Seminar für Sprachwissenschaft
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Grammatikformalismen und Parsing

Prof. Dr. Erhard W. Hinrichs

Frank Richter

Beata Trawiński

(eh,fr,trawinski)@sfs.uni-tuebingen.de

Seminar für Sprachwissenschaft

Abteilung Computerlinguistik

Universität Tübingen

Seminar für Sprachwissenschaft
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

HPSG as a Formal Linguistic Theory I

The grammar of HPSG'94

SS 2002

Fri, April 26th 2002

Two Aspects of HPSG

- HPSG as a lingustic formalism
  (i.e., a set of formal tools for formalizing linguistic analyses of various linguistic phenomena)

- HPSG as a formal linguistic theory
  (i.e., a collection of analyses of various linguistic phenomena encoded in this formalism, i.e. a formal grammar)

Some General Properies

HPSG is a

- generative,
- eclectic,
- comprehensive,
- lexicalist,
- non-derivational,
- constraint-based

linguistic framework.

Pollard and Sag (1994)

The grammar of P&S'94 comes in two versions:

- chapters 1-8 and appendix
  - valence properties are captured by the value of one attribut SUBCAT
  - extraction of an object leaves a trace
- chapter 9
  - valence properties are captured by the value of three attributes, SUBJ, SPR and COMPS
  - extraction of an object does not leave a trace

General Linguistc Assuptions in P&S'94: *signs*

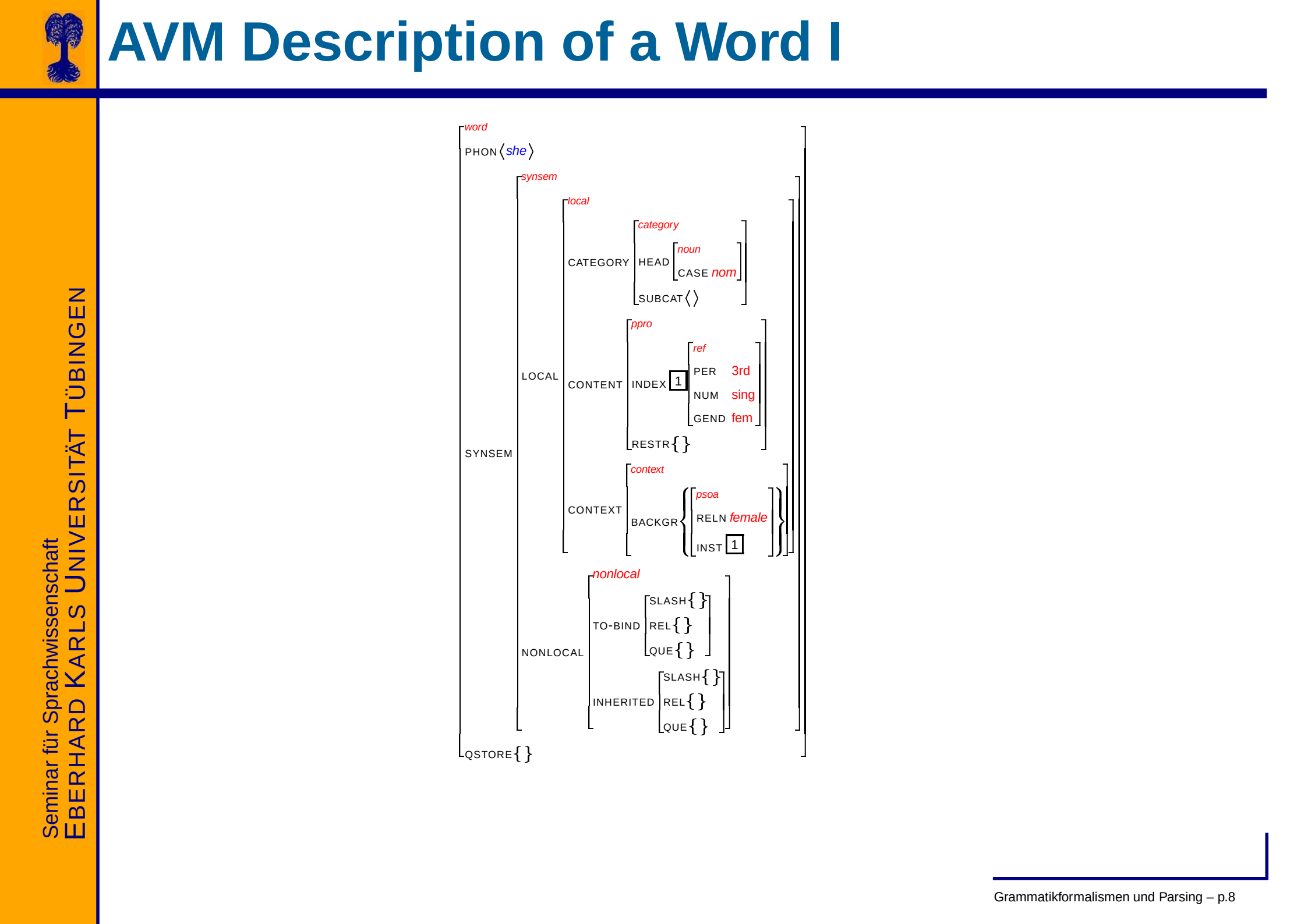
P&S'94 assumes that all linguistic expressions are *signs*, that is structured complexes of

- phonology,
- morphology,
- syntax,
- semantics,
- discourse.

Subtypes of *sign*

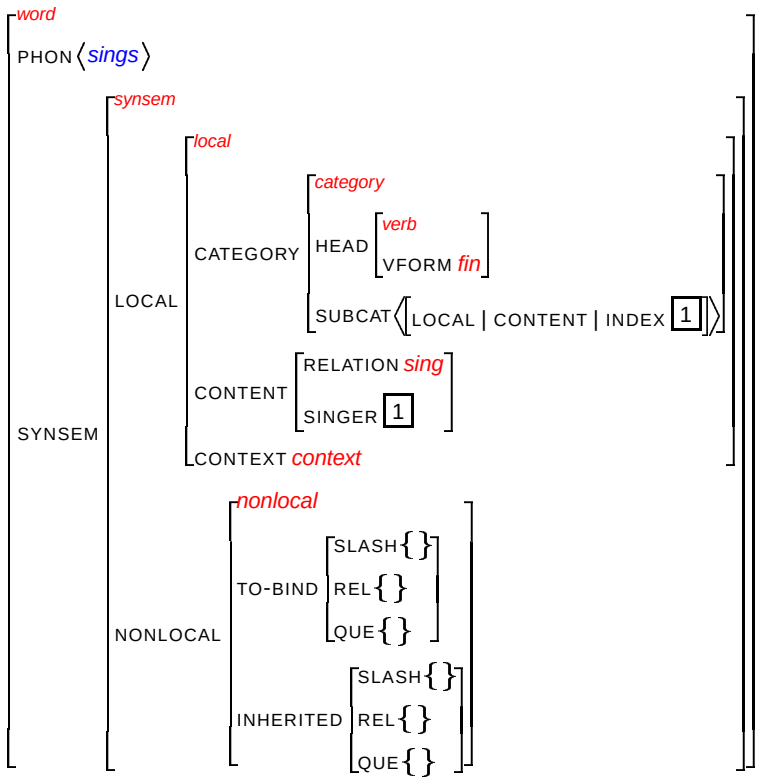
Signs fall into two disjoint subtypes:

- lexical signs (*words*)
- phrasal signs (*phrases*).

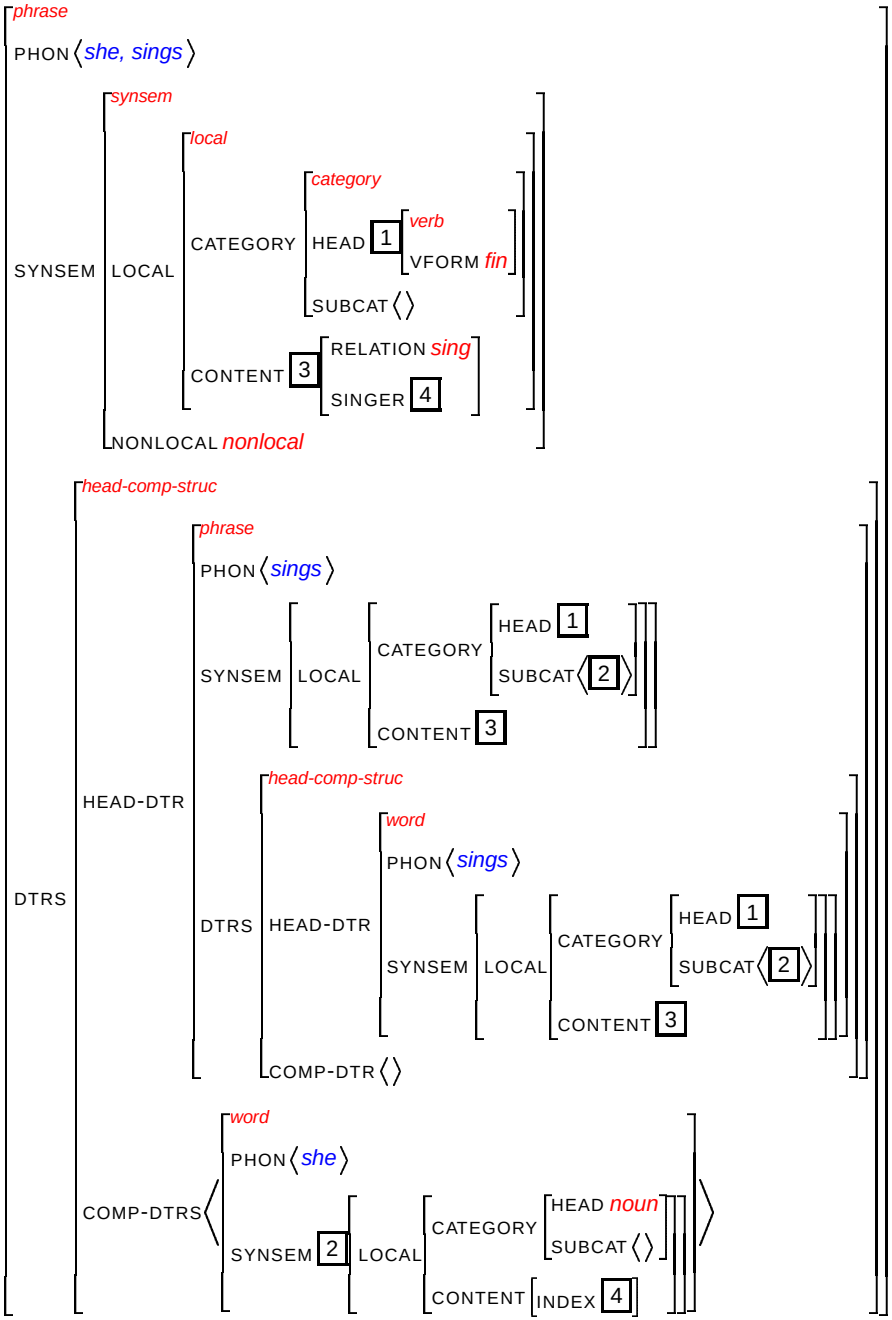


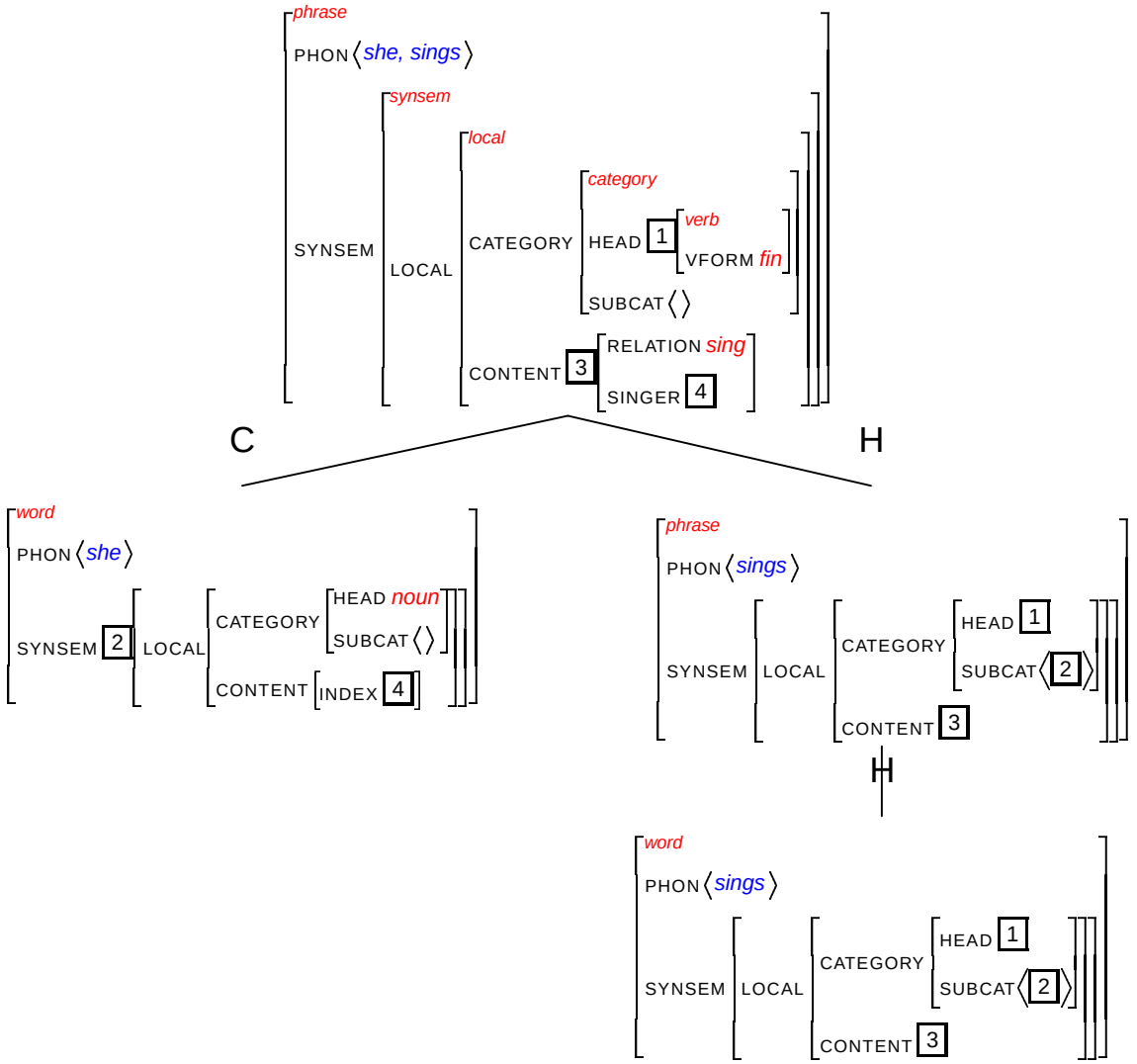
AVM Description of a Word I

AVM Description of a Word II



AVM Description of a Phrase





Tree Structure of the Phrase *She sings*

SCHEMA 1 (HEAD-SUBJECT SCHEMA)

The SYNSEM | LOCAL | CATEGORY | SUBCAT value is $\langle\rangle$, and the DAUGHTERS value is an object of sort *head-comp-struc* whose HEAD-DAUGHTER is a phrase whose SYNSEM | NONLOCAL | TO-BIND | SLASH value is {}, and whose COMPLEMENT-DAUGHTERS value is a list of length one.

SCHEMA 1 (HEAD-SUBJECT SCHEMA) **formalized**

$$\left[\begin{array}{l}\text{SS LOC CAT}\ \left[\begin{array}{l}\text{HEAD}\quad \left[\left[\text{INV}\ \textit{minus}\right] \vee \neg\ \textit{verb}\right] \\ \text{SUBCAT}\ \langle\rangle\end{array}\right] \\ \text{DTRS}\ \left[\begin{array}{l}\textit{head-comp-struc} \\ \text{HEAD-DTR}\ \left[\begin{array}{l}\textit{phrase} \\ \text{SS NONLOC TO-BIND SLASH}\ \{\}\end{array}\right] \\ \text{COMP-DAUGHTERS}\ \langle \textit{object} \rangle\end{array}\right]\end{array}\right]$$

SCHEMA 2 (HEAD-COMPLEMENT SCHEMA)

The SYNSEM | LOCAL | CATEGORY | SUBCAT value is a list of length one, and the daughters value is an object of sort *head-comp-struc* whose HEAD-DAUGHTER value is a word.

SCHEMA 2 (HEAD-COMPLEMENT SCHEMA) **formalized**

$$\begin{bmatrix} \text{SS LOC CAT} & \begin{bmatrix} \text{HEAD} & \left[\left[\text{INV } \textit{minus}\right] \vee \neg\, \textit{verb}\right] \\ \text{SUBCAT} & \langle \textit{object} \rangle \end{bmatrix} \\ \text{DTRS} & \begin{bmatrix} \textit{head-comp-struc} \\ \text{HEAD-DTR } \textit{word} \end{bmatrix} \end{bmatrix}$$

SCHEMA 3 (HEAD-SUBJECT-COMPLEMENT SCHEMA)

The SYNSEM | LOCAL | CATEGORY | SUBCAT value is $\langle\rangle$, and the DAUGHTERS value is an object of sort *head-comp-struc* whose HEAD-DAUGHTER value is a word.

SCHEMA 3 (HEAD-SUBJECT-COMPLEMENT SCHEMA) **formalized**

$$\begin{bmatrix} \text{SS LOC CAT} & \begin{bmatrix} \text{HEAD INV} & \textit{plus} \\ \text{SUBCAT} & \langle\rangle \end{bmatrix} \\ \text{DTRS} & \begin{bmatrix} \textit{head-comp-struc} \\ \text{HEAD-DTR} \ \textit{word} \end{bmatrix} \end{bmatrix}$$

Schema 4 (Head-Marker Schema)

The DAUGHTERS value is an object of sort *head-marker-struc* whose HEAD-DAUGHTER | SYNSEM | NONLOCAL | TO-BIND | SLASH value is {}, and whose MARKER-DAUGHTER | SYNSEM | LOCAL | CATEGORY | HEAD value is of sort *marker*.

SCHEMA 4 (HEAD-MARKER SCHEMA) **formalized**

$$\left[\begin{array}{l}\text{DTRS} \left[\begin{array}{l}\textit{head-marker-struc} \\ \text{HEAD-DTR SS NONLOC TO-BIND SLASH } \{\} \\ \text{MARKER-DTR SS LOC CAT HEAD } \textit{marker}\end{array}\right]\end{array}\right]$$

SCHEMA 5 (HEAD-ADJUNCT SCHEMA)

The DAUGHTERS value is an object of sort *head-adjunct-struc* whose HEAD-DAUGHTER | SYNSEM value is token-identical to its ADJUNCT-DAUGHTER | SYNSEM | LOCAL | CATEGORY | HEAD | MOD value and whose HEAD-DAUGHTER | SYNSEM | NONLOCAL | TO-BIND | SLASH value is {}.

SCHEMA 5 (HEAD-ADJUNCT SCHEMA) **formalized**

$$\left[\begin{array}{l}\text{DTRS} \left[\begin{array}{l}\textit{head-adjunct-struc} \\ \text{HEAD-DTR SS} \quad \boxed{1}\left[\text{NONLOC TO-BIND SLASH } \{\}\right] \\ \text{ADJUNCT-DTR SS LOC CAT HEAD MOD } \boxed{1}\end{array}\right]\end{array}\right]$$

Seminar für Sprachwissenschaft
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

SCHEMA 6 (HEAD-FILLER SCHEMA)

The DAUGHTERS value is an object of sort *head-filler-struc* whose HEAD-DAUGHTER | SYNSEM | LOCAL | CATEGORY value satisfies the description [HEAD *verb*[VFORM *finite*], SUBCAT ⟨⟩], whose HEAD-DAUGHTER | SYNSEM | NONLOCAL | INHERITED | SLASH value contains an element token-identical to the FILLER-DAUGHTER | SYNSEM | LOCAL value, and whose HEAD-DAUGHTER | SYNSEM | NONLOCAL | TO-BIND | SLASH value contains only that element.

SCHEMA 6 (HEAD-FILLER SCHEMA) **formalized**

$$\left[\begin{array}{l}\text{DTRS} \left[\begin{array}{l}\textit{head-filler-struc} \\ \text{FILLER-DTR SS LOC } \boxed{1} \\ \text{HEAD-DTR SS} \left[\begin{array}{l}\text{LOC CAT} \left[\begin{array}{l}\text{HEAD VFORM } \textit{finite} \\ \text{SUBCAT } \langle\rangle\end{array}\right] \\ \text{NONLOC} \left[\begin{array}{l}\text{INHERITED SLASH } \boxed{2} \\ \text{TO-BIND SLASH } \{\boxed{1}\}\end{array}\right]\end{array}\right]\end{array}\right]\end{array}\right]$$

$$\wedge\ \texttt{member}(\boxed{1}, \boxed{2})$$

The ID PRINCIPLE

Every headed phrase must satisfy exactly one of the ID schemata.

$$\left[\text{DTRS } \textit{headed-struc}\right] \rightarrow$$

$$(\text{SCHEMA1} \vee \text{SCHEMA2} \vee \text{SCHEMA3} \vee \text{SCHEMA4} \vee \text{SCHEMA5} \vee \text{SCHEMA6})$$

HEAD FEATURE PRINCIPLE

In a headed phrase, the values of SYNSEM | LOCAL | CATEGORY | HEAD and DAUGHTERS | HEAD-DAUGHTER | SYNSEM | LOCAL | CATEGORY | HEAD are token-identical.

HEAD FEATURE PRINCIPLE **formalized**

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

SUBCATEGORIZATION PRINCIPLE

In a headed phrase, the list value of DAUGHTERS | HEAD-DAUGHTER | SYNSEM | LOCAL | CATEGORY | SUBCAT is the concatenation of the list value of SYNSEM | LOCAL | CATEGORY | SUBCAT with the list consisting of the SYNSEM values (in order) of the elements of the list value of DAUGHTERS | COMPLEMENT-DAUGHTERS.

SUBCATEGORIZATION PRINCIPLE **formalized**

$$\left[\begin{array}{l} \textit{phrase} \\ \text{DTRS}\ \textit{headed-struc} \end{array}\right] \rightarrow$$

$$\left[\begin{array}{l} \text{SYNSEM LOC CAT SUBCAT } \boxed{1} \\ \text{DTRS} \left[\begin{array}{l} \text{HEAD-DTR} \quad \text{SYNSEM LOC CAT SUBCAT } \boxed{2} \\ \text{COMP-DTRS } \boxed{3} \end{array}\right] \end{array}\right]$$

$$\wedge\ \texttt{sign-to-synsem}(\boxed{3},\boxed{4})$$

$$\wedge\ \texttt{append}(\boxed{1},\boxed{4},\boxed{2})$$

MARKING PRINCIPLE

In a headed phrase, the MARKING value is token-identical with that of the MARKER-DAUGHTER if any, and with that of the HEAD-DAUGHTER otherwise.

MARKING PRINCIPLE **formalized**

$$\left[\text{DTRS } \textit{headed-struc}\right] \rightarrow \left(\begin{array}{l} \left[\begin{array}{l}\text{SS LOC CAT MARKING } \boxed{1} \\ \text{DTRS MARKER-DTR SS LOC CAT MARKING } \boxed{1}\end{array}\right] \vee \\ \left(\left[\begin{array}{l}\text{SS LOC CAT MARKING } \boxed{1} \\ \text{DTRS HEAD-DTR SS LOC CAT MARKING } \boxed{1}\end{array}\right] \wedge \neg\left[\text{DTRS } \textit{head-mark-struc}\right]\right) \end{array} \right)$$