Motivation – natürliche Sprachen

(Satz) $\rightarrow$ (Substantivphrase)(Verbphrase)
(Satz) $\rightarrow$ (Substantivphrase)(Verbphrase)(Objektphrase)

(Substantivphrase) $\rightarrow$ (Artikel)(Substantiv)

(Verbphrase) $\rightarrow$ (Verb)(Adverb)

(Substantiv) $\rightarrow$ Hund	(Substantiv) $\rightarrow$ Banane
(Artikel) $\rightarrow$ der	(Artikel) $\rightarrow$ ein
(Verb) $\rightarrow$ geht	(Verb) $\rightarrow$ singt
(Adverb) $\rightarrow$ langsam	

Motivation – natürliche Sprachen

$$
\begin{aligned}
\text{(Satz)} &\Longrightarrow \text{(Substantivphrase)(Verbphrase)}\\
&\Longrightarrow \text{(Substantivphrase)(Verb)(Adverb)}\\
&\Longrightarrow \text{(Substantivphrase) geht (Adverb)}\\
&\Longrightarrow \text{(Substantivphrase) geht langsam}\\
&\Longrightarrow \text{(Artikel)(Substantiv) geht langsam}\\
&\Longrightarrow \text{der (Substantiv) geht langsam}\\
&\Longrightarrow \text{der Hund geht langsam}
\end{aligned}
$$

$$\text{(Satz)} \Longrightarrow \ldots \Longrightarrow \text{ein Banane singt langsam}$$

Motivation – Programmiersprachen

$\langle$unsigned integer$\rangle \rightarrow \langle$digit$\rangle \mid \langle$digit$\rangle\{$digit$\}$
$\langle$unsigned real$\rangle \rightarrow \langle$unsigned integer$\rangle.\langle$digit$\rangle\{$digit$\} \mid$
$\qquad \langle$unsigned integer$\rangle$E$\langle$scale factor$\rangle$
$\langle$scale factor$\rangle \rightarrow \langle$unsigned integer$\rangle \mid \langle$sign$\rangle\ \langle$unsigned integer$\rangle$
$\langle$digit$\rangle \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$
$\langle$sign$\rangle \rightarrow + \mid -$

$\langle$unsigned real$\rangle \Longrightarrow \langle$unsigned integer$\rangle$E$\langle$scale factor$\rangle$
$\Longrightarrow \langle$digit$\rangle\{$digit$\}$E$\langle$scale factor$\rangle$
$\Longrightarrow 3\{$digit$\}$E$\langle$scale factor$\rangle$
$\Longrightarrow$ 314E$\langle$scale factor$\rangle \Longrightarrow$ 314E$\langle$sign$\rangle\langle$unsigned integer$\rangle$
$\Longrightarrow$ 314E–$\langle$unsigned integer$\rangle \Longrightarrow$ 314E–$\langle$digit$\rangle$
$\Longrightarrow$ 314E–2

Regelgrammatik – Definition

Definition: Eine Regelgrammatik (oder kurz Grammatik) ist ein Quadrupel

$$G = (N, T, P, S),$$

wobei

- N und T endliche, disjunkte Alphabete sind ($V = N \cup T$),
- P eine endliche Teilmenge von $(V^* \setminus T^*) \times V^*$ ist, und
- $S \in N$ gilt.

Regelgrammatik – Ableitung und Sprache

Definition: Sei $G = (N, T, P, S)$ eine Regelgrammatik.
Wir sagen, dass aus dem Wort $\gamma \in V^+$ das Wort $\gamma' \in V^*$ erzeugt wird, wenn

$$\gamma = \gamma_1 \alpha \gamma_2, \quad \gamma' = \gamma_1 \beta \gamma_2, \quad \alpha \longrightarrow \beta \in P$$

für gewisse $\gamma_1, \gamma_2 \in V^*$ gelten.

Schreibweise: $\gamma \Longrightarrow \gamma'$

$\Longrightarrow^*$ — reflexiver und transitiver Abschluss von $\Longrightarrow$

Definition: Für eine Grammatik $G = (N, T, P, S)$ definieren wir die von G erzeugte Sprache $L(G)$ durch

$$L(G) = \{w : w \in T^* \text{ } und \text{ } S \Longrightarrow^* w\}.$$

Regelgrammatik – Beispiele I

$G_1 = (\{S, A, B\}, \{a, b\}, \{p_1, p_2, p_3, p_4, p_5\}, S)$
$p_1 = S \to AB,\ p_2 = A \to aA,\ p_3 = A \to \lambda,\ p_4 = B \to Bb,\ p_5 = B \to \lambda$
$L(G_1) = \{a^n b^m : n \geq 0, m \geq 0\}$

$G_2 = (\{S\}, \{a, b\}, \{S \longrightarrow aSb, S \longrightarrow ab\}, S)$
$L(G_2) = \{a^n b^n : n \geq 1\}$

$G_3 = (\{S, A\}, \{a, b\}, \{S \longrightarrow \lambda, S \longrightarrow aS, S \longrightarrow Sb\}, S)$
$L(G_3) = \{a^n b^m : n \geq 0, m \geq 0\}$

$G_4 = (\{S, A\}, \{a, b\}, P_4, S)$
$P_4 = \{S \longrightarrow \lambda, S \longrightarrow aS, S \longrightarrow a, S \longrightarrow A, A \longrightarrow bA, A \longrightarrow b\}$
$L(G_4) = \{a^n b^m : n \geq 0, m \geq 0\}$

Regelgrammatik – Beispiele II

$G_i = (\{S, A, B, B', B''\}, \{a, b, c\}, P_i, S)$ für $i \in \{5, 6\}$

P_5	P_6
	$p_0 = S \to abc$
$p_1 = S \to ABA$	$p_1 = S \to aABbA$
$p_2 = AB \to aAbB'$	$p_2 = AB \to aAbB'$
$p_3 = AB \to abB''$	$p_3 = AB \to abB''$
$p_4 = B'b \to bB'$	$p_4 = B'b \to bB'$
$p_5 = B''b \longrightarrow bB''$	$p_5 = B''b \longrightarrow bB''$
$p_6 = B'A \longrightarrow BAc$	$p_6 = B'A \longrightarrow BAc$
$p_7 = B''A \longrightarrow c$	$p_7 = B''A \longrightarrow cc$
$p_8 = bB \longrightarrow Bb$	$p_8 = bB \longrightarrow Bb$

$L(G_5) = L(G_6) = \{a^n b^n c^n \mid n \geq 1\}$

Regelgrammatik – Beispiele III

$G_7 = (N, T, P, S)$ mit

$$\begin{aligned} N &= \{S\}, \\ T &= \{x, y, z, +, -, \cdot, :, (,)\}, \\ P &= \{S \longrightarrow (S + S), S \longrightarrow (S - S), S \longrightarrow (S \cdot S), S \longrightarrow (S : S), \\ &\qquad S \longrightarrow x, S \longrightarrow y, S \longrightarrow z\}. \end{aligned}$$

$L(G_7)$ besteht aus allen exakt geklammerten arithmetischen Ausdrücken mit den Variablen x, y, z

Regelgrammatik – Beispiele IV

$G_8 = (\{A, I, J\}, T, P, A)$ mit

$$
\begin{aligned}
T &= \{S, P, \mathbf{LOOP}, \mathbf{WHILE}, \mathbf{BEGIN}, \mathbf{END}, :=, \neq, ;\, , (,) \\
&\qquad 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, x, [,]\}, \\
P &= \{A \to x[I] := 0,\ A \to x[I] := x[I],\ A \to x[I] := S(x[I]), \\
&\qquad A \to x[I] := P(x[I]),\ A \to A; A,\ A \to \mathbf{LOOP}\ x[I]\ \mathbf{BEGIN}\ A\ \mathbf{END}, \\
&\qquad A \to \mathbf{WHILE}\ x[I] \neq 0\ \mathbf{BEGIN}\ A\ \mathbf{END}\} \\
&\qquad \cup \{I \to Jx,\ J \to Jx \mid x \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}\} \\
&\qquad \cup \{I \to x,\ J \to x \mid x \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}\}
\end{aligned}
$$

$L(G_8)$ besteht aus allen **LOOP**/**WHILE**-Programmen

Typen von Regelgrammatiken

Definition: Sei $G = (N, T, P, S)$ eine Regelgrammatik. Wir sagen

a) G ist monoton, wenn für alle Regeln $\alpha \longrightarrow \beta \in P$ die Bedingung $|\alpha| \leq |\beta|$ erfüllt ist, wobei als Ausnahme $S \longrightarrow \lambda$ zugelassen ist, wenn $|\beta'|_S = 0$ für alle Regeln $\alpha' \longrightarrow \beta' \in P$ gilt,

b) G ist kontextabhängig, wenn alle Regeln in P von der Form $uAv \longrightarrow uwv$ mit $u, v \in V^*$, $A \in N$ und $w \in V^+$ sind, wobei als Ausnahme $S \longrightarrow \lambda$ zugelassen ist, wenn $|\beta'|_S = 0$ für alle Regeln $\alpha' \longrightarrow \beta' \in P$ gilt,

c) G ist kontextfrei, wenn alle Regeln in P von der Form $A \longrightarrow w$ mit $A \in N$ und $w \in V^*$ sind,

d) G ist regulär, wenn alle Regeln in P von der Form $A \longrightarrow wB$ oder $A \longrightarrow w$ mit $A, B \in N$ und $w \in T^*$ sind.

Typen von Sprachen

Definition: Eine Sprache L heißt <u>monoton</u> (bzw. <u>kontextabhängig</u>, <u>kontextfrei</u> oder <u>regulär</u>), wenn es eine monotone (bzw. kontextabhängige, kontextfreie oder reguläre) Grammatik G mit $L = L(G)$ gibt.

$\mathcal{L}(REG)$	–	Menge der regulären Sprachen
$\mathcal{L}(CF)$	–	Menge der kontextfreien Sprachen
$\mathcal{L}(CS)$	–	Menge der kontextabhängigen Sprachen
$\mathcal{L}(MON)$	–	Menge der monotonen Sprachen
$\mathcal{L}(RE)$	–	Menge der von Regelgrammatiken erzeugbaren Sprachen

Lemma:
$\mathcal{L}(CS) \subseteq \mathcal{L}(MON) \subseteq \mathcal{L}(RE)$ und $\mathcal{L}(REG) \subseteq \mathcal{L}(CF) \subseteq \mathcal{L}(RE)$

Normalformen I

Lemma:
Zu jeder Regelgrammatik $G = (N, T, P, S)$ kann eine Regelgrammatik $G' = (N', T, P', S)$ so konstruiert werden, dass alle Regeln aus P' von der Form

$$\alpha \longrightarrow \beta \text{ oder } A \longrightarrow a \text{ mit } \alpha, \beta \in (N')^*,\ A \in N',\ a \in T$$

sind und $L(G) = L(G')$ gilt. Ist außerdem G eine monotone, kontextabhängige bzw. kontextfreie Grammatik, so ist auch G' monoton, kontextabhängig bzw. kontextfrei.

Normalformen II

Satz:
Zu jeder monotonen Grammatik $G = (N, T, P, S)$ kann eine monotone Grammatik $G' = (N', T, P', S')$ so konstruiert werden, dass jede Regel aus P' von einer der Formen

$$A \longrightarrow BC,\ A \longrightarrow B,\ AB \longrightarrow CB,\ AB \longrightarrow AC,\ B \longrightarrow a \text{ oder } S' \longrightarrow \lambda$$

mit $A \in N',\ B, C \in N' \setminus \{S'\},\ a \in T$ ist und $L(G) = L(G')$ gilt.

Folgerung: $\mathcal{L}(MON) = \mathcal{L}(CS)$.

Normalformen III

Lemma:
Zu jeder kontextfreien Grammatik $G = (N, T, P, S)$ existiert eine kontextfreie Grammatik $G' = (N', T, P', S)$ derart, dass
i) P' keine Regel der Form $A \longrightarrow \lambda$ mit $A \neq S$ enthält,
ii) $|w|_S = 0$ für alle Regeln $A \longrightarrow w \in P'$ gilt, und
iii) $L(G) = L(G')$ ist.

Folgerung: $\mathcal{L}(CF) \subseteq \mathcal{L}(MON)$.

$(\mathcal{L}(REG) \subseteq \mathcal{L}(CF) \subseteq \mathcal{L}(CS) = \mathcal{L}(MON) \subseteq \mathcal{L}(RE))$

Normalformen IV

Lemma:
Zu jeder kontextfreien Grammatik $G = (N, T, P, S)$ kann eine kontextfreie Grammatik $G' = (N, T, P', S)$ so konstruiert werden, dass P' keine Regel der Form $A \longrightarrow B$ mit $A, B \in N$ enthält und $L(G) = L(G')$ gilt.

Satz (CHOMSKY-Normalform):
Zu jeder kontextfreien Grammatik $G = (N, T, P, S)$ kann eine kontextfreie Grammatik $G' = (N', T, P', S)$ so konstruiert werden, dass P' nur Regeln der Form

$$A \longrightarrow BC \quad \text{und} \quad A \longrightarrow a \quad \text{mit} \quad A, B, C \in N',\ a \in T$$

enthält, wobei $S \longrightarrow \lambda$ als Ausnahme zugelassen ist, falls S in keiner rechten Seite einer Regel aus P' vorkommt, und $L(G) = L(G')$ gilt.

Normalformen V

Satz:
Zu jeder regulären Grammatik $G = (N, T, P, S)$ kann eine reguläre Grammatik $G' = (N', T, P', S)$ so konstruiert werden, dass P' nur Regeln der Form

$$A \longrightarrow aB \quad \text{und} \quad A \longrightarrow a \quad \text{mit} \quad A, B \in N',\ a \in T$$

enthält, wobei $S \longrightarrow \lambda$ als Ausnahme zugelassen ist, falls P' keine Regel der Form $A \longrightarrow aS$ enthält, und $L(G) = L(G')$ gilt.

Schleifensätze I

Satz (Schleifensatz / Pumping-Lemma für reguläre Sprachen):
Sei L eine reguläre Sprache. Dann gibt es eine (von L abhängige) Konstante k derart, dass es zu jedem Wort $z \in L$ mit $|z| \geq k$ Wörter u, v, w gibt, die den folgenden Eigenschaften genügen:

i) $z = uvw$,
ii) $|uv| \leq k$, $|v| > 0$, und
iii) $uv^iw \in L$ für alle $i \geq 0$.

Lemma: $L = \{a^nb^n : n \geq 1\} \in \mathcal{L}(CF) \setminus \mathcal{L}(REG)$.

Schleifensätze II

Satz (Schleifensatz / Pumping-Lemma für kontextfreie Sprachen):
Sei L eine kontextfreie Sprache. Dann gibt es eine (von L abhängige) Konstante k derart, dass es zu jedem Wort $z \in L$ mit $|z| \geq k$ Wörter u, v, w, x, y gibt, die folgenden Eigenschaften genügen:
i) $z = uvwxy$,
ii) $|vwx| \leq k$, $|v| + |x| > 0$, und
iii) $uv^iwx^iy \in L$ für alle $i \geq 0$.

Lemma: $L = \{a^nb^nc^n : n \geq 1\} \in \mathcal{L}(MON) \setminus \mathcal{L}(CF)$.

Satz: $\mathcal{L}(REG) \subset \mathcal{L}(CF) \subset \mathcal{L}(CS) = \mathcal{L}(MON) \subseteq \mathcal{L}(REG)$.