# Formal Syntax and Grammar Engineering

## Stephan Oepen

Universitetet i Oslo & CSLI Stanford

`oe@csli.stanford.edu`

## Lilja Øvrelid

Göteborgs Universitet

`lilja.ovrelid@svenska.gu.se`
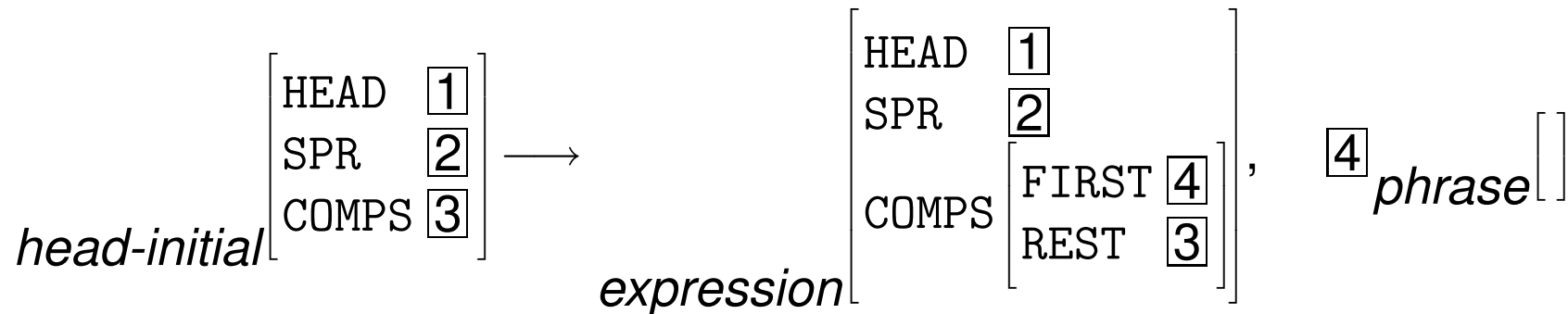
`http://www.delph-in.net/courses/04/fs/`

# (1) Grammar Formalisms

(a) Would it be sufficient to describe natural language simply in terms of sequences of words? If so, sketch a grammar of modern English in that form; if not give a few examples of linguistic phenomena that call for a richer description language.

(b) Is it possible to write a context-free grammar for the language $a^n\, b^n$, i.e. (assuming a vocabulary $\Sigma = \{a, b\}$) all strings of a sequence of $a$ occurrences followed by an equal number of $b$ occurrences? If so, then for extra credit, how about $a^n\, b^n\, c^n$?

(c) Linguists tend to like descriptively 'rich' formalisms, hence often claim that context-free grammars are inadequate (or even insufficient) for the analysis of natural languages. Discuss this claim in a few sentences, preferably giving supporting arguments and examples.

# (2) Unification-Based Grammar

(a) Show the feature structure representation of the following rule as it is used in our LKB grammars:

$$
\textit{head-initial}\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SPR} & \boxed{2} \\ \text{COMPS} & \boxed{3} \end{bmatrix} \longrightarrow \textit{expression}\begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SPR} & \boxed{2} \\ \text{COMPS} & \begin{bmatrix} \text{FIRST} & \boxed{4} \\ \text{REST} & \boxed{3} \end{bmatrix} \end{bmatrix}, \quad \boxed{4}\textit{phrase}\begin{bmatrix} \ \end{bmatrix}
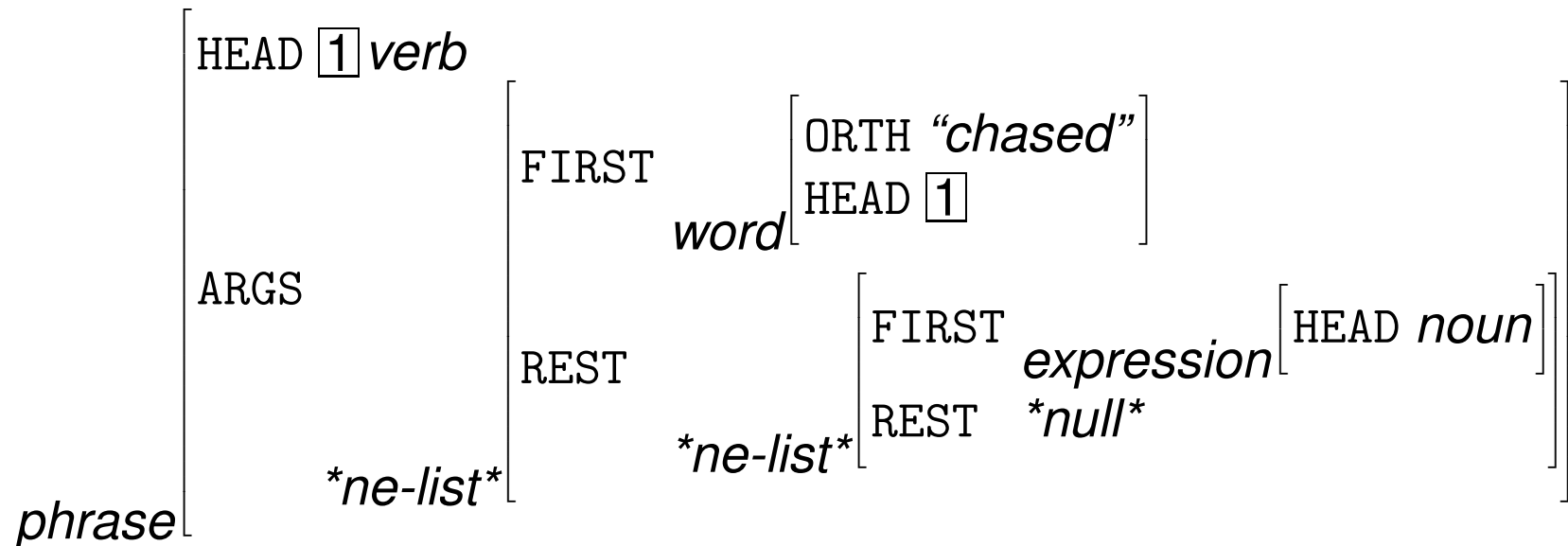$$

(b) What is the (approximate) name of the above rule in our grammar? Sketch its functionality in a few sentences and provide two examples of types of phrases built using this rule.

(c) Why did we choose to implement rules as single feature structures?

# (3) Typed Feature Structures

(a) Draw the following feature structure in DAG notation, i.e. as a directed
acyclic graph of labeled nodes and directed arcs:

$$
phrase\begin{bmatrix}
\text{HEAD } \boxed{1}\ verb \\
\text{ARGS } *ne\text{-}list*\begin{bmatrix}
\text{FIRST } word\begin{bmatrix}\text{ORTH } \text{``chased''} \\ \text{HEAD } \boxed{1}\end{bmatrix} \\
\text{REST } *ne\text{-}list*\begin{bmatrix}\text{FIRST } expression\begin{bmatrix}\text{HEAD } noun\end{bmatrix} \\ \text{REST } *null*\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

(b) In no more than two sentences, comment on the correspondences be-
tween elements of the feature structure and elements of the DAG.

# (4) Linguistic Concepts

(a) Name at least two syntactic properties that characterize (syntactic) heads in the formation of phrases. Use one or two examples.

(b) Sketch a constituent tree for the sentence *the fierce dog chased the cat near the aardvark*. On each node, provide information about its general category (using abbreviatory notions like 'Det', 'N', 'NP', 'VP', et al.) and for each branch of the tree indicate whether the constituent dominated by it acts as a *head*, *specifier*, *complement*, or *modifier*.

(c) What is the main difference between *agreement* vs. *government* relations? Give at least one example of each.

(d) What is the difference between the function of the prepositional phrases (marked by square brackets) in the following two sentences:

   (i) *that cat gave the aardvark [to the dogs]*
   (ii) *the cat chased the aardvark [near the dogs]*