

Automatentheorie und formale Sprachen

Pumpinglemma für kontextfreie Sprachen

Dozentin: Wiebke Petersen

30.6.2009

context-free grammars

Definition

A grammar (N, T, S, P) is *context-free* if all production rules are of the form:

$$A \rightarrow \alpha, \text{ with } A \in N \text{ and } \alpha \in (T \cup N)^*.$$

A language generated by a context-free grammar is said to be context-free.

Chomsky Normal Form

Definition

A grammar is in *Chomsky Normal Form (CNF)* if all production rules are of the form

- 1 $A \rightarrow a$
- 2 $A \rightarrow BC$

with $A, B, C \in T$ and $a \in \Sigma$ (and if necessary $S \rightarrow \epsilon$ in which case S may not occur in any right-hand side of a rule).

Theorem

Each context-free language is generated by a grammar in CNF.

Each context-free language is generated by a grammar in CNF

3 steps

- 1 Adapt the grammar such that terminals only occur in rules of type $A \rightarrow a$.
- 2 Eliminate $A \rightarrow B$ rules.
- 3 Eliminate $A \rightarrow B_1 B_2 \dots B_n$ ($n > 2$) rules.

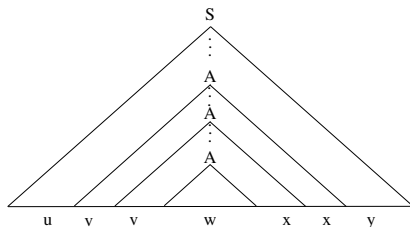
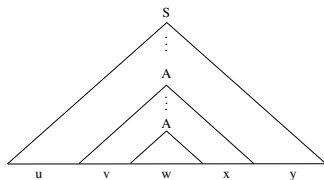
Pumping lemma for context-free languages

pumping lemma

For each context-free language L there exists a $p \in \mathbb{N}$ such that for any $z \in L$: if $|z| > p$, then z may be written as $z = uvwxy$ with

- $u, v, w, x, y \in T^*$,
- $|vwx| \leq p$,
- $vx \neq \epsilon$ and
- $uv^iwx^iy \in L$ for any $i \geq 0$.

Pumping lemma: proof sketch



$|vwx| \leq p$, $vx \neq \epsilon$ and $uv^i wx^i y \in L$ for any $i \geq 0$.

Existence of non context-free languages

- $L_1 = \{a^n b^n c^n\}$
- $L_2 = \{a^n b^m c^n d^m\}$
- $L_3 = \{ww : w \in \{a, b\}^*\}$

Theorem

Context-free languages are closed under intersection with a regular language, i.e., if L_{CF} is a context-free language and L_{reg} is a regular language, then $L_{CF} \cap L_{reg}$ is a context-free language.

Sind natürliche Sprachen kontextfrei?

Nebensatzeinbettung im Schweizerdeutschen

- Jan säit das
mer d'chind em Hans es huus lönd hälfe aastriche
wir die Kinder-AKK Hans-DAT das Haus-AKK ließen helfen anstreichen

NP_1 NP_2 NP_3 VP_1 VP_2 VP_3 "cross serial dependencies"



- *mer d'chind de Hans es huus lönd hälfe aastriche
wir die Kinder-AKK Hans-AKK das Haus-AKK ließen helfen anstreichen

Nebensatzeinbettung im Deutschen

- weil er die Kinder dem Hans das Haus streichen helfen ließ

NP_1 NP_2 NP_3 VP_3 VP_2 VP_1 "nested dependencies"



NL $\not\subseteq$ CF: Beweis Shieber 1985

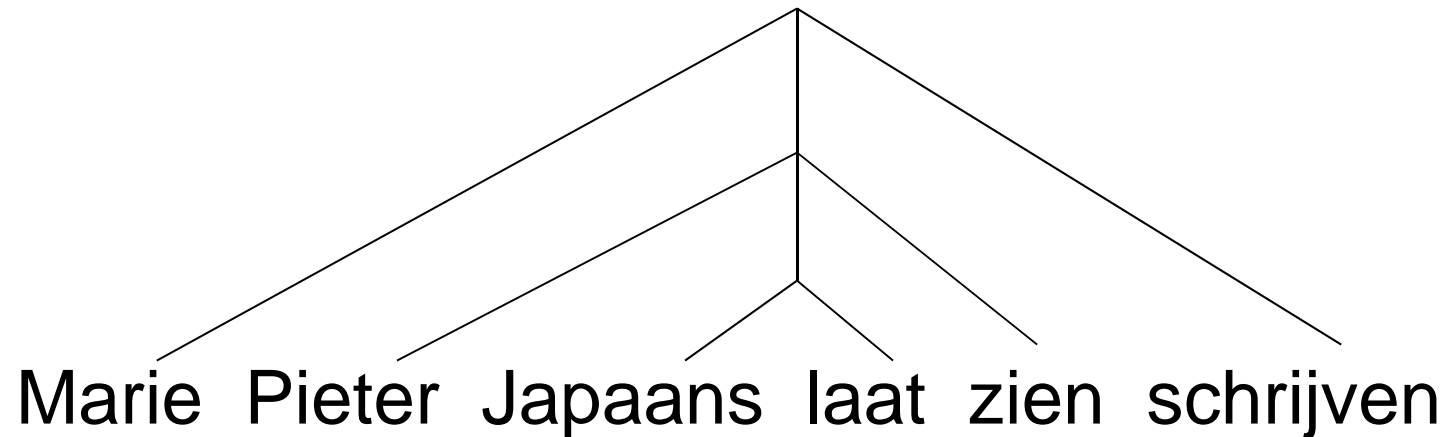
Homomorphismus:

$f(\text{"l\"ond"}) = c$	$f(\text{"es huus"}) = x$
$f(\text{"d'chind"}) = a$	$f(\text{"Jan s\"ait das mer"}) = w$
$f(\text{"em Hans"}) = b$	$f(s) = z \text{ otherwise}$
$f(\text{"h\"alfe"}) = d$	
$f(\text{"aastriiche"}) = y$	

- $f(\text{Schweizerdeutsch}) \cap wa^*b^*xc^*d^*y = wa^mb^nc^md^ny$
- $wa^mb^nc^md^ny$ ist nicht kontextfrei (\rightarrow Pumping Lemma)
- $wa^*b^*xc^*d^*y$ ist regulär
- kontextfreie Sprachen sind abgeschlossen unter
 - Homomorphismen
 - Schnitt mit regulären Sprachen
- Das Schweizerdeutsch ist nicht kontextfrei

Dutch (Huybregts 1976)

dat Jan [Marie Pieter Japaans laat zien schrijven]
dass Jan Marie Pieter Japanisch schreiben
sehen lässt



starke und schwache generative Kapazität

- Die **schwache generative Kapazität** eines linguistischen Formalismus ist die Eigenschaft alle grammatischen Sätze einer Sprache zu generieren
- Die **starke generative Kapazität** eines linguistischen Formalismus ist die Fähigkeit allen grammatischen Sätzen *ihre Struktur* zuzuweisen
- CFG's ????

Hausaufgaben

- 1 Zeigen sie, daß die folgende Aussage nicht wahr ist:
 - Die Schnittmenge zweier kontextfreier Sprachen ist kontextfrei (Beweis über Angabe eines Gegenbeispiels).
- 2 Überlegen sie sich je eine Grammatik zu zwei kontextfreien Beispielsprachen L_1 und L_2 . Bilden sie aus diesen Grammatiken kontextfreie Grammatiken für die Sprachen
 - $L_1 \cup L_2$
 - L_1^*
 - $L_1 \circ L_2$
- 3 Formulieren sie in ihren eigenen Worten den Beweis, dass das Schweizerdeutsche nicht kontext-frei ist.
- 4 Teilen sie mir bitte bis Ende der Woche mit, mit welchen Hausaufgaben sie bisher die größten Schwierigkeiten hatten, damit ich eine Übungssitzung vorbereiten kann.
- 5 Vergessen sie nicht den Vortrag von Marcus Kracht am Freitag um 14 Uhr.

(Gruppenarbeit ist ausdrücklich erwünscht)