22. Die Chomsky-Hierarchie

Durch Einschränkungen der syntaktischen Gestalt der Regeln hat Chomsky eine Reihe von Grammatik-Typen eingeführt. Um diese zu definieren, müssen wir zunächst eine Beschränkung der Regeln mit leerer Konklusion betrachten.

22.1 DEFINITION. Sei $G = (N,T,P,S)$ eine Grammatik. Eine Regel $w \to \lambda$ heißt *λ-Regel*. Die Grammatik G heißt *λ-frei*, wenn P keine λ-Regel enthält, und G heißt *λ-treu*, wenn entweder G λ-frei ist oder wenn $S \to \lambda$ die einzige λ-Regel in P ist und das Axiom in keiner Konklusion einer Regel in P vorkommt.

Für λ-treues G gilt also $\lambda \in L(G)$ genau dann, wenn $S \to \lambda$ in P vorkommt, und in diesem Fall ist $S \Rightarrow \lambda$ die einzige Herleitung des leeren Wortes.

22.2 DEFINITION. (a) Jede Grammatik $G = (N,T,P,S)$ ist *vom Typ 0*.

(b) Ist die Grammatik $G = (N,T,P,S)$ λ-treu, und sind die Nicht-λ-Regeln von G von der Form

$$v \to w \quad \text{mit } |v| \leq |w|,$$

so ist G vom *Erweiterungstyp*.

(c) Ist die Grammatik $G = (N,T,P,S)$ λ-treu, und sind die Nicht-λ-Regeln von G von der Form

$$xXy \to xwy \quad \text{mit } X \in N,\ w \in (N \cup T)^+,\ x,y \in (N \cup T)^*,$$

so ist G *kontextsensitiv* (ks) oder vom *Typ 1*.

(d) Enthält die Grammatik $G = (N,T,P,S)$ nur Regeln der Form

$$X \to w \quad \text{mit } X \in N \text{ und } w \in (N \cup T)^*,$$

so ist G *kontextfrei* (kf) oder vom *Typ 2*.

(e) Enthält die Grammatik $G = (N,T,P,S)$ nur Regeln der Form

$$X \to xYy \quad \text{mit } X \in N,\ Y \in N \cup \lambda,\ x,y \in T^*,$$

so ist G *linear*.

(f) Enthält die Grammatik $G = (N,T,P,S)$ nur Regeln der Form

$$X \to xY \quad \text{mit } X \in N,\ Y \in N \cup \lambda,\ x \in T^*,$$

so ist G *rechtslinear* oder vom *Typ 3*.

(g) Enthält die Grammatik $G = (N, T, P, S)$ nur Regeln der Form

$$X \to Yx \quad \text{mit } X \in N, Y \in N \cup \lambda, x \in T^*,$$

so ist G *linkslinear*.

(h) Die Grammatik G ist *einseitig linear*, wenn sie links- oder rechtslinear ist.

Eine Chomsky-Grammatik vom Typ i nennen wir auch kurz *Chomsky-i-Grammatik* ($i \in \{0, 1, 2, 3\}$).

Durch eine kontextsensitive Regel $xXy \to xwy$ wird eine Variable (X) durch ein nichtleeres Wort (w) ersetzt. Diese Ersetzung ist jedoch nur dann zulässig, wenn die Variable in einem vorgegebenen Kontext (x links und y rechts der Variable) steht. Bei kontextfreien Regeln $X \to w$ wird ebenfalls eine Variable ersetzt. Hier kann die Regel aber in jedem Kontext angewendet werden. Auch verlangt man hier nicht, dass das Wort w, das X ersetzt, nicht leer ist. Kontextfreie Grammatiken müssen daher nicht λ-treu und daher i. Allg. nicht kontextsensitiv sein. Man beachte jedoch, dass kontextsensitive Regeln nicht verkürzend sind. D.h. jede kontextsensitive Grammatik ist auch vom Erweiterungstyp. (Die Umkehrung gilt nicht, so ist z.B. die Regel $XY \to YX$ mit $X, Y \in N$ vom Erweiterungstyp, aber nicht kontextsensitiv.)

Die (einseitig) linearen Grammatiken sind spezielle kontextfreie Grammatiken, bei denen in der Konklusion jeder Regel höchstens eine Variable vorkommt. Die echten Satzformen in Herleitungen aus dem Axiom enthalten also genau eine Variable, die bei rechts-(links-)linearen Grammatiken zudem der letzte (erste) Buchstabe der Satzform ist.

Die Beziehungen zwischen den verschiedenen Grammatik-Typen können wir also wie folgt festhalten:

$$G \text{ kontextsensitiv} \;\Rightarrow\; G \text{ vom Erweiterungstyp} \;\Rightarrow\; G \text{ vom Typ } 0 \tag{22.1}$$

$$G \text{ einseitig linear} \;\Rightarrow\; G \text{ linear} \;\Rightarrow\; G \text{ kontextfrei} \tag{22.2}$$

(Dass keine weiteren Beziehungen zwischen diesen Konzepten gelten, kann man leicht durch Beispiel-Grammatiken belegen.)

Die verschiedenen Grammatiktypen führen zu entsprechenden Sprachklassen.

22.3 DEFINITION. Eine Spache L ist vom *Typ i*, $i \in \{0, 1, 2, 3\}$, (*bzw. vom Erweiterungstyp, kontextsensitiv, kontextfrei, linear, rechtslinear, linkslinear*), falls es eine Grammatik G vom Typ i, $i \in \{0, 1, 2, 3\}$, (bzw. vom Erweiterungstyp, kontextsensitive, kontextfreie, lineare, rechtslineare, linkslineare) gibt, die L erzeugt.

Eine (Chomsky-)Sprache vom Typ i nennen wir auch kurz *Chomsky-i-Sprache*. Mit CH-i, ERW, KS, KF, LIN, RLIN und LLIN bezeichnen wir entsprechend die Klassen der Typ-i, Erweiterungstyp-, kontextsensitiven, linearen, rechtslinearen und linkslinearen Sprachen.

22.4 BEISPIEL. Die Beispiel-Grammatiken und Sprachen in Abschnitt 21 sind von folgenden Typen: Die Grammatik aus den Beispielen 21.4 und 21.7 ist kontextfrei (und λ-frei), aber nicht linear. Die Grammatik in Beispiel 21.8 ist rechtslinear, d.h. $L_1 = \{0^m 1^n : m, n \geq 1\} \in \mathrm{RLIN}$. Die Grammatik in Beispiel 21.9 ist linear aber weder rechts- noch linkslinear, also $L_2 = \{0^n 1^n : n \geq 1\} \in \mathrm{LIN}$. In Beispiel 21.10 haben wir eine kontextfreie, nichtlineare Grammatik G und sehen damit, dass $L_3 = \{0^n 1^n 0^m : n, m \geq$

$1\} \in$ KF gilt. Die Grammatik in Beispiel 21.11 schließlich ist vom Erweiterungstyp, enthält aber nicht kontextsensitive Regeln (z.B. die Regeln der Gruppe (IIa)). Es gilt also $L_4 = \{0^n1^n0^n : n \geq 1\} \in \mathrm{ERW}$.

Die Beziehungen zwischen den verschiedenen Chomsky-Sprachklassen fassen wir in folgendem Hierarchiesatz zusammen.

22.5 SATZ. (CHOMSKY-HIERARCHIESATZ)

$$\begin{array}{c} \mathrm{CH}\ (=\ \mathrm{CH\text{-}0}) \\ \cup\!\!\!\!\not\,\, \\ \mathrm{ERW}\ =\ \mathrm{KS}\ (=\ \mathrm{CH\text{-}1}) \\ \cup\!\!\!\!\not\,\, \\ \mathrm{KF}\ (=\ \mathrm{CH\text{-}2}) \\ \cup\!\!\!\!\not\,\, \\ \mathrm{LIN} \\ \cup\!\!\!\!\not\,\, \\ \mathrm{LLIN}\ =\ \mathrm{RLIN}(=\ \mathrm{CH\text{-}3}) \end{array}$$

Wir werden den Hierarchiesatz in den nächsten drei Abschnitten, in denen wir die Typ-1, Typ-2 und Typ-3-Sprachen genauer betrachten werden, schrittweise beweisen. Hierbei werden wir jedoch die weniger interessante Spachklasse LIN aus Zeitgründen weglassen. Wir werden sehen, dass die Sprachklassen aus Beispiel 1 oben die Trennungen im unteren Bereich bezeugen: $L_2 \in \mathrm{LIN} - \mathrm{RLIN}$, $L_3 \in \mathrm{KF} - \mathrm{LIN}$ (wird nicht gezeigt werden), $L_4 \in \mathrm{KS} - \mathrm{KF}$[1]. Ein Beispiel für eine Sprache in $\mathrm{CH} - \mathrm{KS}$ ist das Halteproblem.

Um das Zusammenfallen von Sprachklassen zu zeigen, benötigen wir folgende (bei Maschinen bereits entsprechend benutzte) Sprechweise.

22.6 DEFINITION. Zwei Grammatiken G und G' sind *äquivalent*, wenn sie dieselbe Sprache erzeugen, d.h. wenn $L(G) = L(G')$ gilt.

Hiermit ist z.B. die Aussage, dass $\mathrm{ERW} = \mathrm{KS}$ gilt, äquivalent zu der Aussage, dass es zu jeder Grammatik G vom Erweiterungstyp eine äquivalente kontextsensitive Grammatik G' gibt und umgekehrt.

Aus den Beobachtungen zu den verschiedenen Grammatik-Typen ergeben sich folgende triviale Beziehungen zwischen den Sprachklassen der Chomsky-Hierarchie.

22.7 LEMMA. *(i)* $\mathrm{KS} \subseteq \mathrm{ERW} \subseteq \mathrm{CH}$

(ii) $\mathrm{LIN} \subseteq \mathrm{KF}$

(iii) $\mathrm{RLIN} \subseteq \mathrm{LIN}$ *und* $\mathrm{LLIN} \subseteq \mathrm{LIN}$

BEWEIS. Dies folgt unmittelbar aus (22.1) und (22.2). □

[1] Da diese Sprachen binär sind, gilt die Echtheit der Inklusionen $\mathrm{RLIN} \subsetneq \mathrm{LIN} \subsetneq \mathrm{KF}$ für Binärsprachen. Man hat jedoch gezeigt, dass für Unärsprachen, die Klassen RLIN, LIN und KF zusammenfallen (s. Literatur). Die Echtheit der Inklusionen $\mathrm{KF} \subsetneq \mathrm{KS} \subsetneq \mathrm{CH}$ gilt dagegen auch für Unärsprachen, wie wir sehen werden.