Syntactic Theory
Head-driven Phrase Structure Grammar (HPSG)

Dr. Dan Flickinger & PD Dr. Valia Kordoni

Department of Computational Linguistics
Saarland University

November 25, 2011

HPSG from a Linguistic Perspective

From a linguistic perspective, an HPSG consists of

- A lexicon
  licensing basic words

- Lexical rules
  licensing derived words

- Immediate dominance (ID) schemata
  licensing constituent structure

- Linear precedence (LP) statements
  constraining word order

- A set of grammatical principles
  expressing generalizations about linguistic objects

The Signature

- Defines the ontology
  - Which kind of objects are distinguished
  - Which properties are modeled
- Consists of
  - Type inheritance hierarchy
  - Appropriate features and constraints on types

Linguistic Description

- Linguistic theories are described using attribute-value matrices (AVMs): the description language of typed feature structures (TFS)
- A set of description statements comprises the constraints on what are the admissible linguistic objects (iff there is a corresponding well-formed TFS satisfying all the constraints)

Description Example

A verb, for example, can specify that its subject be masculine singular:

(1) Ya   spal.
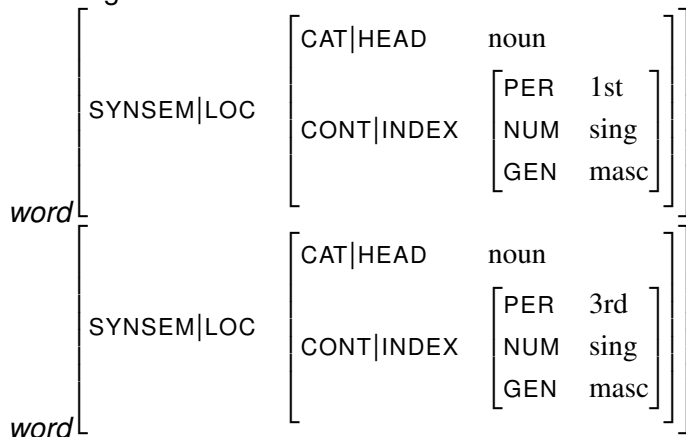   I*masc.sg* slept*masc.sg*

(2) On   spal.
   He*masc.sg* slept*masc.sg*

$$
word \begin{bmatrix} \text{SYNSEM|LOC} & \begin{bmatrix} \text{CAT|HEAD} & noun \\ \text{CONT|INDEX} & \begin{bmatrix} \text{NUM} & sing \\ \text{GEN} & masc \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

This AVM specifies the "partial" constraints on the complete (totally well-typed) feature structure of the subject

Subsumption

The AVM description on the previous slide subsumes both of the
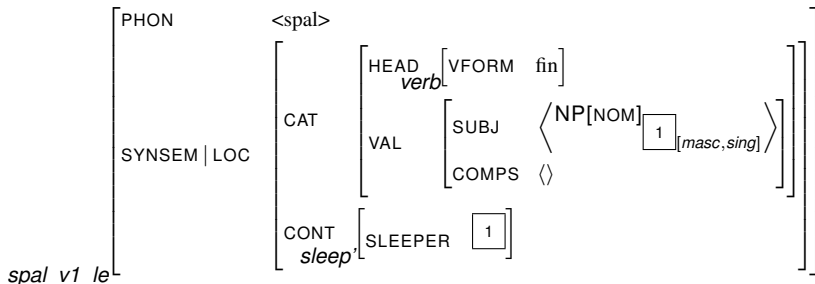following AVMs

$$
\begin{aligned}
word &
\begin{bmatrix}
\text{SYNSEM|LOC}
\begin{bmatrix}
\text{CAT|HEAD} & noun \\[2ex]
\text{CONT|INDEX} &
\begin{bmatrix}
\text{PER} & 1st \\
\text{NUM} & sing \\
\text{GEN} & masc
\end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
word &
\begin{bmatrix}
\text{SYNSEM|LOC}
\begin{bmatrix}
\text{CAT|HEAD} & noun \\[2ex]
\text{CONT|INDEX} &
\begin{bmatrix}
\text{PER} & 3rd \\
\text{NUM} & sing \\
\text{GEN} & masc
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{aligned}
$$

The Lexicon

- The basic lexicon defines the ontologically possible words that are grammatical:

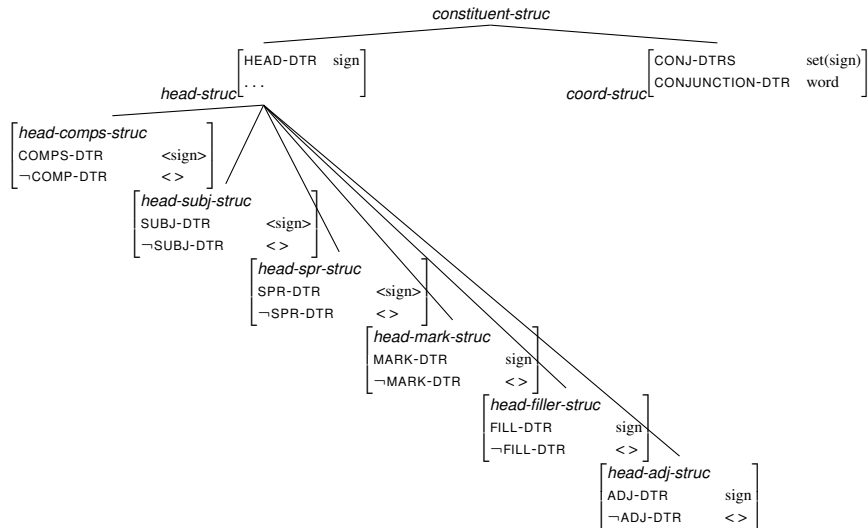 $$word \rightarrow \texttt{lexical\_entry}_1 \vee \texttt{lexical\_entry}_2 \vee \ldots$$

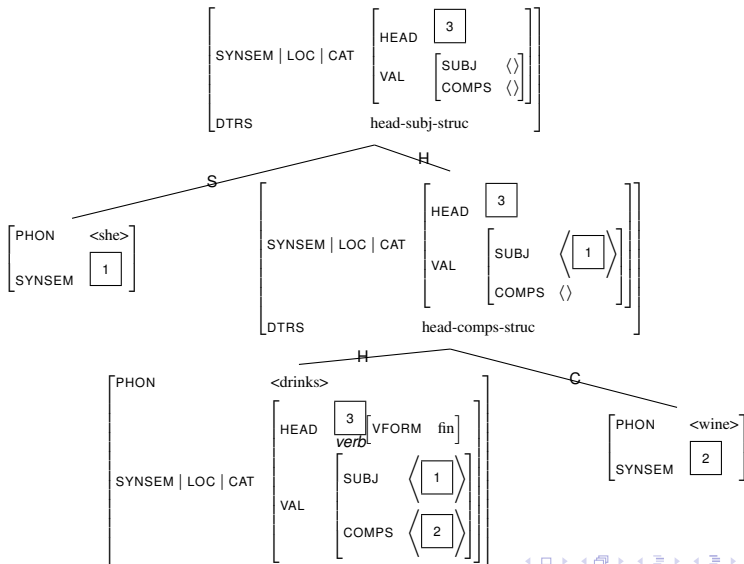- Each lexical entry is described by an AVM, e.g.

Types of Phrases

- Each *phrase* has a DTRS attribute which has a *constituent-structure* value
- This DTRS value corresponds to what we view in a tree as daughters (with additional grammatical role information, e.g. adjunct, complement, etc.)
- By distinguishing different kinds of *constituent-structures*, we can define different kinds of constructions in a language

An Ontology of Phrases

A Sketch of Head-Subject/Complement Structures

Universal Principles

How exactly did the last example work?

- *drink* has head information specifying that it is a finite verb and subcategories for a subject and an object
    - The head information gets percolated up (the HEAD feature principle)
    - The valence information gets "checked off" as one moves up in the tree (the VALENCE principle)

Such principles are treated as linguistic universals in HPSG

HEAD-Feature Principle

HEAD-feature principle

The value of the HEAD feature of any headed phrase is token-identical with the HEAD value of the head daughter

$$
\textit{phrase}\begin{bmatrix} \text{DTRS} & \text{head-struc} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \boxed{1} \\ \text{DTRS} \mid \text{HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \boxed{1} \end{bmatrix}
$$

VALENCE Principle

VALENCE principle

In a headed phrase, for each valence feature F, the F value of the head daughter is the concatenation of the phrase's F value with the list of F-DTR's SYNSEM

$$
\left[\text{DTRS} \quad \text{head-struc} \right] \rightarrow
\begin{bmatrix}
\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{F} \quad \boxed{1} \\
\text{DTRS} \begin{bmatrix}
\text{HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{F} \quad \boxed{1} \oplus \left\langle \boxed{2} \right\rangle \\
\text{F-DTR} \mid \text{FIRST} \mid \text{SYNSEM} \quad \boxed{2}
\end{bmatrix}
\end{bmatrix}
$$

- F can be any one of SUBJ, COMPS, SPR
- $\oplus$ stands for list concatenation:

$$
\textit{elist} \oplus \boxed{1} := \boxed{1} \qquad \left\langle \boxed{1} \; \boxed{2} \right\rangle \oplus \boxed{3} := \left\langle \boxed{1} \; \boxed{2} \oplus \boxed{3} \right\rangle
$$

- When the F-DTR is empty, the F valence feature of the head daughter will be copied to the mother phrase

Semantics Principle

Semantics principle

In a headed phrase, the CONTENT value is token-identical to that of the adjunct daughter if the DTRS value is of sort *head-adj-struc*, and with that of the head daughter otherwise.

$$phrase\begin{bmatrix} \text{DTRS} & \text{head-adj-struc} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CONT} & \boxed{1} \\ \text{DTRS} \mid \text{NON-HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CONT} & \boxed{1} \end{bmatrix}$$

Otherwise:

$$phrase\begin{bmatrix} \text{DTRS} & \text{head-struc} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CONT} & \boxed{1} \\ \text{DTRS} \mid \text{HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CONT} & \boxed{1} \end{bmatrix}$$

Fallout from These Principles

- Note that agreement is handled neatly, simply by the fact that the SYNSEM values of a word's daughters are token-identical to the items on the VALENCE lists
- How exactly do we decide on a syntactic structure?
- Why is the subject checked off at a higher point in the tree?

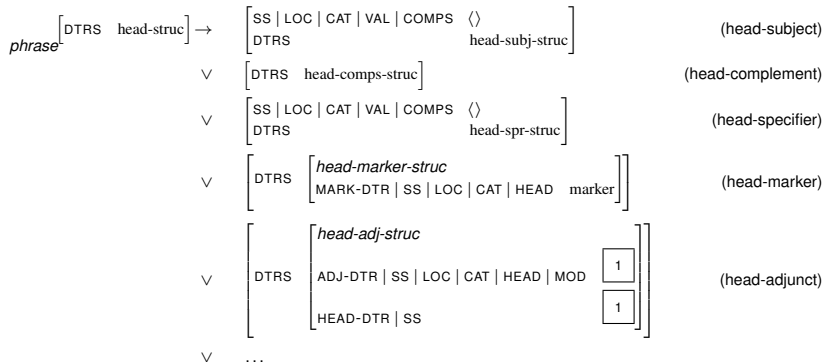# Immediate Dominance (ID) Principle & Schemata

ID Principle

Every headed phrase must satisfy exactly one of the ID schemata

- The exact inventory of valid ID schemata is language-specific
- We will introduce a set of ID schemata for English

Immediate Dominance Schemata (for English)

$$
\textit{phrase}\begin{bmatrix} \text{DTRS} & \text{head-struc} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SS | LOC | CAT | VAL | COMPS} & \langle\rangle \\ \text{DTRS} & \text{head-subj-struc} \end{bmatrix} \quad \text{(head-subject)}
$$

$$
\vee \quad \begin{bmatrix} \text{DTRS} & \text{head-comps-struc} \end{bmatrix} \quad \text{(head-complement)}
$$

$$
\vee \quad \begin{bmatrix} \text{SS | LOC | CAT | VAL | COMPS} & \langle\rangle \\ \text{DTRS} & \text{head-spr-struc} \end{bmatrix} \quad \text{(head-specifier)}
$$

$$
\vee \quad \begin{bmatrix} \text{DTRS} & \begin{bmatrix} \textit{head-marker-struc} \\ \text{MARK-DTR | SS | LOC | CAT | HEAD} & \text{marker} \end{bmatrix} \end{bmatrix} \quad \text{(head-marker)}
$$

$$
\vee \quad \begin{bmatrix} \text{DTRS} & \begin{bmatrix} \textit{head-adj-struc} \\ \text{ADJ-DTR | SS | LOC | CAT | HEAD | MOD} & \boxed{1} \\ \\ \text{HEAD-DTR | SS} & \boxed{1} \end{bmatrix} \end{bmatrix} \quad \text{(head-adjunct)}
$$

$$
\vee \quad \dots
$$

References I

Pollard, C. J. and Sag, I. A. (1994).
Head-Driven Phrase Structure Grammar.
University of Chicago Press, Chicago, USA.