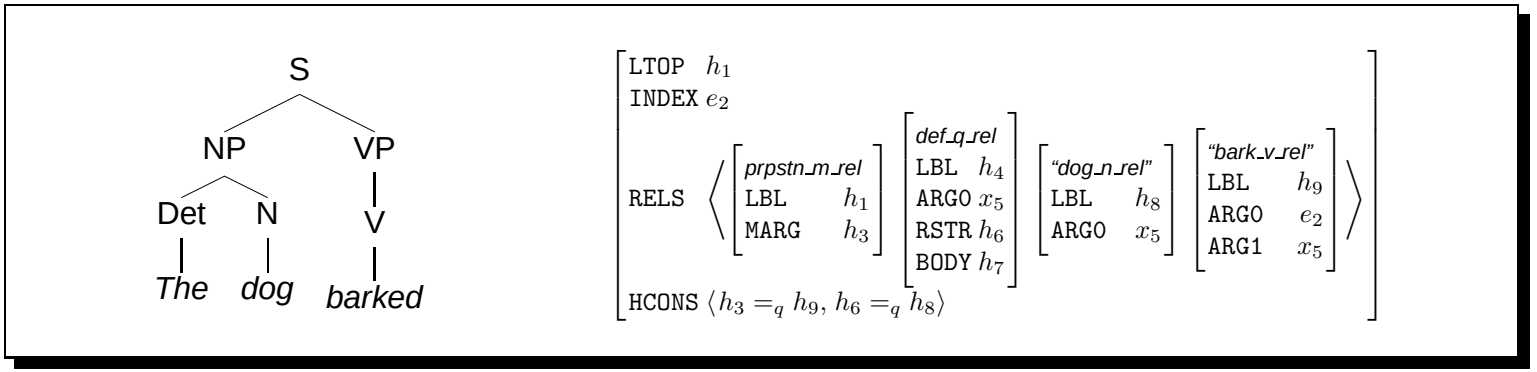

Computational Linguistics (INF2820 — Complexity)

$$\alpha \in C,\ \beta_i \in (C \cup \Sigma)^*,\ \gamma \in (C \cup \Sigma)^+,\ \delta \in \Sigma^+$$

Stephan Oepen

Universitetet i Oslo & CSLI Stanford

oe@ifi.uio.no

Review: Context-Free Grammars

- Formally, a *context-free grammar* (CFG) is a quadruple: $\langle C, \Sigma, P, S \rangle$
- C is the set of categories (aka *non-terminals*), e.g. $\{\text{S}, \text{NP}, \text{VP}, \text{V}\}$;
- Σ is the vocabulary (aka *terminals*), e.g. $\{\text{Kim}, \text{snow}, \text{saw}, \text{in}\}$;
- P is a set of category rewrite rules (aka *productions*), e.g.

$$\begin{aligned} \text{S} &\rightarrow \text{NP VP} \\ \text{VP} &\rightarrow \text{V NP} \\ \text{NP} &\rightarrow \text{Kim} \\ \text{NP} &\rightarrow \text{snow} \\ \text{V} &\rightarrow \text{saw} \end{aligned}$$

- $S \in C$ is the *start symbol*, a filter on complete ('sentential') results;
- for each rule '$\alpha \rightarrow \beta_1, \beta_2, ..., \beta_n$' $\in P$: $\alpha \in C$ and $\beta_i \in C \cup \Sigma$; $1 \leq i \leq n$.

The Chomsky Hierarchy of (Formal) Languages

- (Formal) Languages vary in 'degree of structural complexity' exhibited;
- traditionally: a^* (iteration) vs. $a^n b^n$ (nesting) vs. $a^n b^n c^n$ ('cross-serial');
- Chomsky Hierarchy: inclusion classes of formal languages; Type 0 – 3.

0	unrestricted	$\beta_1 \rightarrow \beta_2$	Turing Machine
1	context-sensitive	$\beta_1 \alpha \beta_2 \rightarrow \beta_1 \gamma \beta_2$	linearly-bounded automaton
2	context-free	$\alpha \rightarrow \beta$	push-down automaton
3	regular	$\alpha \rightarrow \delta \mid \delta\alpha \mid \alpha\delta$	finite-state automaton
$\alpha \in C,\ \beta_i \in (C \cup \Sigma)^*,\ \gamma \in (C \cup \Sigma)^+,\ \delta \in \Sigma^+$			

What is the Formal Complexity of Natural Languages?

- Minimally context-free (center self-embedding, e.g. in relative clauses);
- (Culy; Shieber, 1985): *not* context-free (Bambara, Swiss German);
- (Joshi, 1985): extra class of *mildly* context-sensitive languages (TAG).

Review: Examples of Formal Languages

A Really Complicated Language

[...] *omdat ik Henk de nijlpaarden zag voeren .*

A Really Complicated Language

[...] *omdat ik Jan Henk de nijlpaarden zag helpen voeren .*

Limitations of Context-Free Grammar

Agreement and Valency (For Example)

That dog barks.

**That dogs barks.*

**Those dogs barks.*

The dog chased a cat.

**The dog barked a cat.*

**The dog chased.*

**The dog chased a cat my neighbours.*

The cat was chased by a dog.

**The cat was chased of a dog.*

...

Agreement and Valency in Context-Free Grammars