Dr. Frank Drewes
Studiengang Informatik
MZH 3270, Tel.: 4335
email: drewes@informatik.uni-bremen.de

19. November 1997

Theoretische Informatik 2

– Ergänzungen zum Skript (1) –

Die Chomsky-Hierarchie

Wie in Abschnitt 4 des Skripts gezeigt wurde, haben Chomsky-Grammatiken die ungünstige Eigenschaft, daß das Wortproblem für die erzeugten Sprachen im allgemeinen unentscheidbar ist. Grammatiken, die solche Sprachen erzeugen, sind z.B. für die Definition der Syntax von Programmiersprachen offenbar ungeeignet. Es stellt sich daher die Frage, ob man durch geeignete zusätzliche Bedingungen – insbesondere an die Form der Regeln – zu eingeschränkten Klassen von Chomsky-Grammatiken gelangen kann, die bessere Eigenschaften haben, aber trotzdem noch genügend Allgemeinheit besitzen. Dies führt zur sogenannten *Chomsky-Hierarchie*. Wie der Name andeutet, handelt es sich dabei um eine Hierarchie mehr oder weniger eingeschränkter Typen von Chomsky-Grammatiken. Praktisch der gesamte weitere Inhalt des Kurses *Theoretische Informatik 2* beschäftigt sich mit der Untersuchung dieser Grammatik-Typen bzw. mit den von ihnen definierten Sprachklassen.

Definition (Chomsky-Hierarchie)

Eine Chomsky-Grammatik $G = (N, T, P, S)$ ist

(1a) *monoton*, falls $|u| \leq |v|$ für alle $(u, v) \in P$;

(1b) *kontext-sensitiv*, falls alle Regeln in P die Form $u_1Au_2 ::= u_1vu_2$ haben, wobei u_1, u_2 und v Wörter sind, $v \neq \lambda$, und $A \in N$;

(2) *kontextfrei*, falls $u \in N$ für alle Regeln $(u, v) \in P$;

(3) *rechtslinear* (auch *regulär* genannt), falls für alle Regeln $(u, v) \in P$ gilt, daß $u \in N$ und entweder $v \in T^*$ oder $v = v'B$ mit $v' \in T^+$ und $B \in N$. (Hierbei bezeichnet T^+ die Menge $T^* \setminus \{\lambda\}$.)

Monotone und kontext-sensitive Gammatiken werden auch *Grammatiken vom Typ 1* genannt, kontextfreie werden als *Typ-2-* und rechtslineare als *Typ-3-Grammatiken* bezeichnet. Allgemeine Chomsky-Grammatiken werden Grammatiken vom Typ 0 genannt. Eine Sprache L ist vom Typ i wenn es eine Grammatik dieses Typs gibt, die L erzeugt.

Der folgende Satz zeigt in seinem ersten Teil, warum die obige Typeneinteilung nicht zwischen monotonen und kontext-sensitiven Grammatiken unterscheidet. Der zweite Teil liefert die Berechtigung, von einer *Hierarchie* zu sprechen.

Satz

1. Monotone und kontext-sensitive Grammatiken erzeugen dieselbe Klasse von Sprachen.
2. Sei $\mathcal{L}_i$ die Menge aller Sprachen des Typs i ($i \in \{0, \ldots, 3\}$). Dann gilt
   - $\mathcal{L}_1 \subsetneqq \mathcal{L}_0$ (Monotonie bzw. Kontext-Sensitivität ist eine echte Einschränkung),
   - $\{L \setminus \{\lambda\} \mid L \in \mathcal{L}_2\} \subsetneqq \mathcal{L}_1$ (bis auf die bei monotonen Grammatiken offenbar nicht bestehende Möglichkeit, das leere Wort zu erzeugen, ist Kontextfreiheit eine echte Einschränkung gegenüber Kontext-Sensitivität),
   - $\mathcal{L}_3 \subsetneqq \mathcal{L}_2$ (Rechtslinearität ist eine echte Einschränkung gegenüber Kontextfreiheit).

Bemerkung Die inzwischen bereits in der Vorlesung behandelte Entscheidbarkeit des Wortproblems für Typ-1-Sprachen (siehe auch Abschnitt 5 des Skripts) ist der erste der oben erwähnten Sätze, aus denen sich die Aussagen des obigen Satzes (in diesem Fall $\mathcal{L}_1 \subsetneqq \mathcal{L}_0$) als Folgerungen ableiten lassen. Zusammen mit der Unentscheidbarkeit des Wortproblems im allgemeinen Fall ergibt sich nämlich aus der Entscheidbarkeit für $\mathcal{L}_1$ die Ungleichheit $\mathcal{L}_1 \neq \mathcal{L}_0$ (und $\mathcal{L}_1 \subseteq \mathcal{L}_0$ gilt ohnehin per Definition).