Syntactic Theory

Head-driven Phrase Structure Grammar (HPSG)

Yi Zhang

Department of Computational Linguistics
Saarland University

December 17th, 2009

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 AVMs: description language of TFS
- A set of description statements comprises the constraints on what are the admissible linguistic objects (iff there is 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$_{masc.sg}$

$$\textit{word}\left[\begin{array}{ll}\text{SYNSEM}|\text{LOC} & \left[\begin{array}{ll}\text{CAT}|\text{HEAD} & \text{noun} \\ \text{CONT}|\text{INDEX} & \left[\begin{array}{ll}\text{NUM} & \text{sing} \\ \text{GEN} & \text{masc}\end{array}\right]\end{array}\right]\end{array}\right]$$

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

$$\textit{word}\left[\begin{array}{l}\text{SYNSEM}|\text{LOC}\left[\begin{array}{ll}\text{CAT}|\text{HEAD} & \text{noun}\\ \text{CONT}|\text{INDEX} & \left[\begin{array}{ll}\text{PER} & \text{1st}\\ \text{NUM} & \text{sing}\\ \text{GEN} & \text{masc}\end{array}\right]\end{array}\right]\end{array}\right]$$

$$\textit{word}\left[\begin{array}{l}\text{SYNSEM}|\text{LOC}\left[\begin{array}{ll}\text{CAT}|\text{HEAD} & \text{noun}\\ \text{CONT}|\text{INDEX} & \left[\begin{array}{ll}\text{PER} & \text{3rd}\\ \text{NUM} & \text{sing}\\ \text{GEN} & \text{masc}\end{array}\right]\end{array}\right]\end{array}\right]$$

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.

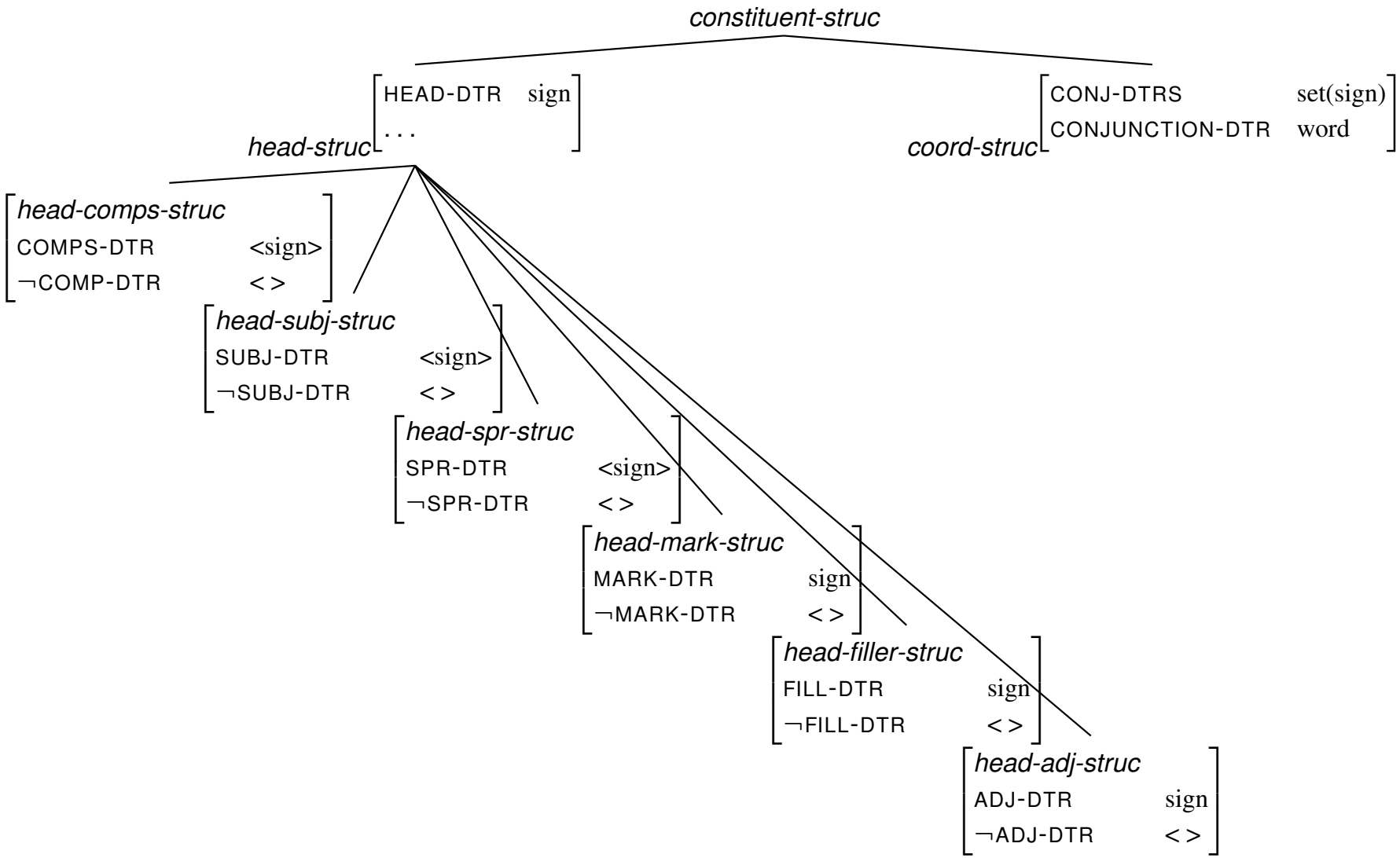
spal_v1_le

$$\begin{bmatrix} \text{PHON} & \text{<spal>} \\ \text{SYNSEM} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{VFORM} & \text{fin} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \left\langle \text{NP[NOM]}\boxed{1}_{[\textit{masc},\textit{sing}]} \right\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{sleep'} \\ \text{SLEEPER} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

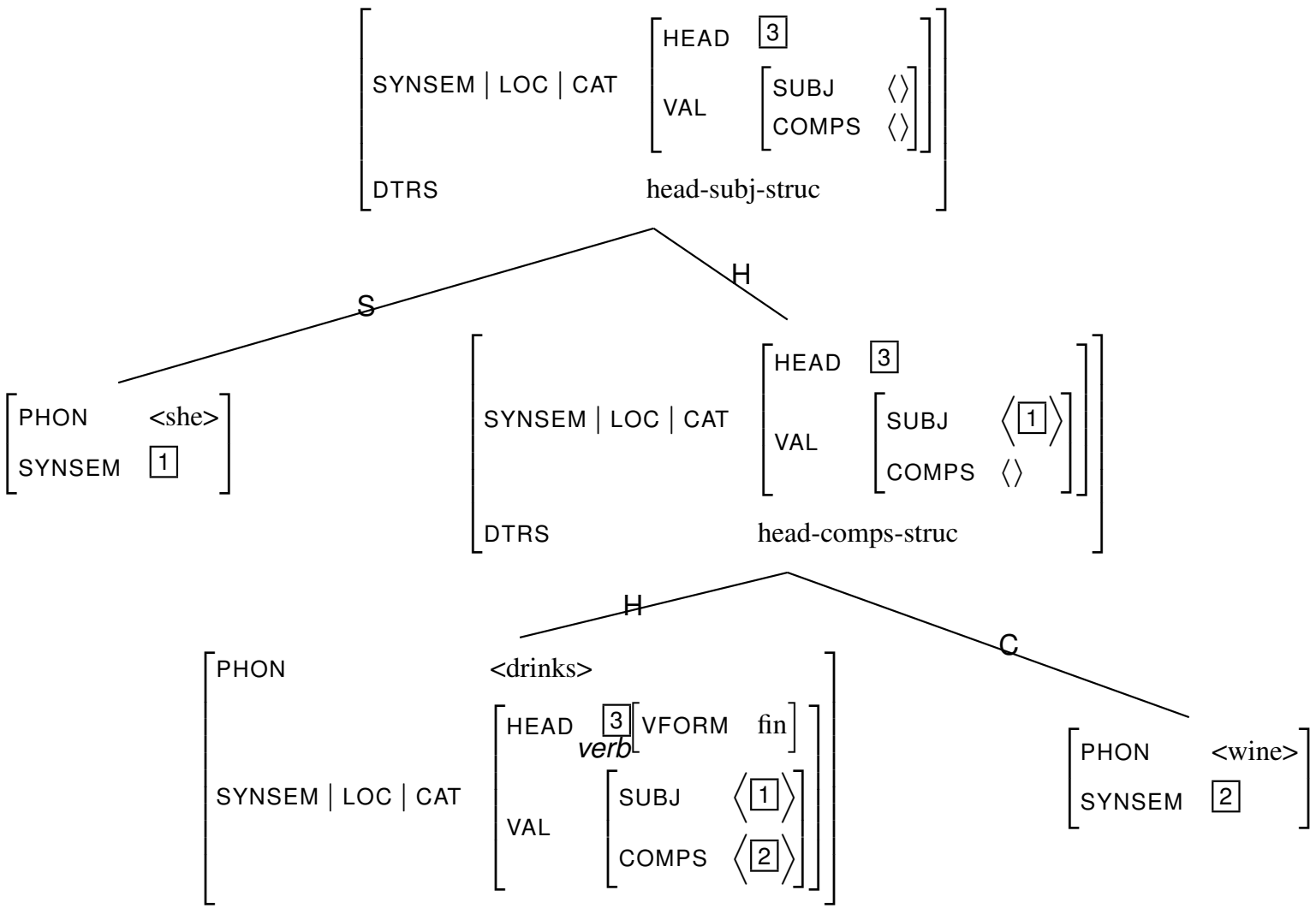
Types of Phrases

- 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

$$\textit{phrase}\left[\text{DTRS}\quad \text{head-struc}\right] \rightarrow \left[\begin{array}{ll}\text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \boxed{1} \\ \text{DTRS} \mid \text{HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \boxed{1}\end{array}\right]$$

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

$$\left[\text{DTRS}\quad \text{headed-structure}\right] \rightarrow \left[\begin{array}{ll} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{F} & \boxed{1} \\ \text{DTRS} & \left[\begin{array}{ll} \text{HEAD-DTR} \mid \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{F} & \boxed{1} \oplus \langle \boxed{2} \rangle \\ \text{F-DTR} \mid \text{FIRST} \mid \text{SYNSEM} & \boxed{2} \end{array}\right] \end{array}\right]$$

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

$$\textit{elist} \oplus \boxed{1} := \boxed{1}$$

$$\left\langle \boxed{1} \middle| \boxed{2} \right\rangle \oplus \boxed{3} := \left\langle \boxed{1} \middle| \boxed{2} \oplus \boxed{3} \right\rangle$$

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

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 the subject is checked off at a higher point in the tree?

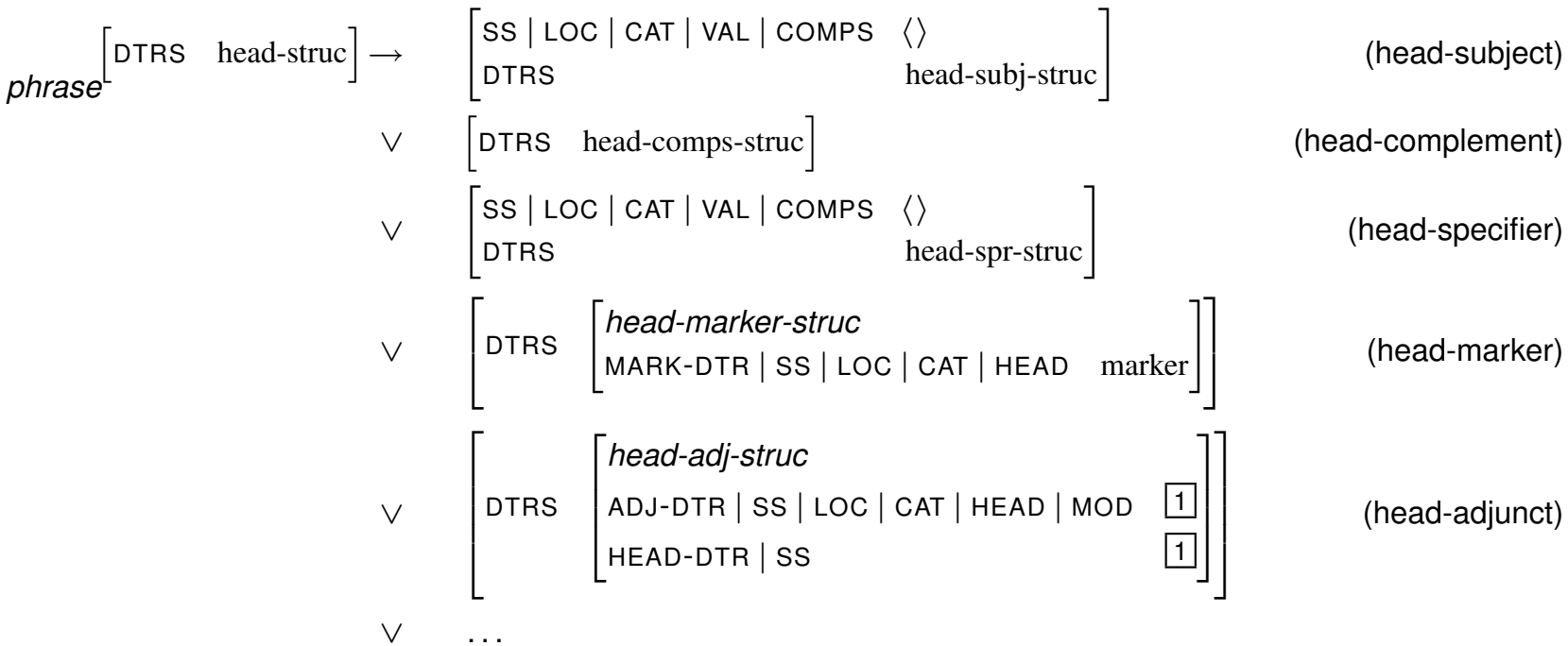
Immediate Dominance (ID) Principle & Schemata

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)



References I

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