Computerlinguistik I

Vorlesung im WiSe 2018/19
(M-GSW-09)

Prof. Dr. Udo Hahn

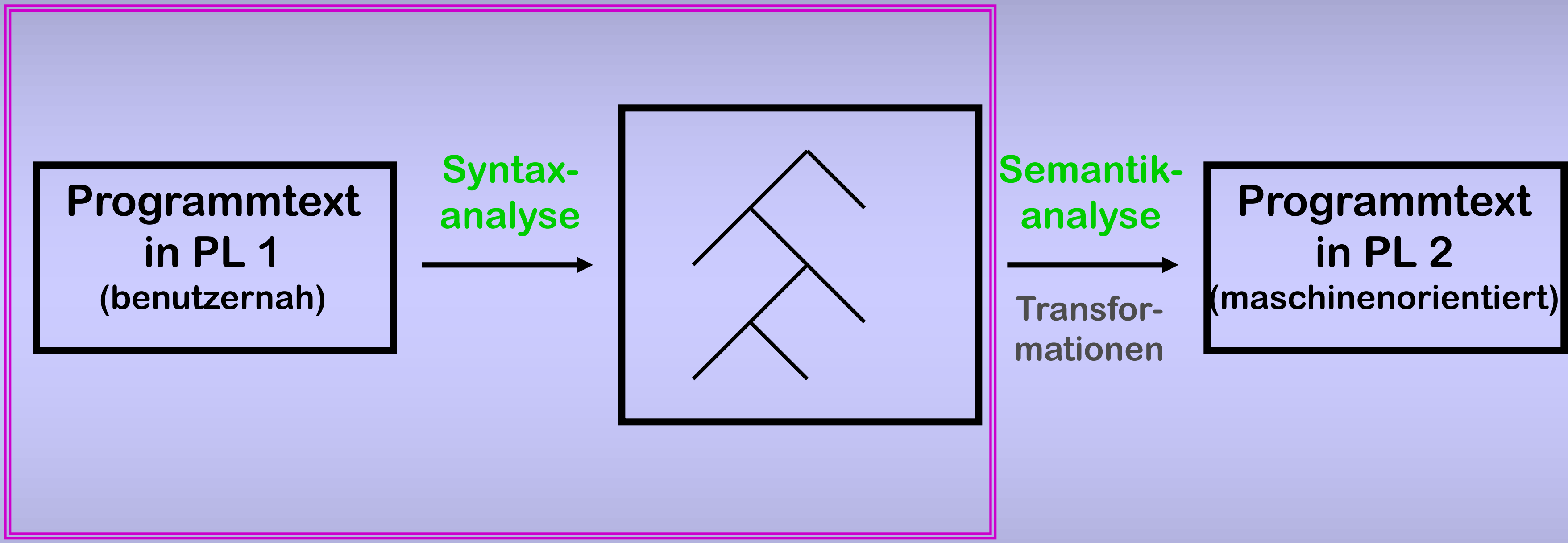
Lehrstuhl für Computerlinguistik
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

http://www.julielab.de

Syntaxanalyse

- Formale Analyse von Ausdrücken einer Sprache
  - Computerlinguistik
    - Formale Analyse von Wörtern oder Sätzen einer natürlichen Sprache (z.B. des Deutschen)
  - Informatik
    - Formale Analyse von Ausdrücken einer forma-len Sprache (z.B. einer Programmiersprache)

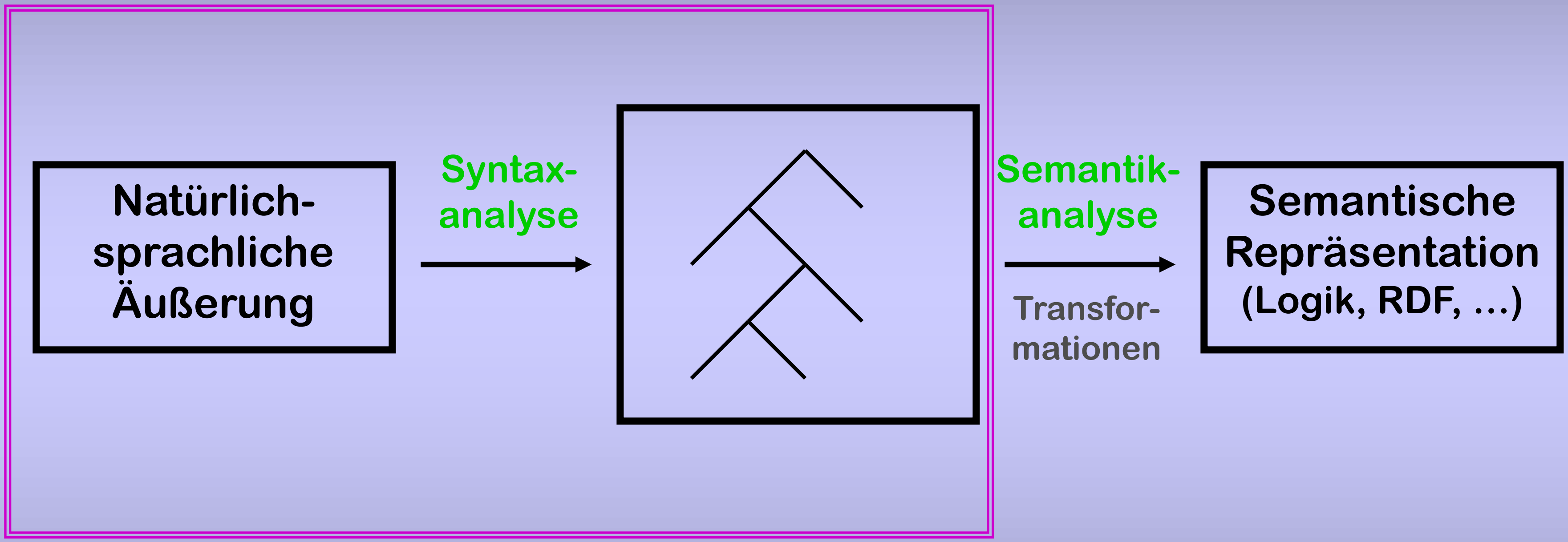
Analyse von Programmen



Aufgaben der Syntaxanalyse:

1. Syntaktisch korrekte Programme werden als korrekt erkannt
2. Syntaktisch unkorrekte Programme werden zurück gewiesen: Fehlererkennung und -diagnose

Analyse von natürlichsprachlichen Äußerungen



Aufgaben der Syntaxanalyse:

1. Syntaktisch korrekte Äußerungen werden als korrekt erkannt
2. Syntaktisch unkorrekte Äußerungen werden zurück gewiesen:

   Aber: Robustheit im Umgang mit paragrammatischen Äußerungen ist wünschenswert !

Beziehung zwischen Informatik und Computerlinguistik

- Informatik besitzt umfangreichen Methodenfundus
  - präzise beschriebene Analyseverfahren
  - Charakterisierung der formalen Eigenschaften dieser Verfahren (Entscheidbarkeit, Berechnungskomplexität)
  - mathematische Beschreibung der „Hintergrundtheorie“ (formale Grammatiken, formale Sprachen, Automaten)
- Übernahme und Adaption an NL in CL

Grundbegriffe zu formalen Sprachen (Wh.)

- **Potenzen** von **Wörtern** ω bzw. **Wort-mengen** $\mathcal{F}$ sind gegeben durch:
  - $\omega^0 := \varepsilon$ $\quad \omega^1 := \omega$ $\quad \omega^i := \omega^{i-1}\,\omega$, für $i \geq 1$
  - $\mathcal{F}^0 := \{\varepsilon\}$ $\quad \mathcal{F}^1 := \mathcal{F}$ $\quad \mathcal{F}^i := \mathcal{F}^{i-1}\,\mathcal{F}$, für $i \geq 1$

Beispiele für Potenzen von Wortmengen:

- DSilbe := {ba, di, ko}
- $\text{DSilbe}^0 = \{\varepsilon\}$, $\text{DSilbe}^1 = \{$ba, di, ko$\}$
- $\text{DSilbe}^2 = \text{DSilbe}^1$ DSilbe

  = { baba, badi, bako, diba, didi, diko , koba, kodi, koko}
- $\text{DSilbe}^3 = \text{DSilbe}^2$ DSilbe

Grundbegriffe zu formalen Sprachen (Wh.)

- Die Plushülle bzw. Sternhülle einer Wortmenge $\mathcal{F}$ werden definiert durch:

$$\mathcal{F}^{+} := \bigcup_{i \geq 1} \mathcal{F}^{i} \qquad \mathcal{F}^{*} := \bigcup_{i \geq 0} \mathcal{F}^{i}$$

- Ist Σ ein Alphabet, dann ist Σ^* die Gesamtheit aller Wörter über Σ. Jede Teilmenge dieser Sternhülle, $\mathcal{L} \subseteq \Sigma^*$, heißt formale Sprache über Σ .

Grundbegriffe zu formalen Grammatiken

- Eine **formale Grammatik** $\mathcal{G}$ ist ein 4-tupel

$$\mathcal{G} = (N, T, P, S)$$

mit

  - N: das Alphabet der **Nicht-Terminalsymbole**
  - T: das Alphabet der **Terminalsymbole**
  - P: eine endliche Menge von **Produktionen** der Form

    $\alpha \rightarrow \gamma$ (gesprochen: *„α produziert γ“*) mit

    $\alpha \in (N \cup T)^* \, N \, (N \cup T)^*$ und

    $\gamma \in (N \cup T)^*$
  - S: das **Startsymbol**, $S \in N$

$\mathcal{V} = N \cup T$, bezeichnet das **Gesamtalphabet** ($N \cap T = \emptyset$)

Beziehung zwischen formalen Grammatiken & formalen Sprachen

- Eine formale Grammatik $\mathcal{G}$ **erzeugt** eine formale Sprache. Der Erzeugungsprozess ist festgelegt durch eine auf $\mathcal{V}^*$ definierte Relation „$\Rightarrow$“ (gesprochen: *„ist direkt ableitbar nach“*). Für u, v, $\gamma \in \mathcal{V}^*$ und $\alpha \in \mathcal{V}^*$ N $\mathcal{V}^*$ gilt:

u α v $\Rightarrow$ u γ v genau dann, wenn $\alpha \rightarrow \gamma \in$ P

Grundbegriffe zu formalen Grammatiken

- Die transitive Hülle der Relation „$\Rightarrow$“ schreibt man

  $\overset{+}{\Rightarrow}$ (gesprochen: „ist nichttrivial ableitbar nach“)

- Die reflexive und transitive Hülle der Relation „$\Rightarrow$“ schreibt man

  $\overset{*}{\Rightarrow}$ (gesprochen: „ist ableitbar nach“)

Grundbegriffe zu formalen Grammatiken

- **Man schreibt $s \stackrel{n}{\Rightarrow} z$, um auszudrücken, dass $s_0, s_1, \ldots, s_n$ existieren mit**

  $s = s_0$, $s_i \Rightarrow s_{i+1}$ **für** $0 \leq i < n$ **und** $s_n = z$

- **Damit ist also**

  $s \stackrel{+}{\Rightarrow} z$ **g.d.w.** $s \stackrel{n}{\Rightarrow} z$ **für ein** $n \geq 1$ **und**

  $s \stackrel{*}{\Rightarrow} z$ **g.d.w. entweder** $s = z$ **oder** $s \stackrel{+}{\Rightarrow} z$

Grundbegriffe zu formalen Grammatiken

- Die von der formalen Grammatik $\mathcal{G}$ = (N, T, P, S) erzeugte formale Sprache $\mathcal{L}(\mathcal{G})$ ist wie folgt definiert:

$$\mathcal{L}(\mathcal{G}) := \{ \tau \mid \tau \in T^*, S \overset{*}{\Rightarrow} \tau, \text{ S ist Startsymbol von } \mathcal{G} \}$$

τ heißt auch Wort der Sprache $\mathcal{L}(\mathcal{G})$.

Grundbegriffe zu formalen Grammatiken

- Zwei formale Grammatiken $\mathcal{G}_1$ und $\mathcal{G}_2$, $\mathcal{G}_1 \neq \mathcal{G}_2$, heißen äquivalent, wenn sie dieselbe Sprache erzeugen, d.h.:

$$\mathcal{L}(\mathcal{G}_1) = \mathcal{L}(\mathcal{G}_2)$$

Grundbegriffe zu formalen Grammatiken

- Abhängig von der Form der zugelassenen Produktionen definiert man vier Typen von formalen Grammatiken:
  - Eine Grammatik $\mathcal{G}$ heißt Typ-0-Grammatik, wenn die Gestalt der Produktionen nicht weiter eingeschränkt ist. D.h., sie haben die Form

$$\alpha \longrightarrow \gamma$$

mit

$$\alpha \in (N \cup T)^* \, N \, (N \cup T)^* \text{ und}$$
$$\gamma \in (N \cup T)^*$$

Grundbegriffe zu formalen Grammatiken

– Eine Grammatik $\mathcal{G}$ heißt Typ-1-Grammatik (kontextsensitive Grammatik), wenn P nur Produktionen der Gestalt

$$\alpha \rightarrow \gamma$$

mit

$$\alpha \in (N \cup T)^* \, N \, (N \cup T)^* \text{ und}$$
$$\gamma \in (N \cup T)^*$$
$$|\alpha| \leq |\gamma|$$

(sog. *non-shrinking rules*) und eventuell die Produktion $S \rightarrow \varepsilon$ enthält (wobei letztere nur zugelassen ist, wenn das Startsymbol S in keiner Produktion auf der rechten Seite auftritt).

Grundbegriffe zu formalen Grammatiken

– Eine Grammatik $\mathcal{G}$ heißt **Typ-2-Grammatik** (**kontextfreie Grammatik**), wenn P nur Produktionen enthält der Gestalt

$$A \rightarrow \gamma \quad \text{mit } A \in N \text{ und } \gamma \in (N \cup T)^*$$

Grundbegriffe zu formalen Grammatiken

- Eine Grammatik $\mathcal{G}$ heißt Typ-3-Grammatik (reguläre Grammatik), wenn P nur Produktionen der Gestalt

  $$A \longrightarrow \gamma \quad \text{mit } A \in N \text{ und } \gamma \in N\,T^* \cup T^*$$

  (sog. linkslineare Produktionen) oder nur Produktionen der Gestalt

  $$A \longrightarrow \gamma \quad \text{mit } A \in N \text{ und } \gamma \in T^*\,N \cup T^*$$

  (sog. rechtslineare Produktionen) enthält.
- Man spricht dann auch entsprechend von linkslinearen bzw. rechtslinearen Grammatiken.
- Eine reguläre Grammatik darf nicht Regeln nach beiden Produktionsregelmustern mischen.

Beispiel einer rechtslinearen Grammatik

$\mathcal{G}$-3 = (N, T, P, S) mit

N = { S, A, B }

T = { a, b }

P = { S $\longrightarrow$ aA,

A $\longrightarrow$ aA,

A $\longrightarrow$ bbB,

B $\longrightarrow$ bB,

B $\longrightarrow$ b }

$\mathcal{L}(\mathcal{G}$-3) = { abbb, aabbb, aaabbb, aaaabbb, abbbb, … }

= $a^n b^m$, $n \geq 1, m \geq 3$

Beispiel einer rechtslinearen Grammatik

$\mathcal{G}\text{-}3 = (N, T, P, S)$ mit

$N = \{ S, A, B \}$

$T = \{ a, b \}$

$P = \{ S \longrightarrow aA,$
$A \longrightarrow aA,$
$A \longrightarrow bbB,$
$B \longrightarrow bB,$
$B \longrightarrow b \}$

$S \Rightarrow aA$	mit	$S \rightarrow aA \in P$
$aA \Rightarrow abbB$	mit	$A \rightarrow bbB \in P$
$abbB \Rightarrow abbb$	mit	$B \rightarrow b \in P$

$\mathcal{L}(\mathcal{G}\text{-}3) = \{ abbb, aabbb, aaabbb, aaaabbb, abbbb, \ldots \}$

$= a^n b^m$, $n,m \geq 1$

Beispiel einer kontextfreien Grammatik

$\mathcal{G}$-2 = (N, T, P, S) mit

N = { S }

T = { a, b }

P = { $S \longrightarrow aSb$,

$S \longrightarrow ab$ }

$\mathcal{L}(\mathcal{G}\text{-2})$ = {ab, aabb, aaabbb, aaaabbbb, … }

= $a^n b^n$, $n \geq 1$

Beispiel einer kontextfreien Grammatik

$\mathcal{G}$-2 = (N, T, P, S) mit

N = { S }

T = { a, b }

P = { S $\longrightarrow$ aSb,

S $\longrightarrow$ ab }

$\mathcal{L}(\mathcal{G}\text{-2})$ = {ab, aabb, aaabbb, aaaabbbb, ... }

= $a^n b^n$, $n \geq 1$

S $\Rightarrow$ aSb mit S $\rightarrow$ aSb $\in$ P

aSb $\Rightarrow$ aabb mit S $\rightarrow$ ab $\in$ P

Beispiel einer kontextfreien Grammatik

$\mathcal{G}$-2 = (N, T, P, S) mit

N = { S }

T = { a, b }

P = { $S \longrightarrow aSb$,

$S \longrightarrow ab$ }

$\mathcal{L}(\mathcal{G}\text{-}2)$ = {ab, aabb, aaabbb, aaaabbbb, ... }

= $a^n b^n$, $n \geq 1$

$S \overset{4}{\Rightarrow} aaaabbbb$ $S \overset{*}{\Rightarrow} aaaabbbb$

$\underset{s_0}{\boxed{S}} \Rightarrow \underset{s_1}{\boxed{aSb}} \Rightarrow \underset{s_2}{\boxed{aaSbb}} \Rightarrow \underset{s_3}{\boxed{aaaSbbb}} \Rightarrow \underset{s_4}{\boxed{aaaabbbb}}$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$,

$bB \longrightarrow bb$,

$C \longrightarrow c$ }

$\mathcal{L}(\mathcal{G}$-1) = { abc, aabbcc, aaabbbccc, ... }

= $a^n b^n c^n$, $n \geq 1$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \overset{*}{\Rightarrow} aaabbbccc$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \Rightarrow aSBC$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \Rightarrow aSBC \Rightarrow aaSBCBC$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { S $\longrightarrow$ aSBC, S $\longrightarrow$ aBC,

CB $\longrightarrow$ XB, XB $\longrightarrow$ XC, XC $\longrightarrow$ BC,

aB $\longrightarrow$ ab, bB $\longrightarrow$ bb, C $\longrightarrow$ c }

S $\Rightarrow$ aSBC $\Rightarrow$ aaSBCBC $\Rightarrow$ aaaBCBCBC

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \Rightarrow aSBC \Rightarrow aaSBCBC \Rightarrow aaaBCBCBC \Rightarrow$
$\ldots \Rightarrow aaaBBCCBC$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \Rightarrow aSBC \Rightarrow aaSBCBC \Rightarrow aaaBCBCBC \Rightarrow \ldots \Rightarrow aaaBBCCBC \Rightarrow \ldots \Rightarrow aaaBBCBCC$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { S $\longrightarrow$ aSBC, S $\longrightarrow$ aBC,

CB $\longrightarrow$ XB, XB $\longrightarrow$ XC, XC $\longrightarrow$ BC,

aB $\longrightarrow$ ab, bB $\longrightarrow$ bb, C $\longrightarrow$ c }

S $\Rightarrow$ aSBC $\Rightarrow$ aaSBCBC $\Rightarrow$ aaaBCBCBC $\Rightarrow$
… $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCBCC $\Rightarrow$
… $\Rightarrow$ aaaBBBCCC

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { S $\longrightarrow$ aSBC, S $\longrightarrow$ aBC,

CB $\longrightarrow$ XB, XB $\longrightarrow$ XC, XC $\longrightarrow$ BC,

aB $\longrightarrow$ ab, bB $\longrightarrow$ bb, C $\longrightarrow$ c }

S $\Rightarrow$ aSBC $\Rightarrow$ aaSBCBC $\Rightarrow$ aaaBCBCBC $\Rightarrow$
… $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCCBC $\Rightarrow$
… $\Rightarrow$ aaaBBBCCC $\Rightarrow$ aaabBBCCC

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { S $\longrightarrow$ aSBC, S $\longrightarrow$ aBC,

CB $\longrightarrow$ XB, XB $\longrightarrow$ XC, XC $\longrightarrow$ BC,

aB $\longrightarrow$ ab, bB $\longrightarrow$ bb, C $\longrightarrow$ c }

S $\Rightarrow$ aSBC $\Rightarrow$ aaSBCBC $\Rightarrow$ aaaBCBCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBBCCC $\Rightarrow$ aaabBBCCC $\Rightarrow$ aaabbBCCC

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

$S \Rightarrow aSBC \Rightarrow aaSBCBC \Rightarrow aaaBCBCBC \Rightarrow$
$\ldots \Rightarrow aaaBBCCBC \Rightarrow \ldots \Rightarrow aaaBBCCBC \Rightarrow$
$\ldots \Rightarrow aaaBBBCCC \Rightarrow aaabBBCCC \Rightarrow$
$aaabbBCCC \Rightarrow aaabbbCCC$

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { S $\longrightarrow$ aSBC, S $\longrightarrow$ aBC,

CB $\longrightarrow$ XB, XB $\longrightarrow$ XC, XC $\longrightarrow$ BC,

aB $\longrightarrow$ ab, bB $\longrightarrow$ bb, C $\longrightarrow$ c }

S $\Rightarrow$ aSBC $\Rightarrow$ aaSBCBC $\Rightarrow$ aaaBCBCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBCCBC $\Rightarrow$ … $\Rightarrow$ aaaBBBCCC $\Rightarrow$ aaabBBCCC $\Rightarrow$ aaabbBCCC $\Rightarrow$ aaabbbCCC $\Rightarrow$ … $\Rightarrow$ aaabbbccc

Beispiel einer kontextsensitiven Grammatik

$\mathcal{G}$-1 = (N, T, P, S) mit

N = { S, B, C, X }

T = { a, b, c }

P = { $S \longrightarrow aSBC$, $S \longrightarrow aBC$,

$CB \longrightarrow XB$, $XB \longrightarrow XC$, $XC \longrightarrow BC$,

$aB \longrightarrow ab$, $bB \longrightarrow bb$, $C \longrightarrow c$ }

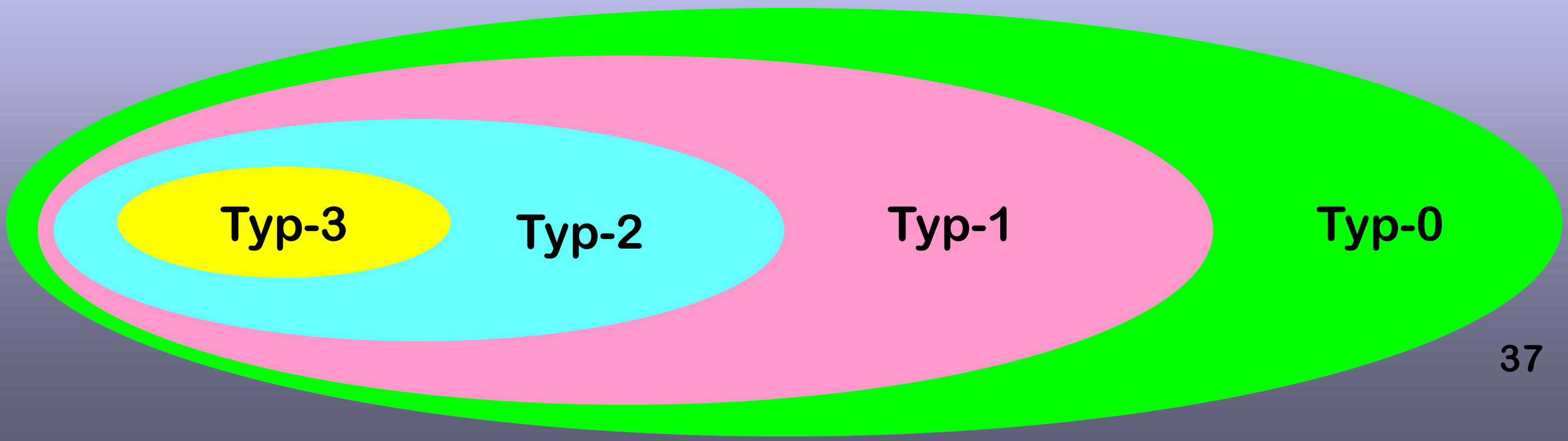
S ⇒ aSBC ⇒ aaSBCBC ⇒ aaaBCBCBC ⇒ … ⇒ aaaBBCCBC ⇒ … ⇒ aaaBBCCBC ⇒ … ⇒ aaaBBBCCC ⇒ aaabBBCCC ⇒ aaabbBCCC ⇒ aaabbbCCC ⇒ … aaabbbccc

Typen formaler Sprachen (und ihr Bezug zu Grammatik-Typen)

– Eine formale Sprache heißt vom Typ 0, 1, 2 oder 3, wenn sie von einer Grammatik des entsprechenden Typs erzeugt werden kann.
  - Eine Typ-1-Sprache heißt auch kontextsensitive Sprache.
  - Eine Typ-2-Sprache heißt auch kontextfreie Sprache.
  - Eine Typ-3-Sprache heißt auch reguläre Sprache.

Chomsky-Hierarchie formaler Sprachen

– Für jedes Alphabet Σ (mit mindestens zwei Symbolen) ist die Menge der Typ-i-Sprachen über Σ für $i = 0, 1, 2$ jeweils (echte) Obermenge der Typ-[i+1]-Sprachen über Σ.
Die damit gegebene Hierarchie von formalen Sprachen heißt Chomsky-Hierarchie.



Formale Eigenschaften natürlicher Sprachen

– Natürliche Sprachen werden im Folgenden als formale Sprachen (Mengen von Wörtern) betrachtet

– <u>Problemstellung</u>:

Welcher Typ formaler Sprachen charakterisiert natürliche Sprachen ?

Natürliche Sprachen als reguläre Sprachen

- **Natürliche Sprachen sind ausdrucksstärker als reguläre Sprachen (Typ-3)**
  - **Beweisbar durch Pumping-Theorem und die Tatsache, dass reguläre Sprachen unter Mengen-schnitt abgeschlossen sind ($L_{r1} \cap L_{r2} = L_{r3}$)**
  - **Illustration durch Sprachdaten**

Sind Natürliche Sprachen regulär ?

Satzbeispiele:

(1) The cat waited.

(2) The cat the dog admired waited.

(3) The cat the dog the ant bit admired waited.

Satzformel:

(„the“ N)^{n} (V$_{transitiv}$)$^{n-1}$ („waited“ [V$_{intransitiv}$])^{1}

bzw. $a^{n} b^{n-1} x$

Beweisidee (Abschluss unter Schnitt):

Englisch$_{reg}$ [als Menge von Ketten] $\cap$ $N^{*} V^{*} x = N^{n} V^{n-1} x$

Aber: $a^{n} b^{n}$ ist eine CFL (pumping lemma)!

Natürliche Sprachen als kontextfreie Sprachen

- <u>Einige</u> natürliche Sprachen sind (etwas) ausdrucksstärker als kontextfreie Sprachen
  - <u>Beweisbar</u> durch Pumping Theorem und die Tatsache, dass kontextfreie Sprachen unter Mengenschnitt mit regulären Sprachen abgeschlossen sind ($L_{cf1} \cap L_{r2} = L_{cf3}$)
  - Illustration durch Sprachdaten

Sind alle Natürlichen Sprachen kontextfrei ?

Satzbeispiele (cross-serial dependencies):

(1) Jan säit.

(2) Jan säit das mer em Hans hälfed.

(3) Jan säit das mer em Hans es huus hälfed aastriiche.

(4) Jan säit das mer d'chind em Hans es huus lönd hälfe aastriiche.

(5) Jan sagte, dass wir - die Kinder$_{AKK}$ – dem Hans$_{DAT}$ – das Haus lassen$_{AKK}$ - helfen$_{DAT}$ - anzustreichen

Satzformel:

Jan säit das mer („d'chind"$_{AKK}$)^n („em Hans"$_{DAT}$)^m es huus („lönd"$_{AKK}$)^n („hälfe"$_{DAT}$)^m aastriiche.

bzw. $w\ a^n\ b^m\ x\ c^n\ d^m\ y$

Beweisidee (Abschluss unter Schnitt):

SchweizerDeutsch$_{cf}$ [als Menge von Ketten] $\cap$ Jan säit das mer $(N_{AKK})^* (N_{DAT})^*$ es huus $(V_{AKK})^* (V_{DAT})^*$ aastriiche
= Jan säit das mer $(N_{AKK})^n (N_{DAT})^m$ es huus $(V_{AKK})^n (V_{DAT})^m$ aastriiche.

Aber: $w\ a^n\ b^m\ x\ c^n\ d^m\ y$ ist eine CSL (pumping lemma)!

Natürliche Sprachen als formale Sprachen

- NLs sind keine Typ-3-Sprachen
- NLs sind überwiegend (Englisch, Deutsch, Französisch, Spanisch, …) Typ-2-Sprachen
- Einige wenige NLs sind sicher (Schweizer Deutsch) bzw. vermutlich (Niederländisch, Bambara [Mali]) milde Typ-1-Sprachen