

Eine kontextfreie Chomsky-Grammatik

Ein *Semi-Thue-System* über einem Alphabet Σ ist gegeben durch eine endliche Menge von *Regeln* (auch *Produktionen* genannt). Jede Regel hat die Form $x \rightarrow y$ mit Wörtern $x, y \in \Sigma^*$. Bei einer *Chomsky-Grammatik* wird angenommen, dass das Alphabet Σ in zwei Teilmengen zerlegt ist, nämlich die Menge der *terminalen Zeichen* und die Menge der *nichtterminalen Zeichen* (auch *syntaktische Variablen* genannt), wobei eines der nichtterminalen Zeichen als Startsymbol S ausgezeichnet ist.

Noam Chomsky hat seine Grammatiken eingeführt zur Beschreibung der Syntax von Sätzen und Satzteilen (auch Phrasen genannt) in einer natürlichen Sprache wie etwa Englisch oder Deutsch. Phrasen können z.B. einzelne Wörter der natürlichen Sprache wie der Artikel (englisch: determiner) „the“, Nominalphrasen (englisch: noun phrase) wie die Phrase „the man“ oder sogar ganze Sätze sein.

Die Idee, die Noam Chomsky hatte, ist, dass es die Produktionen es erlauben, aus einem nichtterminalen Zeichen eine bestimmte Art von Phrasen herzuleiten. Zum Beispiel könnte es ein nichtterminales Zeichen geben, aus dem man Nominalphrasen herleiten kann. Dieses nichtterminale Zeichen hat dann in gewissem Sinne die Bedeutung „Nominalphrase“.

Beispiel Als Beispiel betrachten wir eine rudimentäre Grammatik der englischen Sprache. Der Einfachheit halber stellen wir jedes Wort der englischen Sprache, das in unserem Beispiel vorkommt als ein Zeichen unseres Alphabets Σ dar. Z.B. wird das englische Wort „the“ dargestellt als Zeichen $\boxed{\text{the}}$ des Alphabets Σ . Insgesamt haben wir die folgenden Zeichen:

$\boxed{\text{the}}, \boxed{\text{apple}}, \boxed{\text{man}}, \boxed{\text{eats}}, S, \mathcal{NP}, \mathcal{VP}, D, N, V$

Die Menge, die aus diesen zehn Zeichen besteht, sei unser Alphabet Σ .

Terminale Zeichen

$\boxed{\text{the}}, \boxed{\text{apple}}, \boxed{\text{man}}, \boxed{\text{eats}}$

Nichtterminale Zeichen und ihre jeweilige Bedeutung:

S	sentence
\mathcal{NP}	noun phrase
\mathcal{VP}	verb phrase
D	determiner
N	noun
V	verb

S ist das *Startsymbol*

Regeln :

$$S \rightarrow NPVP$$

$$NP \rightarrow DN$$

$$VP \rightarrow VNP$$

$$VP \rightarrow V$$

$$D \rightarrow \boxed{\text{the}}$$

$$N \rightarrow \boxed{\text{apple}}$$

$$N \rightarrow \boxed{\text{man}}$$

$$V \rightarrow \boxed{\text{eats}}$$

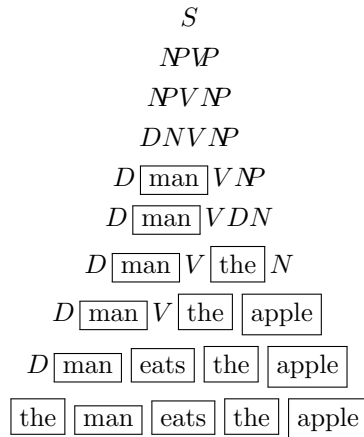
Anwendung einer Regel auf ein Wort des Alphabets der Zeichen der Grammatik bedeutet, dass in dem Wort ein Vorkommen des Zeichens auf der linken Seite der Regel durch das Wort auf der rechten Seite der Regel ersetzt wird. So kann man z.B. im Wort $NPVP$ das erste Vorkommen des Zeichens NP mit Hilfe der zweiten Regel durch das Wort DN ersetzen und erhält dann das Wort $DNVP$.

Durch wiederholte Anwendung der Regeln kann man nun aus dem Startsymbol S einen Satz der englischen Sprache *herleiten*. Da die nichtterminalen Zeichen nur Hilfszeichen jeweils zur Bezeichnung einer bestimmten Art von Phrasen sind, sollen am Ende nur noch terminale Zeichen dastehen. Mathematisch exakter ausgedrückt:

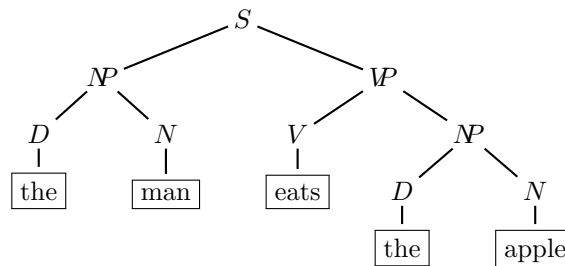
Eine *Herleitung* eines aus terminalen Zeichen bestehenden Wortes w ist eine endliche Folge (w_0, \dots, w_n) von Wörtern mit

1. $w_0 = S$ (Wort, das nur aus dem Startsymbol besteht)
2. w_{i+1} ergibt sich durch Anwendung einer Regel auf w_i $(0 \leq i < n)$
3. $w_n = w$.

Eine Herleitung des Wortes the man eats the apple (über dem Alphabet der terminalen Zeichen) in dieser Grammatik (Wörter der Herleitung untereinander geschrieben):



Der Syntaxbaum (parse tree) zu dieser Herleitung:



In unserem Beispiel hat jede Produktion die Form $v \rightarrow w$, wobei v ein Wort ist, das nur aus einem nichtterminalen Zeichen besteht. Wir sprechen von einer *kontextfreien* Grammatik. In allgemeinen (Typ 0) Chomsky-Grammatiken kann die linke Seite einer Produktion aus mehreren terminalen und nichtterminalen Zeichen bestehen. Im Gegensatz zu kontextfreien Chomsky-Grammatiken eignen sich allgemeine Chomsky-Grammatiken nicht zur Aufstellung eines Syntaxbaumes. Andererseits können allgemeine Chomsky-Grammatiken beliebige (rekursiv) aufzählbare Sprachen erzeugen, was mit kontextfreien Chomsky-Grammatiken nicht möglich ist. Ebenso kann man mit allgemeinen Chomsky-Grammatiken einen Berechenbarkeitsbegriff definieren, der alle partiellberechenbaren Funktionen umfasst.