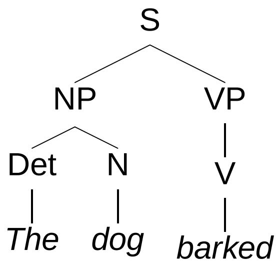

$$\begin{bmatrix} \texttt{LTOP} & h_1 \\ \texttt{INDEX} & e_2 \\ \texttt{RELS} & \left\langle \begin{bmatrix}\textit{prpstn\_m\_rel} & \\ \texttt{LBL} & h_1 \\ \texttt{MARG} & h_3\end{bmatrix} \begin{bmatrix}\textit{def\_q\_rel} & \\ \texttt{LBL} & h_4 \\ \texttt{ARG0} & x_5 \\ \texttt{RSTR} & h_6 \\ \texttt{BODY} & h_7\end{bmatrix} \begin{bmatrix}\textit{“dog\_n\_rel”} & \\ \texttt{LBL} & h_8 \\ \texttt{ARG0} & x_5\end{bmatrix} \begin{bmatrix}\textit{“bark\_v\_rel”} & \\ \texttt{LBL} & h_9 \\ \texttt{ARG0} & e_2 \\ \texttt{ARG1} & x_5\end{bmatrix} \right\rangle \\ \texttt{HCONS} & \langle h_3 =_q h_9,\, h_6 =_q h_8 \rangle \end{bmatrix}$$

Computational Linguistics (INF2820 — Beyond CFGs)

$$\textit{phrase}\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SPR} & \langle\rangle \\ \texttt{COMPS} & \boxed{3} \end{bmatrix} \longrightarrow \boxed{2}\,\textit{phrase}\begin{bmatrix} \texttt{SPR} & \langle\rangle \\ \texttt{COMPS} & \langle\rangle \end{bmatrix},\ \textit{phrase}\begin{bmatrix} \texttt{HEAD} & \boxed{1} \\ \texttt{SPR} & \left\langle\boxed{2}\right\rangle \\ \texttt{COMPS} & \boxed{3} \end{bmatrix}$$

~~Wilhelm~~ Stephan Oepen & Jan Tore Lønning

Universitetet i Oslo

`oe@ifi.uio.no`

INF2880 — What We Are About to Do (and Why)

Course Outline

- Extend understanding of (natural) language as a system of rules;
- learn how to *formalize* grammars through typed feature structures;
- solve practical exercises: immediate gratification (risk of late hours).

Three Interacting Components

- **formal syntax** learn and practice (basic) notions of formal syntactic theory; by and large framework-independent and common sense;
- **grammar engineering** formalize linguistic theories with complex interactions of multiple phenomena; implementation and debugging;
- **processing** understand standard parsing algorithms; unification of typed feature structures; fundamentals of unification-based parsing.

Grammar Engineering from a CS Perspective

Implementation Goals

- Translate linguistic analysis into computational formalism: formal model;
- computational grammar provides mapping between form and meaning;
- assign correct analyses to grammatical, reject ungrammatical inputs;
- parsing and generation algorithms: apply mapping in either direction.

Analogy to (Object-Oriented) Programming

- Computational system with observable behavior: immediately testable;
- typed feature structures as a specialized (OO) programming language;
- make sure that all the pieces fit together; revise – test – revise – test ...

Comments on Background Literature

Natural Language Processing and Computational Grammar

(1) Jurafsky, Daniel and Martin, James H.: *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd Edition).* Upper Saddle River, NJ: Prentice Hall (2008).

(2) Sag, Ivan A. Tom Wasow, and Emily M. Bender: *Syntactic Theory. A Formal Introduction (2nd Edition).* Stanford, CA: CSLI Publications (2003);

(3) Copestake, Ann: *Implementing Typed Feature Structure Grammars.* Stanford, CA: CSLI Publications (2001).

Selected chapters from (2) and (3) are available as a 'course pack' (*kompendium*) from Akademika; in store starting Thursday, March 17.

Course Logistics Ahead of Us (And a Moral Appeal)

Recap: How to Define Grammatical Categories

Word Classes or Parts of Speech (PoS)

cat, *dog*, *neighbours*, ...	**noun (N)**
adore, *barks*, *chased*, *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 (Pron)**
and, *neither ... nor*, *because*, ...	**conjunction (C)**

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

Recap: More Grammatical Categories

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.

...

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, ...

...

Limitations of (Our) Context-Free Grammars

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

Structured Categories in a Unification Grammar

- All categories in the grammar are (typed) feature structures (aka TFSs);
- specific TFS configurations may correspond to ‘traditional’ categories;

→ labels like ‘S’ or ‘NP’ are mere abbreviations, not elements of the theory.

$$\mathit{word}\begin{bmatrix} \texttt{HEAD} & \mathit{noun} \\ \texttt{SPR} & \left\langle \begin{bmatrix} \texttt{HEAD}\ \mathit{det} \end{bmatrix} \right\rangle \\ \texttt{COMPS} & \langle\rangle \end{bmatrix}$$

‘N’

‘lexical’

$$\mathit{phrase}\begin{bmatrix} \texttt{HEAD} & \mathit{verb} \\ \texttt{SPR} & \langle\rangle \\ \texttt{COMPS} & \langle\rangle \end{bmatrix}$$

‘S’

‘maximal’

$$\mathit{phrase}\begin{bmatrix} \texttt{HEAD} & \mathit{verb} \\ \texttt{SPR} & \left\langle \begin{bmatrix} \texttt{HEAD}\ \mathit{noun} \end{bmatrix} \right\rangle \\ \texttt{COMPS} & \langle\rangle \end{bmatrix}$$

‘VP’

‘intermediate’

Preliminary Words on Specifiers and Complements

Interaction of Lexicon and Phrase Structure Schemata

$$\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \boxed{3} \end{bmatrix} \longrightarrow \boxed{2}\begin{bmatrix} \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix}, \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SPR} & \langle\boxed{2}\rangle \\ \text{COMPS} & \boxed{3} \end{bmatrix}$$

$$\begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \longrightarrow \textit{Kim} \qquad \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \left\langle \begin{bmatrix} \text{HEAD} & \textit{noun} \\ \text{SPR} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \longrightarrow \textit{sleeps}$$