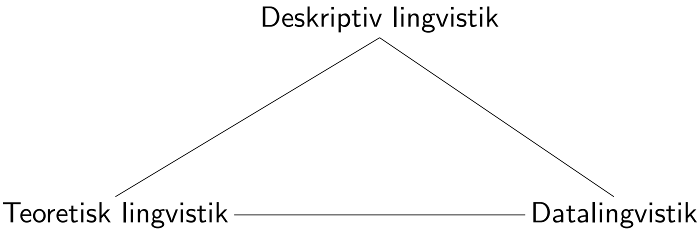
Grammatikteorier og mulige sprog

Anders Søgaard

Center for Language Technology
University of Copenhagen
Njalsgade 80
DK-2300 Copenhagen S
Email: anders@cst.dk

October 23 2007

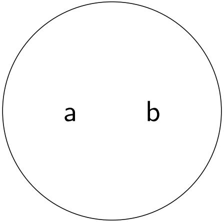


- Mild kontekstsensitivitet, og
- effektiv genkendelighed.

[FW06]: Et uendeligt hierarki af unifikationsbaserede grammatikker, der genererer Weirhierarkiet?

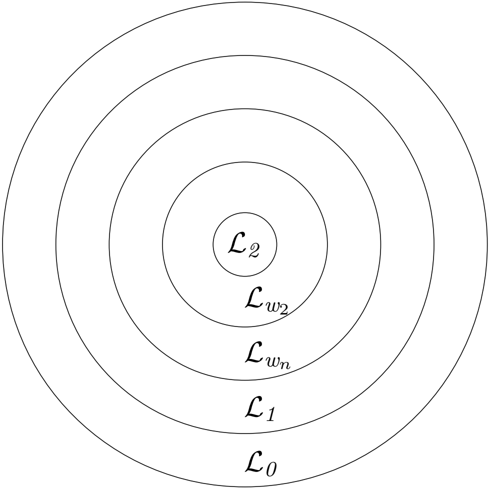
Teorier som værktøjskasser

Et sprog er en delmængde af mulige kombinationer af vokabolariet. Med hensyn til vokabolariet

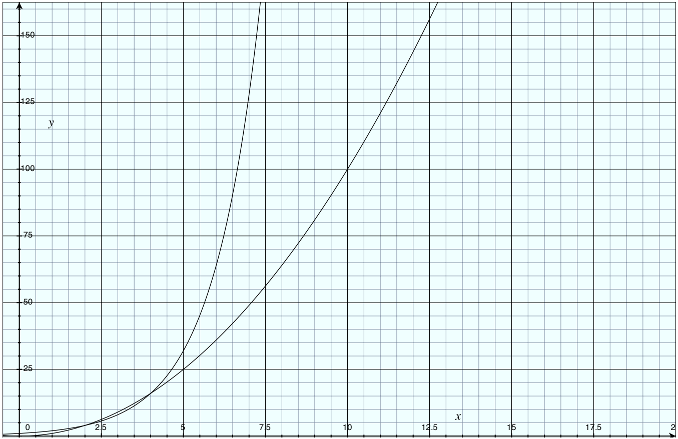


kan defineres en lang række sprog af varierende kompleksitet: $\{a\}, \{a^+\}, \{a^n b^n | n \geq 1\}, \ldots$

Chomsky- og Weirhierarkiet



Kompleksitet



Unifikationsbaserede grammatikker, kort fortalt

$$\begin{bmatrix} \text{CAT} & \mathit{np} \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{AGR}_1 & \mathit{agr}_1 \\ \vdots & \\ \text{AGR}_n & \mathit{agr}_n \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & \mathit{d} \\ \text{AGR} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & \mathit{n} \\ \text{AGR} & \boxed{1} \end{bmatrix}$$

- Kongruens,
- dependenser, and
- komposition.

Eksempler: funktionel grammatik, HPSG, konstruktionsgrammatik, LFG, PATR.

Theorem

Genkendelsesproblemet for unifikationsbaseret grammatik er uafgørbart.

Theorem

Unifikationsbaseret grammatik kan bruges til at beskrive alle type 0-sprog.

Definition

En unifikationsbaseret grammatik er off-line parsable, hvis alle ikkeforgrenende udvidelser er acykliske.

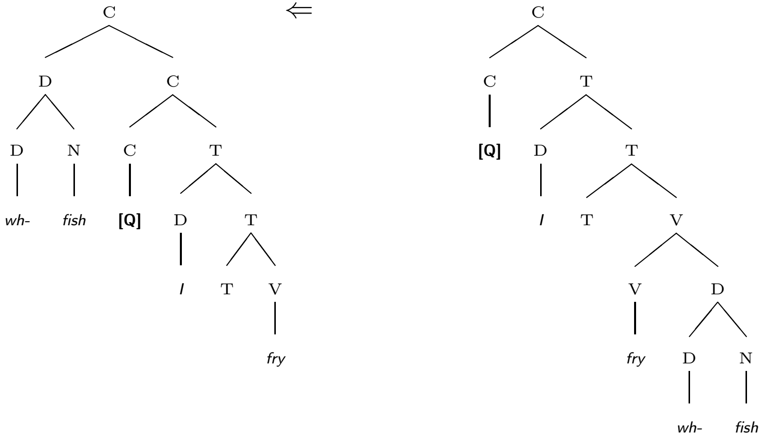
Theorem

Genkendelsesproblemet for off-line parsable unifikationsbaseret grammatik er NP-komplet.

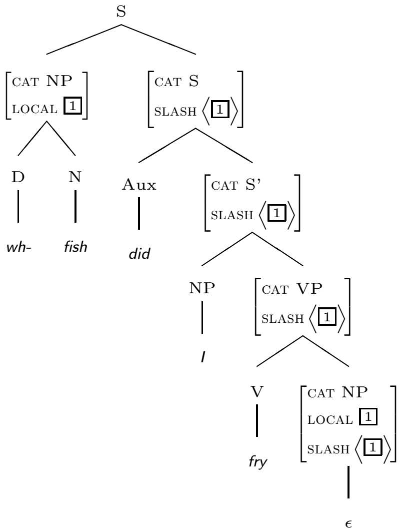
Theorem

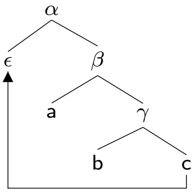
Off-line parsable unifikationsbaseret grammatik beskriver kun type 1-sprog, men herunder sprog udenfor Wierhierarkiet.

wh-movement (Minimalisme)



wh-movement (Unifikationsbaseret grammatik)





$\Longrightarrow$

$$\boxed{0}\begin{bmatrix}\text{DOWN} & \boxed{1}\end{bmatrix}$$

$$\boxed{0}\begin{bmatrix}\alpha & \\ \text{DOWN} & \begin{bmatrix}\epsilon & \\ \text{RIGHT} & \begin{bmatrix}\beta & \\ \text{DOWN} & \begin{bmatrix} a & \\ \text{RIGHT} & \begin{bmatrix}\gamma & \\ \text{DOWN} & \begin{bmatrix} b & \\ \text{RIGHT} & \boxed{1}\; c\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Theorem

Enhver analyse, som kan udtrykkes i Minimalisme, kan også udtrykkes i unifikationsbaseret grammatik.

Theorem

Ethvert sprog, som kan beskrives i Minimalisme, kan også beskrives i unifikationsbaseret grammatik.

Mildt kontekstsensitiv unifikationsbaseret grammatik?

Definition

Weirhierarkiet beskrives af k-degree multiple kontekstfrigrammatikker (k-MCFG) [SMFK91].

Example

Den kontekstfri grammatik med produktionsreglerne $S \rightarrow NP\ VP, NP \rightarrow John, VP \rightarrow snorker$ er ækvivalent med en 1-degree MCFG med produktionsreglerne:

$$\begin{array}{rcl} S(X_1X_2) & \rightarrow & NP(X_1)\,VP(X_2) \\ NP(John) & & VP(snorker) \end{array}$$

Example

En 2-MCFG med produktionsreglerne

$$\begin{array}{rcl} S(X_1X_2X_3X_4) & \rightarrow & A(X_1, X_3)B(X_2, X_4) \\ A(aX_1, cX_2) & \rightarrow & A(X_1, X_2) \\ B(bX_1, dX_2) & \rightarrow & B(X_1, X_2) \\ A(a, c) & & B(b, d) \end{array}$$

beskriver sproget $\{a^n b^m c^n d^m \mid m, n \geq 1\}$.

Non-deterministic copying

$$\begin{bmatrix} \text{CAT} & a_0 \\ \text{REST} & \begin{bmatrix} \text{ATTR}_1 & [\ldots] \\ \ldots & \\ \text{ATTR}_i & \boxed{1}\,[\ldots] \\ \ldots & \\ \text{ATTR}_j & \boxed{2}\,[\ldots] \\ \ldots & \\ \text{ATTR}_n & [\ldots] \end{bmatrix} \end{bmatrix} \rightarrow \ldots \begin{bmatrix} \text{CAT} & a_i \\ \text{REST} & \boxed{2} \end{bmatrix} \begin{bmatrix} \text{CAT} & a_j \\ \text{REST} & \boxed{1} \end{bmatrix} \ldots$$

Konsistens

En mængde produktionsregler er konsistente hvis og kun hvis de ikke inkluderer produktioner:

$$\begin{bmatrix}\text{REST.ATTR } \alpha\end{bmatrix} \rightarrow \ldots, \begin{bmatrix}\text{REST.ATTR } \boxed{1}\end{bmatrix} \rightarrow \ldots \begin{bmatrix}\ldots \boxed{1} \ldots\end{bmatrix} \ldots$$

eller

$$\begin{bmatrix}\text{REST.ATTR } \alpha\end{bmatrix} \rightarrow \ldots, \begin{bmatrix}\text{REST.ATTR } \beta\end{bmatrix} \rightarrow \ldots$$

hvor $\alpha \neq \beta$.

Deterministic copying og right-linearity

- En non-deterministic copying unifikationsbaseret grammatik er deterministic copying hvis og kun hvis alle A-produktionsregler er inkonsistente.
- En deterministic copying unifikationsbaseret grammatik er right-linear hvis og kun hvis der for enhver konsistent delmængde af regler, og ethvert træk ATTR og kategoriel værdi a findes højst én strukturdeling mellem ATTR og REST i en A-datter.

E.g. er flg. umuligt:

$$\begin{bmatrix} \text{CAT} & a_0 \\ \text{REST} & \begin{bmatrix} \text{ATTR} & \boxed{1} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & a_1 \\ \text{REST} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & a_1 \\ \text{REST} & \boxed{1} \end{bmatrix}$$

Example

En 2MCFG med produktionsreglerne

$$
\begin{array}{lrcl}
r_0 : & S(X_1 X_2 X_3 X_4) & \to & A(X_1, X_3) B(X_2, X_4) \\
r_1 : & A(aX_1, cX_2) & \to & A(X_1, X_2) \\
r_2 : & B(bX_1, dX_2) & \to & B(X_1, X_2) \\
r_3 : & A(a, c) & & \\
r_4 : & B(b, d) & &
\end{array}
$$

er ækvivalent med

$$
\begin{bmatrix} \text{CAT} & s \\ \text{REST} & \begin{bmatrix} \text{ATTR}_1 & \boxed{1} \\ \text{ATTR}_2 & \boxed{2} \\ \text{PROD} & r_0 \end{bmatrix} \end{bmatrix} \to \begin{bmatrix} \text{CAT} & a_1 \\ \text{REST} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & b_1 \\ \text{REST} & \boxed{2} \end{bmatrix} \begin{bmatrix} \text{CAT} & a_2 \\ \text{REST} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & b_2 \\ \text{REST} & \boxed{2} \end{bmatrix}
$$

$$
\begin{bmatrix} \text{CAT} & a_1 \\ \text{REST} & \begin{bmatrix} \text{ATTR}_1 & \boxed{1} \\ \text{PROD} & r_1 \end{bmatrix} \end{bmatrix} \to a \begin{bmatrix} \text{CAT} & a_1 \\ \text{REST} & \boxed{1} \end{bmatrix}, \quad \begin{bmatrix} \text{CAT} & a_2 \\ \text{REST} & \begin{bmatrix} \text{ATTR}_1 & \boxed{1} \\ \text{PROD} & r_1 \end{bmatrix} \end{bmatrix} \to c \begin{bmatrix} \text{CAT} & a_2 \\ \text{REST} & \boxed{1} \end{bmatrix}
$$

$$
\begin{bmatrix} \text{CAT} & b_1 \\ \text{REST} & \begin{bmatrix} \text{ATTR}_2 & \boxed{1} \\ \text{PROD} & r_2 \end{bmatrix} \end{bmatrix} \to b \begin{bmatrix} \text{CAT} & b_1 \\ \text{REST} & \boxed{1} \end{bmatrix}, \quad \begin{bmatrix} \text{CAT} & b_2 \\ \text{REST} & \begin{bmatrix} \text{ATTR}_2 & \boxed{1} \\ \text{PROD} & r_2 \end{bmatrix} \end{bmatrix} \to d \begin{bmatrix} \text{CAT} & b_2 \\ \text{REST} & \boxed{1} \end{bmatrix}
$$

$$
\begin{bmatrix} \text{CAT} & a_1 \\ \text{REST.PROD} & r_3 \end{bmatrix} \to a, \begin{bmatrix} \text{CAT} & a_2 \\ \text{REST.PROD} & r_3 \end{bmatrix} \to c, \begin{bmatrix} \text{CAT} & b_1 \\ \text{REST.PROD} & r_4 \end{bmatrix} \to b, \begin{bmatrix} \text{CAT} & b_2 \\ \text{REST.PROD} & r_4 \end{bmatrix} \to d
$$

Theorem

Hierarkiet af k-reentrant right-linear unifikationsbaserede grammatikker beskriver nøjagtigt Weirhierarkiet.

Noam Chomsky.
Aspects of the theory of syntax.
MIT Press, Cambridge, Massachusetts, 1965.

Daniel Feinstein and Shuly Wintner.
Highly constrained unification grammars.
In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 1089–1096, Sydney, Australia, 2006.

Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag.
Generalized phrase structure grammar.
Harvard University Press, Cambridge, Massachusetts, 1985.

Hiroyuki Seki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami.
On multiple context-free grammars.
Theoretical Computer Science, 88(2):191–229, 1991.