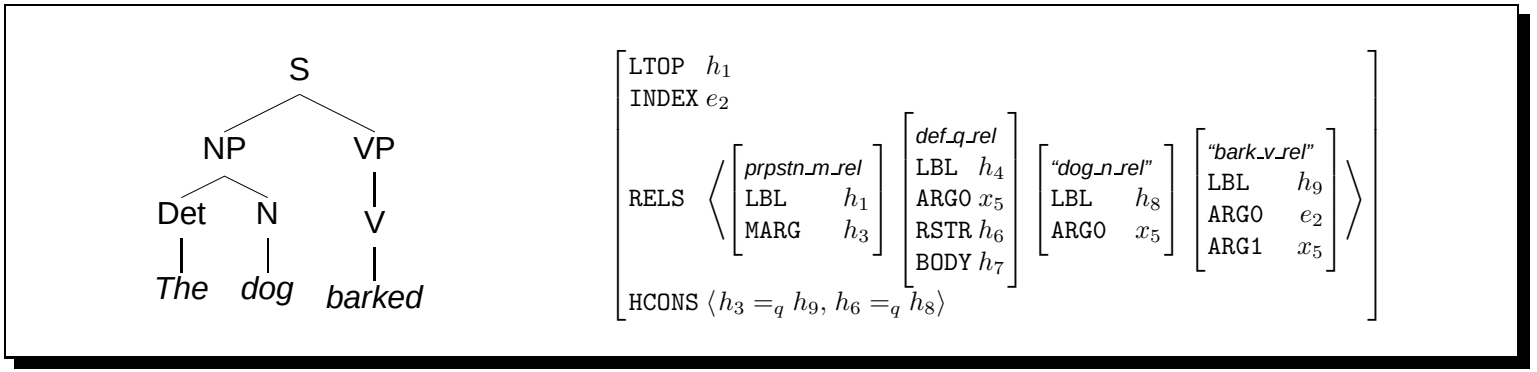


Computational Linguistics (INF2820 — Syntax)

S $\longrightarrow$ *NP VP; S* $\longrightarrow$ *S PP; S* $\longrightarrow$ *VP*

Stephan Oepen

Universitetet i Oslo & CSLI Stanford

oe@ifi.uio.no

Candidate Theories of Grammar (1 of 3)

Language as a Set of Strings

The dog barks.
The angry dog barks.
The fierce dog barks.
The fierce angry dog barks.
The angry fierce dog barks.
The dog chased a cat.
A dog chased the cat.
The dog chased a black cat.
The dog chased a young cat.
The dog of my neighbours chased a cat.
A dog chased the cat of my neighbours.
The cat of my neighbours was chased by a dog.
...

Grammatical Categories (1 of 2)

Word Clases or Parts of Speech (PoS)

cat, *dog*, *neighbour(s)*, ...	**noun (N)**
adore, *bark(s)*, *chase(d)*, *was*, ...	**verb (V)**
fierce, *angry*, *black*, *young*, ...	**adjective (A)**
quickly, *probably*, *not*, ...	**adverb (Adv)**
a, *the*, *my*, *that*, ...	**determiner (D)**
of, *by*, *on*, *at*, *under*, ...	**preposition (P)**
she, *mine*, *those*, *what*, ...	**pronoun (Pro)**
and, *neither ... nor*, *because*, ...	**conjunction (C)**

$$\textit{the} \left\{ \begin{array}{c} \textit{cat} \\ \textit{dog} \\ {}^{*}\textit{adore} \end{array} \right\} \quad \textit{Kim likes to} \left\{ \begin{array}{c} \textit{bark} \\ \textit{chase dogs} \\ {}^{*}\textit{cat} \end{array} \right\} \quad \textit{a} \left\{ \begin{array}{c} \textit{fierce} \\ \textit{angry} \\ {}^{*}\textit{quickly} \end{array} \right\} \textit{cat}$$

Grammatical Categories (2 of 2)

Number — Person — Case — Gender

That dog barks. — Those dogs bark.
I bark. — You bark. — They bark. — Sam shaved himself.
We bark. — You bark. — Those dogs bark.
I saw her. — She saw me. — My dog barked.
...

How many distinct verb forms according to number and person?

Tense — Aspect — Mood

The dog barks. — The dog barked — The dog will bark.
The dog has barked. — The dog is barking.
If I were a carpenter, ...
...

Candidate Theories of Grammar (2 of 3)

Language as a Sequence of Words

cat, *dog*, *neighbour(s)*, ...	**noun (N)**
adore, *bark(s)*, *chase(d)*, *was*, ...	**verb (V)**
fierce, *angry*, *black*, *young*, ...	**adjective (A)**
a, *the*, *my*, *that*, ...	**determiner (D)**
of, *by*, *on*, *at*, *under*, ...	**preposition (P)**

Regular Expressions

$$\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\ \mathrm{V}\ \left(\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\right)^{*}$$

Candidate Theories of Grammar (2 of 3)

Language as a Sequence of Words

cat, *dog*, *neighbour(s)*, ...	**noun (N)**
adore, *bark(s)*, *chase(d)*, *was*, ...	**verb (V)**
fierce, *angry*, *black*, *young*, ...	**adjective (A)**
a, *the*, *my*, *that*, ...	**determiner (D)**
of, *by*, *on*, *at*, *under*, ...	**preposition (P)**

Regular Expressions

$$\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\ \mathrm{V}\ \left(\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\right)^{*}$$

$$\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\ \left(\mathrm{P}\ \mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\right)^{*}\ \mathrm{V}\ \left(\mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\ \left(\mathrm{P}\ \mathrm{D}^{?}\ \mathrm{A}^{*}\ \mathrm{N}^{+}\right)^{*}\ \right)^{*}$$

Candidate Theories of Grammar (3 of 3)

Mildly Mathematically: 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.

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

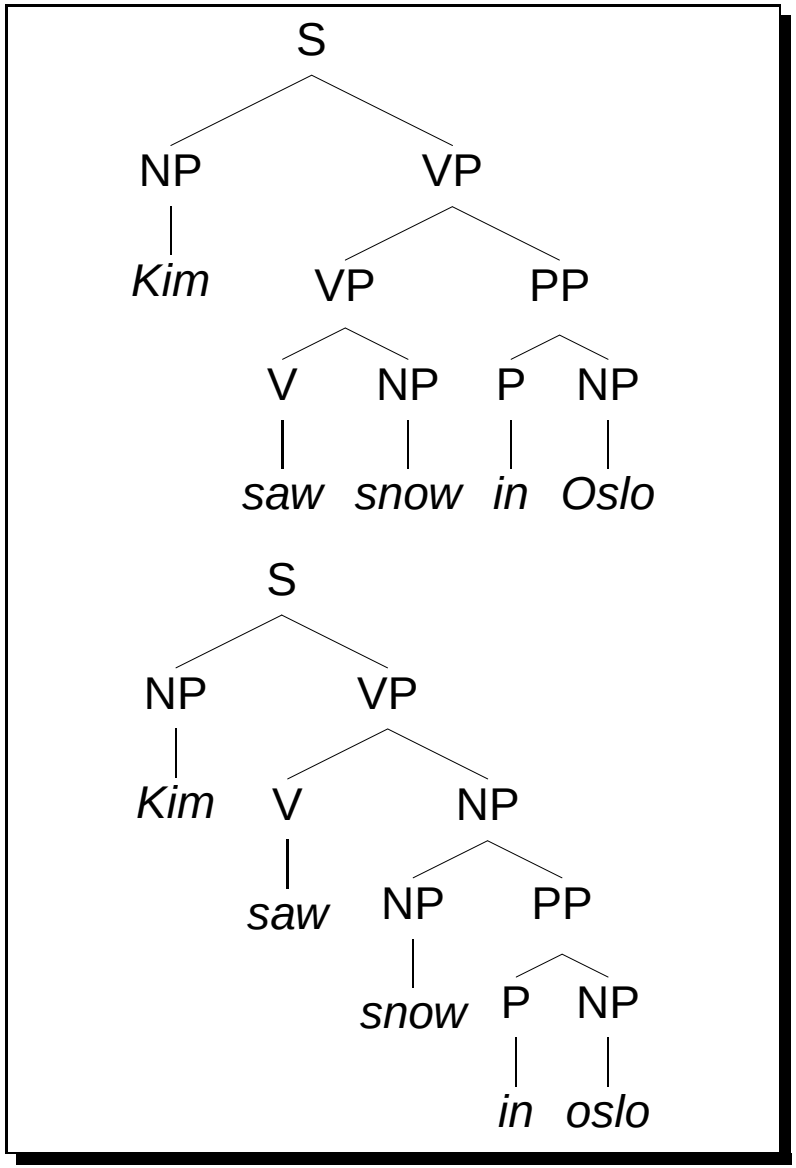
- $S \in C$ is the *start symbol*, a filter on complete ('sentential') results;
- for each rule '$\alpha \rightarrow \beta_1, \beta_2, ..., \beta_n$' $\in P$: $\alpha \in C$ and $\beta_i \in C \cup \Sigma$; $1 \leq i \leq n$.

Parsing: Recognizing the Language of a Grammar

$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$VP \rightarrow VP\ PP$
$NP \rightarrow NP\ PP$
$PP \rightarrow P\ NP$
$NP \rightarrow Kim \mid snow \mid Oslo$
$V \rightarrow saw$
$P \rightarrow in$

All Complete Derivations

- are rooted in the start symbol S;
- label internal nodes with categories $\in C$, leafs with words $\in \Sigma$;
- instantiate a grammar rule $\in P$ at each local subtree of depth one.



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.*

...