Kontekstfri grammatikker

Eksempel:

[Chomsky, 1956]

$$
\begin{aligned}
&\textit{sentence} \rightarrow \textit{subject verb object}\\
&\textit{subject} \rightarrow \textit{person}\\
&\textit{person} \rightarrow \texttt{Morten} \mid \texttt{Ole} \mid \texttt{Henrik}\\
&\textit{verb} \rightarrow \texttt{spurgte} \mid \texttt{sparkede}\\
&\textit{object} \rightarrow \textit{thing} \mid \textit{person}\\
&\textit{thing} \rightarrow \texttt{fodbolden} \mid \texttt{computeren}
\end{aligned}
$$

- **Nonterminal**-symboler:
  sentence, subject, person, verb, object, thing
- **Terminal**-symboler:
  `Morten`, `Ole`, `Henrik`, `spurgte`, `sparkede`, `fodbolden`, `computeren`
- **Start**-symbol: *sentence*
- Eksempel på **derivation**:
  sentence $\Rightarrow$ *subject verb object* $\Rightarrow$ ... $\Rightarrow$ `Ole spurgte computeren`

Formel definition af CFG'er

En *kontekstfri grammatik* (CFG) er et 4-tupel

$G = (V, \Sigma, S, P)$ hvor

- V er en endelig mængde af **nonterminal**-symboler
- Σ er et alfabet af **terminal**-symboler og $V \cap \Sigma = \emptyset$
- $S \in V$ er et **start**-symbol
- P er en endelig mængde af **produktioner** på form $A \rightarrow \alpha$ hvor $A \in V$ og $\alpha \in (V \cup \Sigma)^*$

Derivationer

- "$\Rightarrow$" repræsenterer ét derivations-trin, hvor en nonterminal erstattes ifølge en produktion
- dvs. "$\Rightarrow$" er en relation over mængden $(V \cup \Sigma)^*$
- Hvis $\alpha_1, \alpha_2 \in (V \cup \Sigma)^*$ og $(A \rightarrow \gamma) \in P$
  (dvs. grammatikken indeholder produktionen $A \rightarrow \gamma$)
  så gælder

$$\alpha_1 A \alpha_2 \Rightarrow \alpha_1 \gamma \alpha_2$$

("$\Rightarrow$" er i denne sammenhæng **ikke** et "logisk medfører" tegn)

Sproget af en kontekstfri grammatik

- Definer relationen "$\Rightarrow^*$" som den *refleksive transitive lukning* af "$\Rightarrow$", dvs.

$$\alpha \Rightarrow^* \beta \quad \text{hvis og kun hvis} \quad \alpha \underbrace{\Rightarrow \ldots \Rightarrow \ldots \Rightarrow}_{\text{0 eller flere derivationstrin}} \beta$$

- **Sproget** af G defineres som
$L(G) = \{ x \in \Sigma^* \mid S \Rightarrow^* x \}$

- Et sprog $L \subseteq \Sigma^*$ er **kontekstfrit**
hvis og kun hvis
der findes en CFG G hvor $L(G)=L$

Eksempel 1

Sproget $A = \{ a^n b^n \mid n \geq 0 \}$ kan beskrives af en CFG $G=(V,\Sigma,S,P)$ hvor

- $V = \{S\}$
- $\Sigma = \{a,b\}$
- $P = \{S \to aSb,\ \ S \to \Lambda\}$

alternativ notation:
$S \to aSb \mid \Lambda$

dvs. $L(G) = A$ (bevis følger...)

Bevis for korrekthed

Påstand: $L(G) = A$

Bevisskitse: (udnyt at $x \in L(G) \Leftrightarrow S \Rightarrow^* x$)

- $L(G) \subseteq A$: givet x hvor $x \in L(G)$, lav **induktion i antal derivationsskridt** i $S \Rightarrow^* x$
- $A \subseteq L(G)$: givet x hvor $x \in A$, lav **induktion i længden** af x

Eksempel 2

Sproget *pal* $= \{\, x \in \{0,1\}^* \mid x = reverse(x) \,\}$

kan beskrives af en CFG $G=(V,\Sigma,S,P)$ hvor

- $V = \{S\}$

$\forall\ \Sigma = \{0,1\}$

- $P = \{S \rightarrow \Lambda\ ,$
  $S \rightarrow 0,$
  $S \rightarrow 1,$
  $S \rightarrow 0S0,$
  $S \rightarrow 1S1\}$

alternativ notation:
$S \rightarrow \Lambda \mid 0 \mid 1 \mid 0S0 \mid 1S1$

Hvorfor navnet "kontekstfri"?

- $\alpha_1 A \alpha_2 \Rightarrow \alpha_1 \gamma \alpha_2$
  hvis grammatikken indeholder
  produktionen $A \rightarrow \gamma$

- dvs. γ kan substituere A
  uafhængigt af konteksten (α_1 og α_2)

Anvendelser af kontekstfri grammatikker

- **Praktisk:** til beskrivelse af syntaks for programmeringssprog
  (ofte med BNF-notationen)

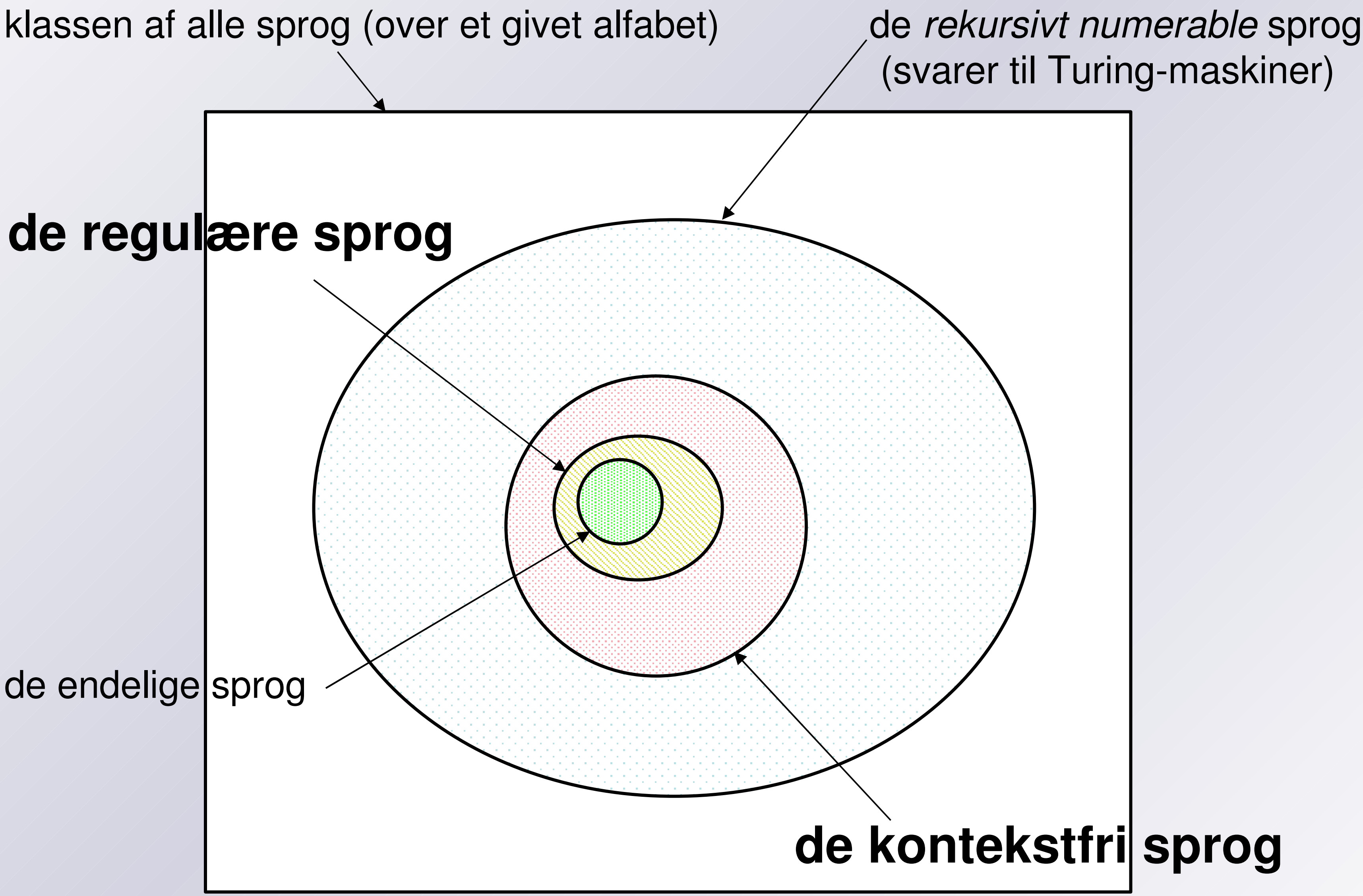
- **Teoretisk:** som karakteristik af en vigtig klasse af formelle sprog

En kontekstfri grammatik for Java

http://www.daimi.au.dk/dRegAut/JavaBNF.html

En tekst er et **syntaktisk korrekt Java-program** hvis den kan deriveres af denne grammatik

Klasser af formelle sprog



Øvelser

- [Martin] 6.1 (a+b+e)
- [Martin] 6.9 (a-c)

Resume

- Regulære sprog:
  - lukkethed under $\cup$, $\cap$, ‘, $\cdot$, *, homomorfi og invers homomorfi
  - “pumping”-lemmaet
  - beslutningsproblemer: *membership, emptiness, finiteness, subset, equality*
- Kontekstfri grammatikker:
  - definition af kontekstfri grammatikker og sprog