folgendermaßen ausgedrückt. Eine Theorie *hat* empirischen Gehalt genau dann wenn es eine Menge  $X$  von möglichen partiellen potentiellen Modellen gibt, die in eine Modellmenge eingepasst werden kann. *Basistheorien*, die zu interessanten Spezialisierungen führen, haben oft keinen empirischen Gehalt. Dies scheint auch auf  $\mathbf{T}(LS)$  zuzutreffen.

#### 4) Diskussion

Einige zentrale Punkte dieser Theorie sind in unserer Rekonstruktion aus strukturalistischer Sicht gut erkennbar. Erstens postuliert Chomsky in *SynStr* und in vielen seiner weiteren Werken, dass *alle* natürlichen Sprachen die gleiche Struktur haben. Wir konnten diese Struktur durch die zwei obigen Definitionen ziemlich durchsichtig machen. Da dieser Punkt oft mit vielen Worten umschrieben wird, möchten wir diesen Inhalt in unserer metatheoretischen Notation noch einmal kurz formulieren. Eine natürliche Sprache wird in  $\mathbf{T}(LS)$  durch ein intendiertes System  $z \in \mathbf{I}$  dargestellt. Mit

der Theorie  $\mathbf{T}(LS)$  wird behauptet, das der Bereich *aller* intendierten Systeme, d.h. aller natürlichen Sprachen, in eine theoretisch induzierte Menge  $\mathbf{r}(\mathbf{M})$  approximativ eingepaßt werden kann, oder noch gröber, dass jedes intendierte System approximativ zu einem Modell von *SynStr* passt. Die Frage, ob diese Behauptung empirisch mehr oder weniger bestätigt ist, können wir nur an die linguistischen Spezialisten weiterreichen.

Zweitens, betont Chomsky den Aspekt der Erzeugung eines, vieler oder sogar unendlich vieler Sätze. Die linguistischen Werke,<sup>26</sup> die vor der Veröffentlichung von *SynStr* publiziert wurden, betonen mehr den Aspekt der Analyse. In *SynStr* wurde zum ersten Mal der Erzeugungsprozess, das linguistische Produktionssystem, in den Mittelpunkt einer linguistischen Theorie gestellt, während die vorherigen Ansätze die Analyse der Sätze ins Zentrum stellen. Die fast triviale Feststellung, dass beide Aspekte: Erzeugung und Analyse, zusammengehören, möchten wird trotzdem noch einmal kurz formulieren. Wenn eine Person einen Satz (genauer: einen gesprochenen Satz) ausspricht, erzeugt sie eine Reihe von Äußerungen und auch einen Satz. Die Reihe von Äußerungen und der gesprochene Satz existiert vorher nicht. Der gesprochene Satz ist aus den verschiedenen, geäußerten Lauten zusammengesetzt. Umgekehrt hört die Person einen gesprochenen Satz und sie oder einige Linguisten zerlegen ihn in eine Reihe von Äußerungen oder Lauten. Der Zerlegungsprozess wird normalerweise nicht ins Bewußtsein der Person treten. Daher interessierten sich die Linguisten am Anfang besonders für solche Analysen. Es wäre aber zuviel gesagt, dass die Vorläufer den Erzeugungsaspekt nicht gekannt hätten. Insbesondere scheint es uns aus wissenschaftshistorischer Sicht viel zu früh, um zu beurteilen, ob *SynStr* eine echte, neue linguistische Weltansicht hervorgebracht hat.

Einige wissenschaftstheoretische Punkte, finden wir noch interessanter. Erstens entdeckten wir, dass die meisten der hier formulierten Hypothesen von  $\mathbf{T}(LS)$  in *SynStr* nur implizit erscheinen. Mit anderen Worten werden diese Hypothesen in *SynStr* nicht explizit formuliert. Dies trifft auf die Hypothesen 2-2, 2-3, 2-4, 2-8, 2-9, 2-10 und 2-11 zu. Zweitens, hält sich Chomsky bedeckt in der Frage, was eine Phrase ist. In *SynStr*<sup>27</sup> werden nur Phrasenstrukturen und Phrasenregeln erörtern. Ob die Sätze in der morphologischen Ebene in Phrasen zerlegt werden, scheint natürlich implizit richtig zu sein, aber die Frage bleibt im Raum, was die Entität der Art „Phrase“ sozusagen „ontologisch“ ist.

Drittens hatte Chomsky in seiner Dissertation viel Arbeit aufgewendet, um die verschiedenen linguistischen Ebenen systematisch zu klären, ob dies gelungen ist, möchten wir hier nicht beurteilen.<sup>28</sup> In *SynStr* werden diese verschiedenen Ebenen zwar an einigen Stellen ins Spiel gebracht. Aber in diesem Werk werden die Erzeugungsprozesse „in einem Guß“ beschrieben, ohne die vielen Stellen zu diskutieren, in denen die verschiedenen Produktionsregeln in mehreren Ebenen liegen. Auch in dem

<sup>26</sup> Etwa (Bloomfield, 1933) -der Vater des *american structuralism*- oder (Harris, 1951).

<sup>27</sup> Dasselbe gilt auch für das nächste Buch von Chomsky (Chomsky, 1965).

<sup>28</sup> Eine Publikation der Dissertation erschien erst viele Jahre nach *SynStr*, nämlich in (Chomsky, 1975).

zentralen Beispiel der englischen Sprache werden die Ebenen nur flüchtig erörtert. Aus wissenschaftstheoretischer Sicht enthält eine empirische Theorie normalerweise mehrere Grundmengen. Ob verschiedene Grundmengen auch zu verschiedenen Ebenen Anlaß geben, läßt sich nur in jedem Einzelfall ontologisch diskutieren. Wir haben in unserer Rekonstruktion die verschiedenen Ebenen systematisch, genauso wie Chomsky, nicht benutzt.

Viertens involviert der Gesamtprozess der Produktion eines Satzes alle Ebenen. Mit anderen Worten gibt es viele Produktionsregeln, die mehrere Ebenen benutzen. Eine der trickreichsten Regeln in *SynStr* ist die Regel, die auf Seite 39 beschrieben wird. Wir können leicht unterscheiden zwischen „reinen“ Regeln, die nur auf einer einzigen Ebene wirken und den „Mehrebenenregeln“. Eine reine Regel zerlegt etwa in der deutschen Sprache die Nominalphrase *der Baum* in den Artikel (determiner) *der* und das Substantiv *Baum*.

Fünftens wurden schon in *SynStr* Regeln formuliert und erörtert, die sich nicht unter ein einziges Produktionssystem subsumieren lassen. Zum Beispiel gibt es eine Regel, die zwei Ausgangssätze zu einem neuen Satz (ähnlich einer Konjunktion) vereinigen. In der deutschen Sprache kann man etwa den Satz *Peter geht nach Hause* und *Peter kocht* durch die von Chomsky beschriebene Regel (22.) zu *Peter geht nach Hause und kocht* zusammenführen. Diese Art von Regeln lassen sich ohne große Mühe formulieren, wenn wir das Produktionssystem mehrfach anwenden. So werden mehrere beendete Prozesse erzeugt. Mit den so erzeugten letzten Listen der Prozesse lassen sich mit weiteren Regeln „zweiter Stufe“ weitere Ableitungen bilden. Es wäre kein großer Aufwand, die Rekonstruktion in dieser Weise zu verallgemeinern. Sie wird allerdings auch komplexer. Wir haben hier nur den einfachsten Fall in Anlehnung an die Strategie, die auch Chomsky benutzt, formalisiert.

Ein sechster, metatheoretischer Punkt betrifft die Anzahl der Regeln, die für eine Sprache angemessen ist. In *SynStr* wird ein interessantes Regelsystem mit 26 Regeln formuliert. Einerseits ist es klar, das diese Regeln nur ein Fragment der englischen Sprache wiedergeben. Weiter ist klar, das dieses Fragment sehr viele Regeln enthält, die mit „etc.“ abgekürzt sind. Zum Beispiel muss es in einem Gesamtprozess eines Satzes eine Regel geben, die die Phrasenart *Substantiv* in ein bestimmtes Wort umwandelt, z.B. *Haus, Garten, Person, Frau, Wort* etc. In der deutsche Sprache sind die wichtigen Regeln etwa im Duden aufgelistet.<sup>29</sup> Andererseits schreiben die meisten Linguisten -insbesondere Chomsky-, dass eine natürliche Sprache durch eine endliche Anzahl von Regeln abgedeckt werden kann, aber das Produktionssystem eine unendliche Menge von Sätze dieser Sprache erzeugen kann. Wir meinen, dass die hier gestellte und oft diskutierte Frage, durch unsere Rekonstruktion einen interessanten Aspekt bekommt. Es gibt natürlich mengentheoretisch mögliche partielle potentielle Modelle, die aus unendlich vielen Elementen bestehen. Aber ob in einem intendierten System die Menge  $gS$  unendlich viele Elemente enthält, ist damit nicht entschieden. Unserer Meinung nach sind diese Mengen  $gS$  „natürlich“ endlich. Diese Frage sollte

---

<sup>29</sup>Z.B. (Duden, 1980), Band 1.

wissenschaftstheoretisch unter dem Aspekt der Approximation diskutiert werden.

Schließlich wollten wir das Problem des empirischen Gehalts  $\mathbf{T}(LS)$  beantworten. Leider haben wir es unter dem Zeitconstraint der Festschrift nicht lösen können. Inhaltlich sieht man sofort, dass die hier formulierten Hypothesen sehr viel Raum für „unsinnige“ Produktionsregeln bieten.

## Bibliographie

- Aristotle. 1949. *Categoriae et Liber de Interpretatione*. Hrg. L. Minio-Paluello, Oxford.
- Balzer, W. 1986. Theoretical Terms: A New Perspective. *The Journal of Philosophy* 83, No.2, 71-90.
- Balzer, W., Moulines, C.U., Sneed J.D. 1987. *An Architectonic for Science*. Dordrecht, Reidel.
- Balzer, W., Moulines, C.U. (eds.). 1996. *Structuralist Theory of Science - Focal Issues, New Results*. Berlin - New York, Walter de Gruyter.
- Balzer, W., Sneed J.D., Moulines, C.U. (eds.) 2000. *Structuralist Knowledge Representation Paradigmatic Examples*. Amsterdam - Atlanta, Rodopi.
- Bloomfield, L. 1933. *Language*. New York, Holt, Rinehalt and Winston.
- Chomsky, N. 1957. *Syntactic Structures*. Berlin - New York, Mouton de Gruyter, (2nd edition 2002).
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass., MIT.
- Chomsky, N. 1975. *The Logical Structure of Linguistic Theory*. New York und London, Plenum Press.
- Duden, Band I. 1980. *Rechtschreibung der deutschen Sprache und der Fremdwörter*, Mannheim, BI.
- Diederich, W., Ibarra, A., Mormann, T. (1989). Bibliography of Structuralism 1971-1988. *Erkenntnis* 30, 387-407.
- Diederich, W., Ibarra, A., Mormann, T. (1994). Bibliography of Structuralism II 1989-1994 and Additions. *Erkenntnis* 41. 403-418.
- Gonzalo, A. 2001. Cambios modeloteóricos en la lingüística chomskiana. Una reconstrucción desde la concepción estructural de la ciencia. Dissertation, Universidad de Buenos Aires.
- Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago, University of Chicago Press.
- Newmeyer, F. 1996. *Generative Linguistics. A Historical Perspective*. London and New York, Routledge.
- Peris Viñé, L.M. 1990. First Steps on the Reconstruction of Chomskyan Grammar. In: Diez, A. et al. (eds.) *Structures of Mathematical Theories*, San Sebastian, Servicio Editorial Universidad del Pais Vasco, 83-87.
- Quesada, D. 1993. Grammar as a Theory. An Analysis of the Standard Model of Syntax within the Structural Program. In: Diez, A. et al. (eds.) *Structures*

- of Mathematical Theories*, San Sebastian, Servicio Editorial Universidad del Pais Vasco, 175-182.
- Sneed, J.D. 1971. *The Logical Structure of Mathematical Physics*. Dordrecht, Reidel.
- Thue, A. 1914. Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln. Videnskapselskapeta Skrifter, *J. Math. Naturw. Kl.*, 3-34.