Parsing with HPSG

- Lecture 3-

Syntactic formalisms for natural language parsing

FI MU autumn 2011

Overview on syntactic formalisms

- Unification based grammars

  : HPSG, LFG, TAG, UCG...

- Dependency based grammars

  : Tesnière model; Meaning-Text of Mel'čuk...

- Application based grammars

  : CG, CCG, ACG, CTL...

Heritage of HPSG

- GPSG
  - linear order/ hierarchy order

    feature structure for representation of information
- LFG
  - Lexicon contains
  - Lexical rules
- CG
  - Subcategorization

Key points of HPSG

- Monostratal theory without derivation
  - Sharing a given information without movement and transformation
  - One representation for different levels of analysis : phonology, syntax, semantic
  - Constraint-based analyses
- Unification of given informations
- Computational formalism

Syntactic representation in HPSG

- **Typed feature structure**
  - consists of a couple “**attribute/value**”
  - the types are organized into a hierarchy

    ex: sign>phrase, case>nominative
  - feature structure is a directed acyclic graph (DAG), with arcs representing features going between values

Features

- Basic element of structure in HPSG
- Should be appropriate to a type

- Most frequently used features
  - PHON
  - SYNSEM
  - LOC/NON-LOC
  - CAT
  - CONTEXT
  - CONTENT
  - HEAD
  - SUJ
  - COMPS
  - S-ARG
  - QUANT
  - ....

Types

- Types are attributed to features -> typed features
  - *sign*
  - *synsem*
  - *head*
  - *phrase*
  - *content*
  - *Index*
  - *....*
- Each of these feature values is itself a complex object:
  - The type *sign* has the features PHON and SYNSEM appropriate for it
  - The feature SYNSEM has a value of type *synsem*
  - This type itself has relevant features (LOCAL and NON-LOCAL)

- *sign* is the basic type in HPSG used to describe lexical items (of type word) and phrases (of type phrase).

  - All signs carry the following two features:
    - PHON encodes the phonological representation of the sign
    - SYNSEM syntax and semantics

$$_{\textit{sign}}\begin{bmatrix} \textsf{PHON} & \text{list(phon-string)} \\ \textsf{SYNSEM} & \text{synsem} \end{bmatrix}$$

- In attribute-value matrix (AVM) form, here is the skeleton of an object:

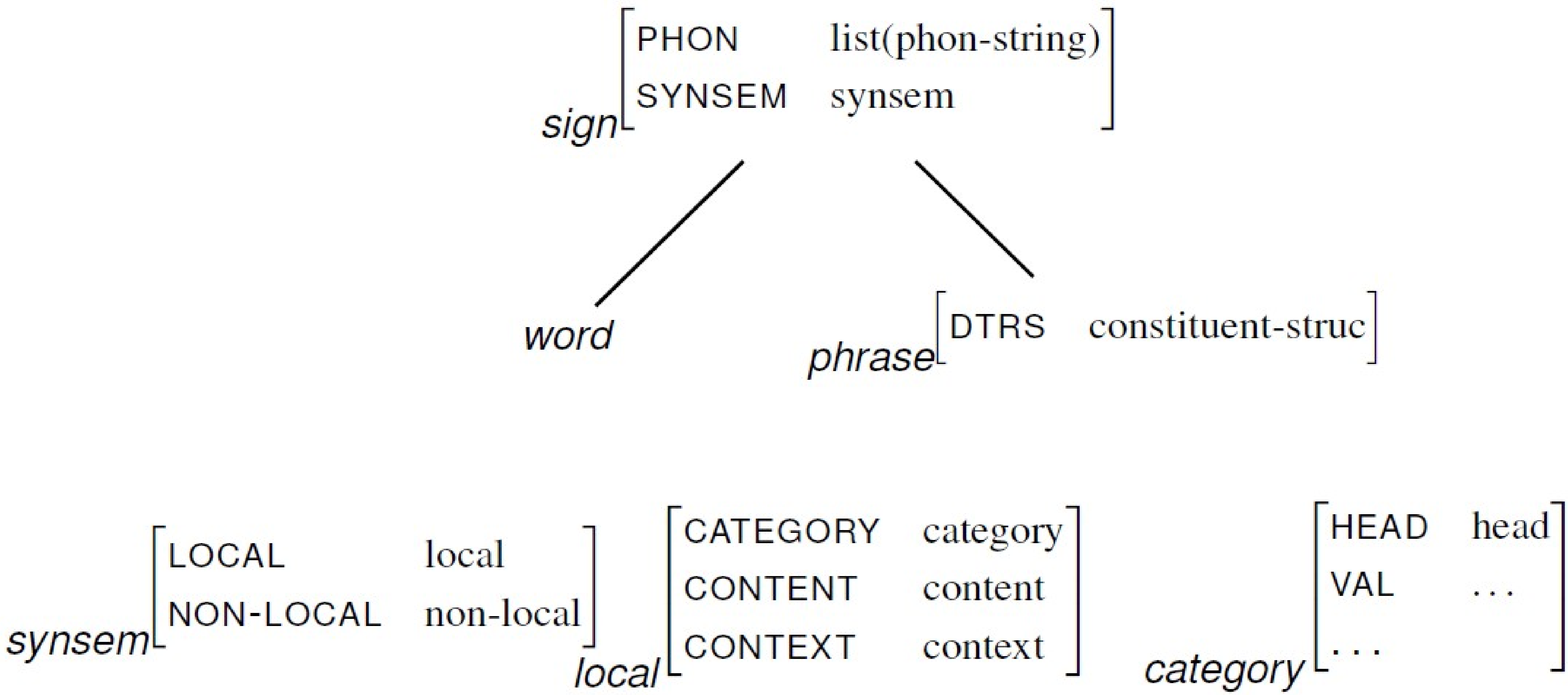
$$\begin{bmatrix} \textit{sign} & \\ \text{PHON} & \textit{list}\left(\text{PHON}\right) \\ \text{SYNSEM} & \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL} & \textit{local} \\ \text{NON-LOCAL} & \textit{non-local} \end{bmatrix} \\ \text{DTRS} & \textit{list}\left(\text{SIGN}\right) \end{bmatrix}$$

Structure of *signs* in HPSG

- *synsem* introduces the features LOCAL and NON-LOCAL
- *local* introduces CATEGORY (CAT), CONTENT (CONT) and CONTEXT(CONX)
- *non-local* will be discussed in connection with unbounded dependencies
- *category* includes the syntactic category and the grammatical argument of the word/phrase

Description of an object in HPSG:

lexical sign and phrasal sign

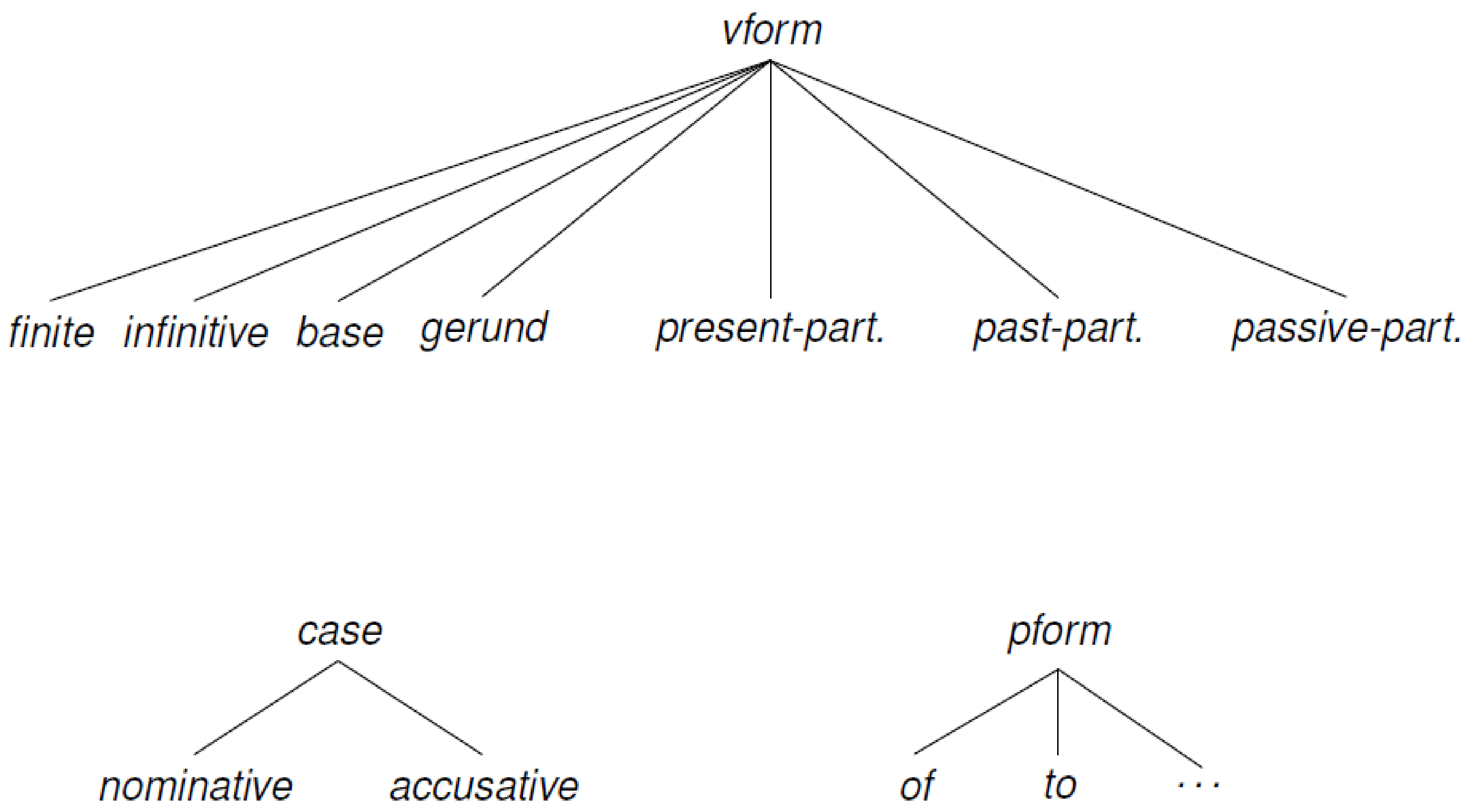


CATEGORY

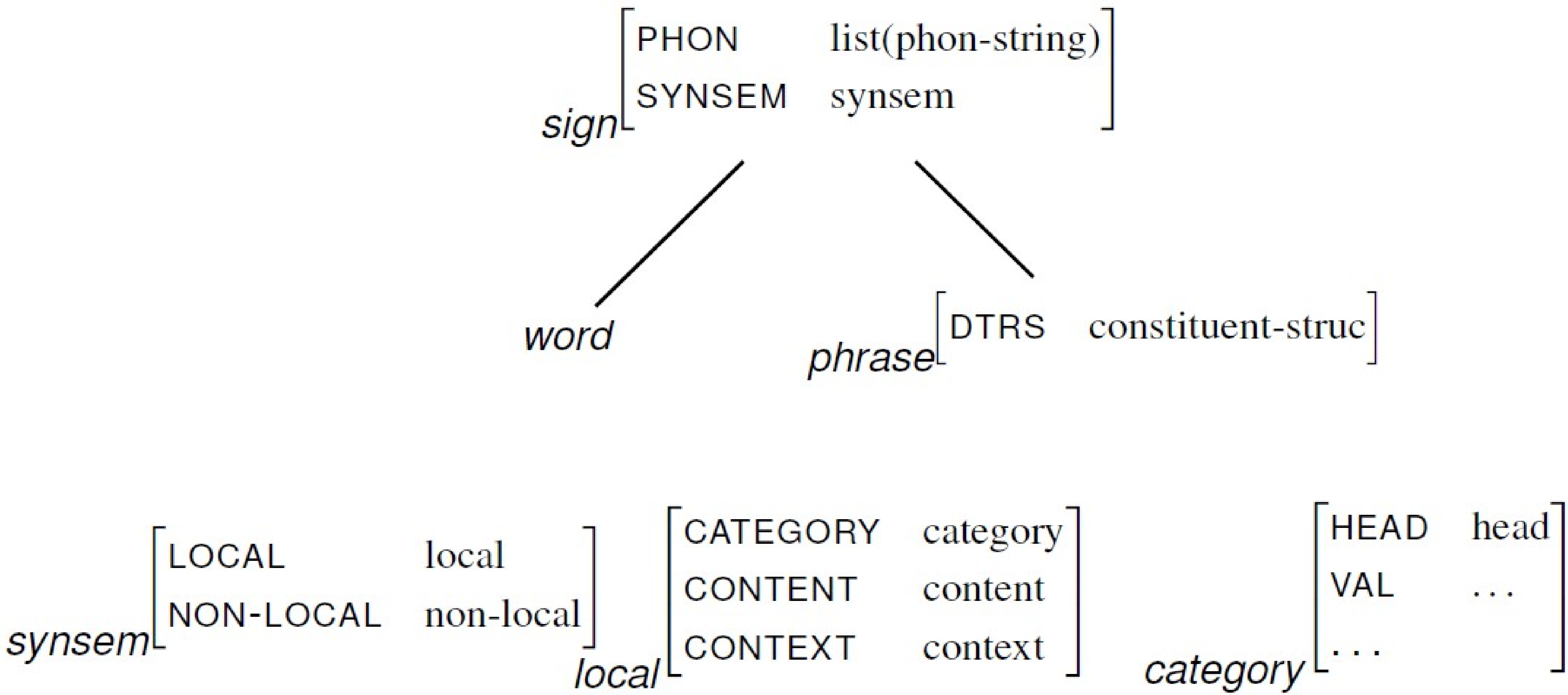
- CATEGORY encode the *sign*'s syntactic category
  - Given via the feature [HEAD head], where *head* is the supertype for noun, verb, adjective, preposition, determiner, marker; each of these types selects a particular set of head features
  - Given via the feature [VALENCE ...], possible to combine the *signs* with the other *signs* to a larger phrases

$$\left[\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{VALENCE} \;\; {}_{\textit{valence}}\!\begin{bmatrix} \text{SUBJECT} & \text{list(synsem)} \\ \text{SPECIFIER} & \text{list(synsem)} \\ \text{COMPLEMENTS} & \text{list(synsem)} \end{bmatrix}\right]$$

Sub-categorization of *head* type



Description of an object in HPSG

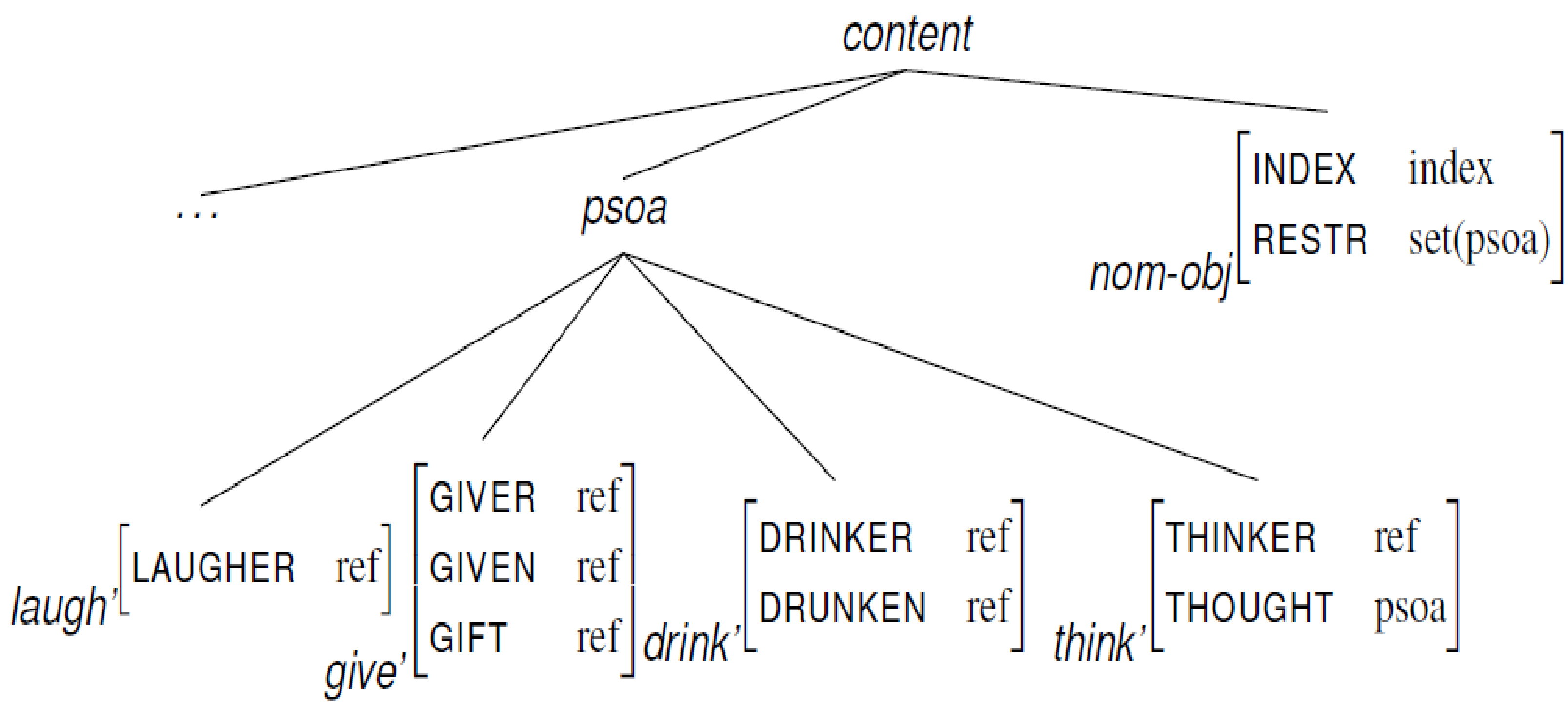


Semantic representation : CONTENT (& CONTEXT) feature

- Semantic interpretation of the *sign* is given as the value to CONTENT
  - *nominal-object*: an individual/entity (or a set of them), associated with a referring index, bearing agreement features → INDEX, RESTR
  - *Parameterized-state-of-affairs (psoa)*: a partial situate; an event relation along with role names for identifying the participants of the event→ BACKGR
  - *quantifier*: *some, all, every, a, the*, . . .

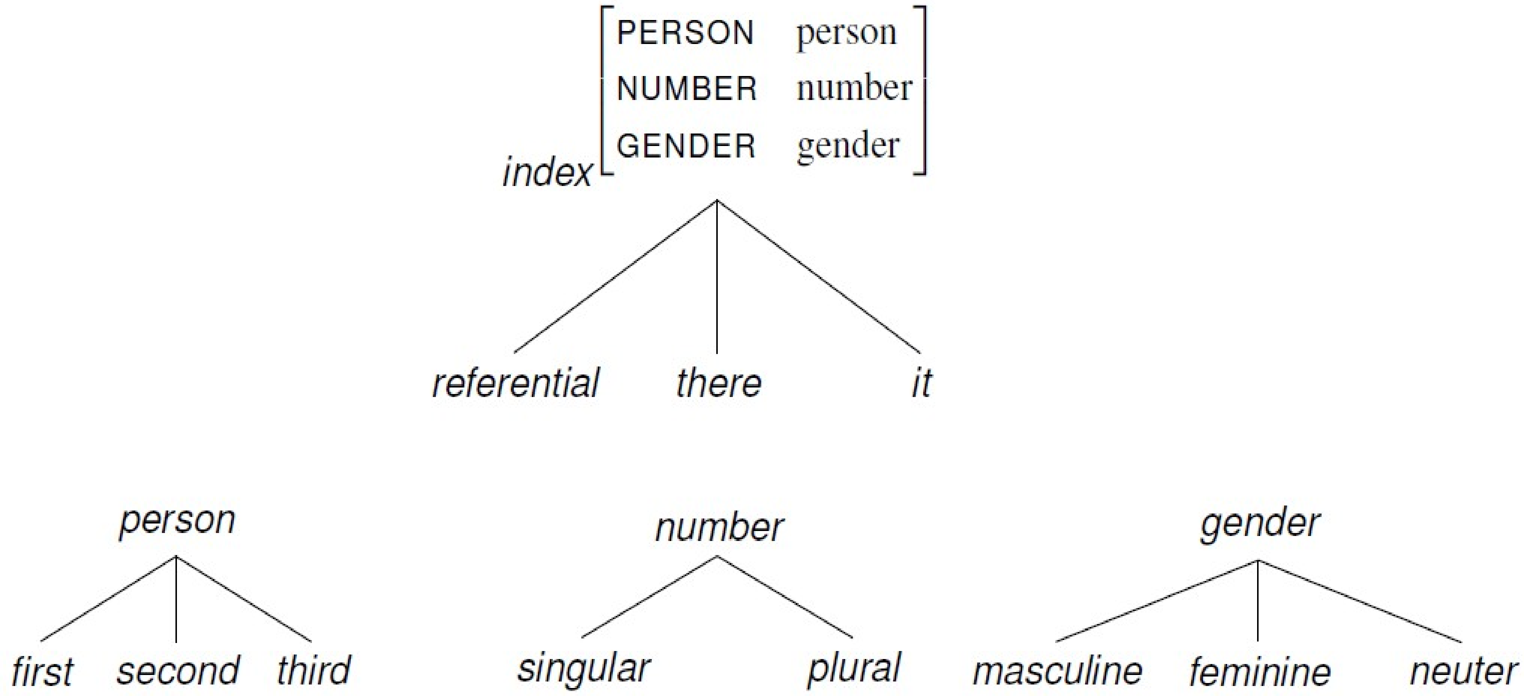
- Note: many of these have been reformulated by “Minimal Recursion Semantics (MRS)” which allows underspecification of quantifier scopes, though a in-depth discussion of MRS is beyond the scope of this class

Sub-categorization of *content* type

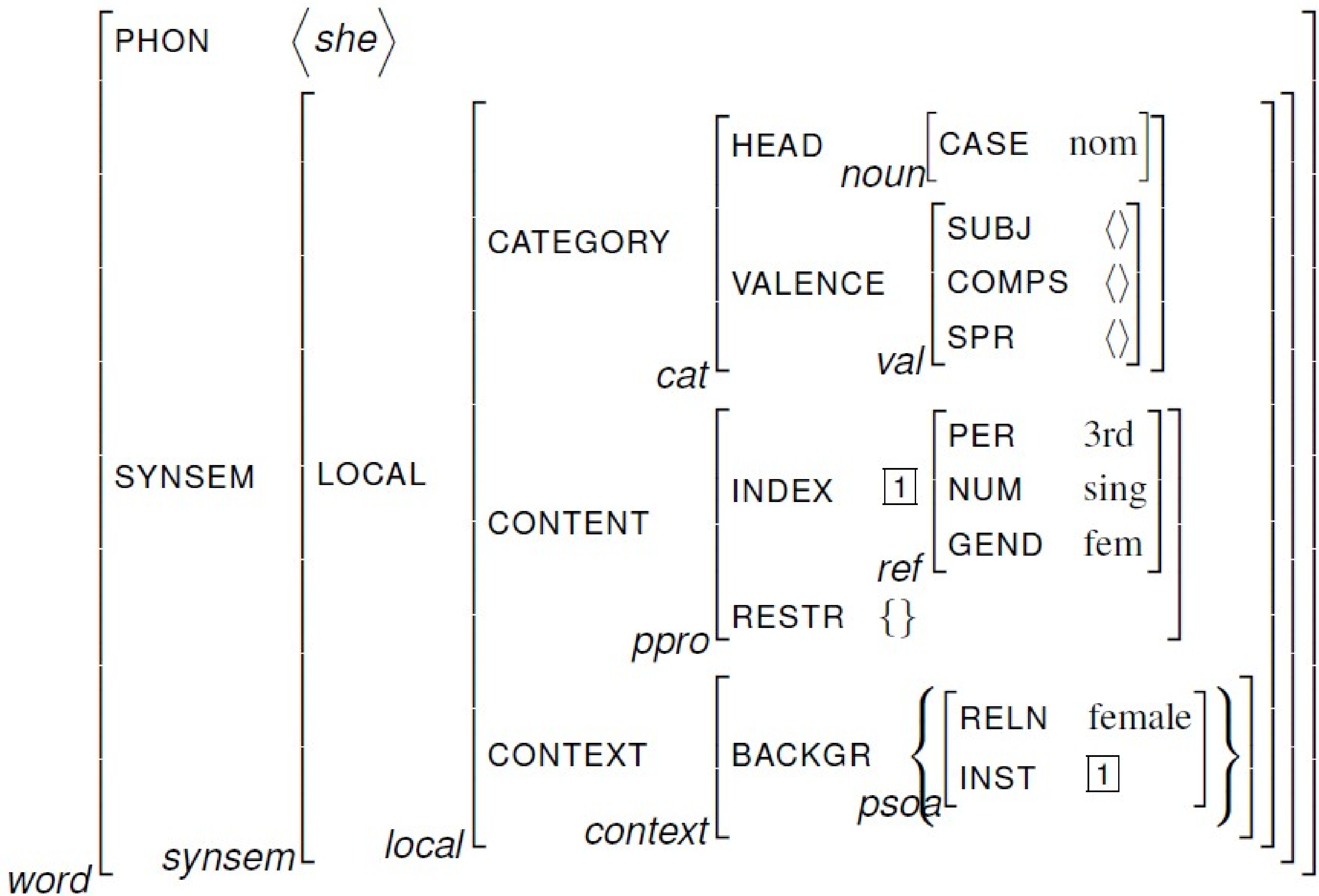


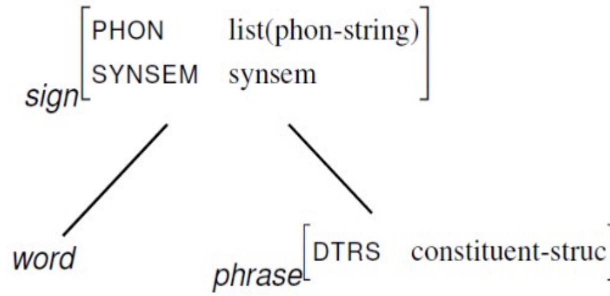
Note: Semantic restriction on the index are represented as a value of RESTR. RESTR is an attribute of a nominal object. The value of RESTR is a set of psoa. In turn, RESTR has the attribute of REL whose value can either be referential indices or psoas.

Sub-categorization of *index* type



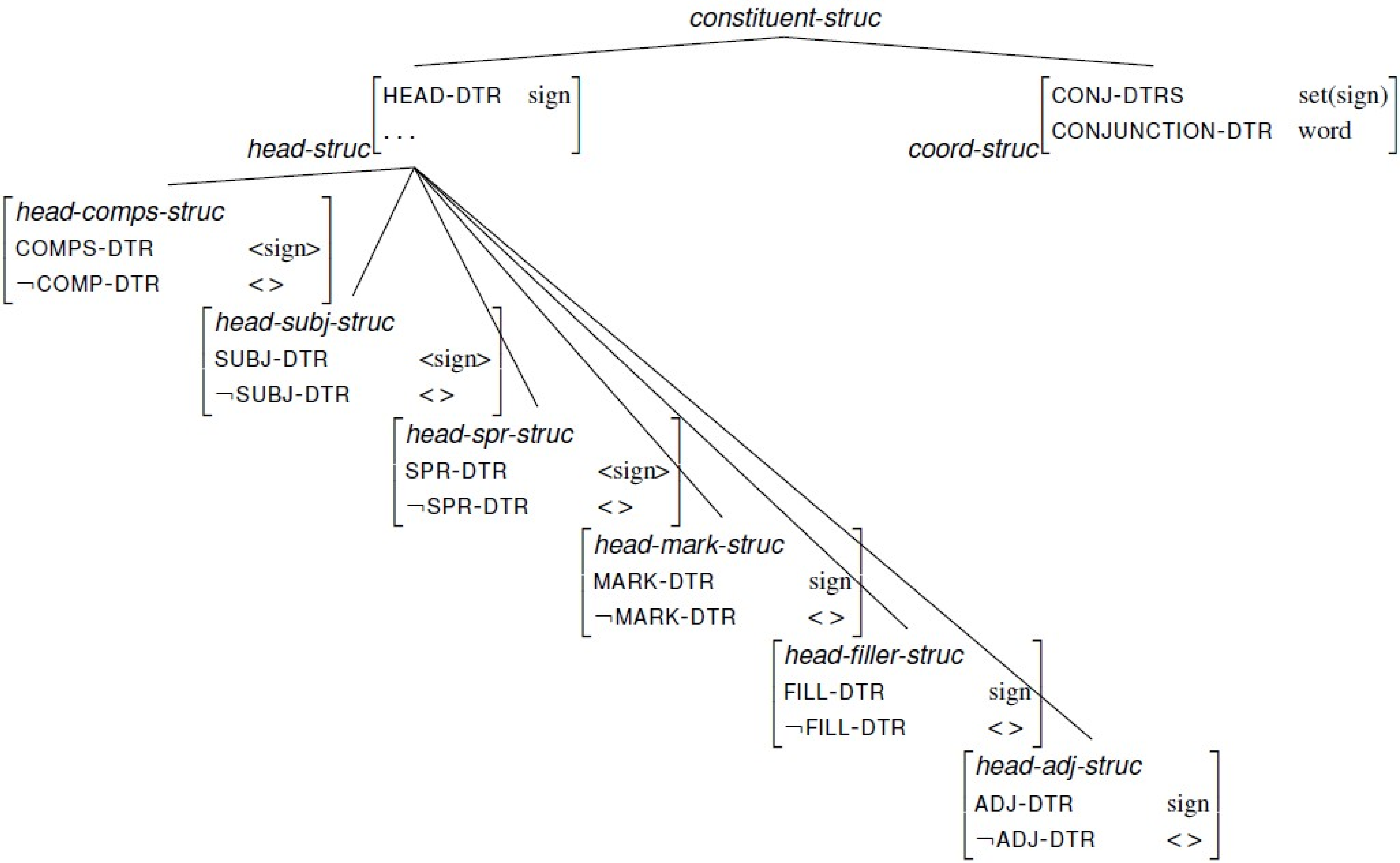
Lexical input of *She*



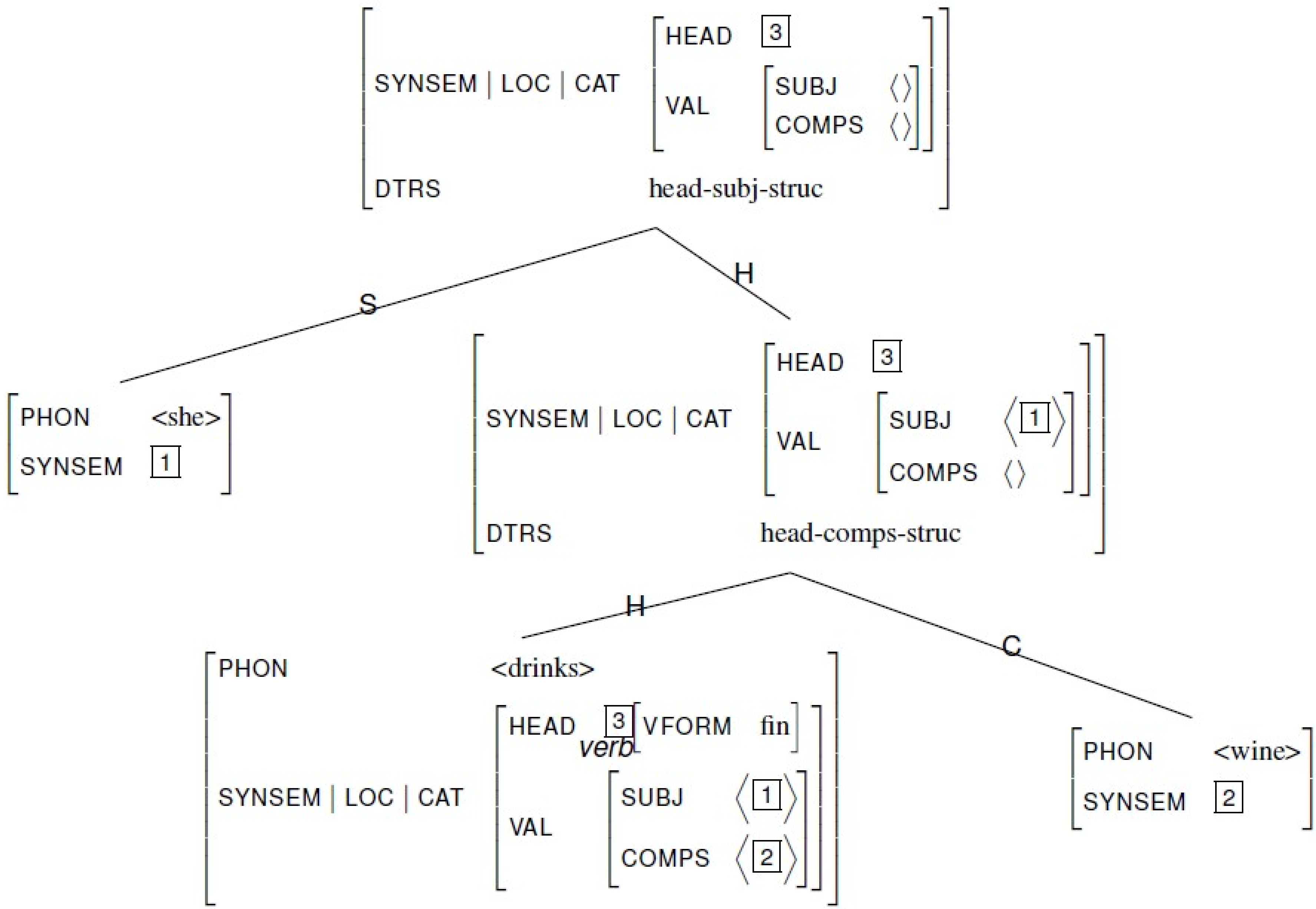


- Each *phrase* has a DTRS attribute which has a <u>constituent-structure</u> value
- This DTRS value corresponds to what we view in a tree as daughters (with additional grammatical role information, e.g. adjunct, complement, etc.)
- By distinguishing different kinds of <u>constituent-structures</u>,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

Structure of *phrase*



head-subject/complement structure



Questions! (1)

- 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

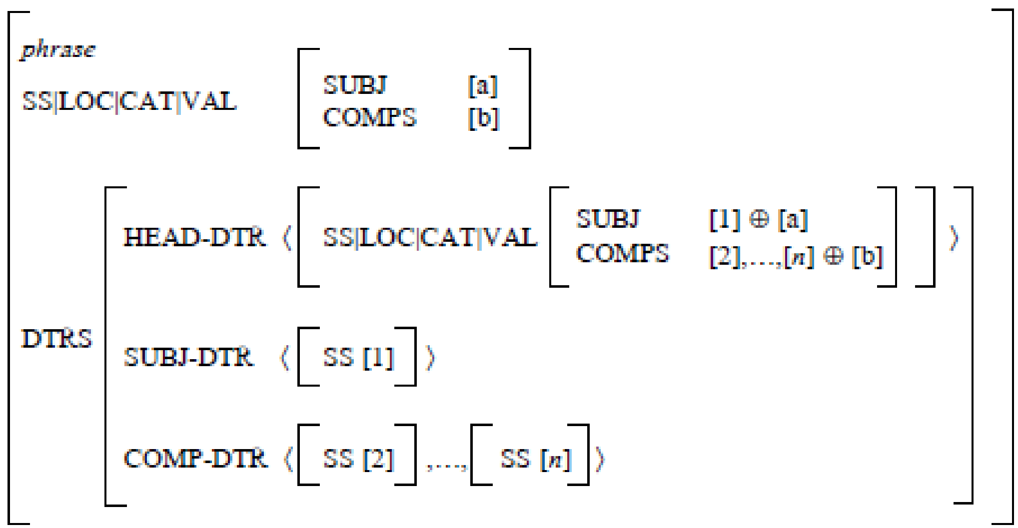
HEAD-feature principle

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

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

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* (Pollard and Sag, 1994:348)



- Note:

  Valence Principle constrains the way in which information is shared between phrases and their head daughters.

  – F can be any one of SUBJ, COMPS, SPR
  – When the F-DTR is empty, the F valence feature of the head daughter will be copied to the mother phrase

Questions! (2)

- 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?

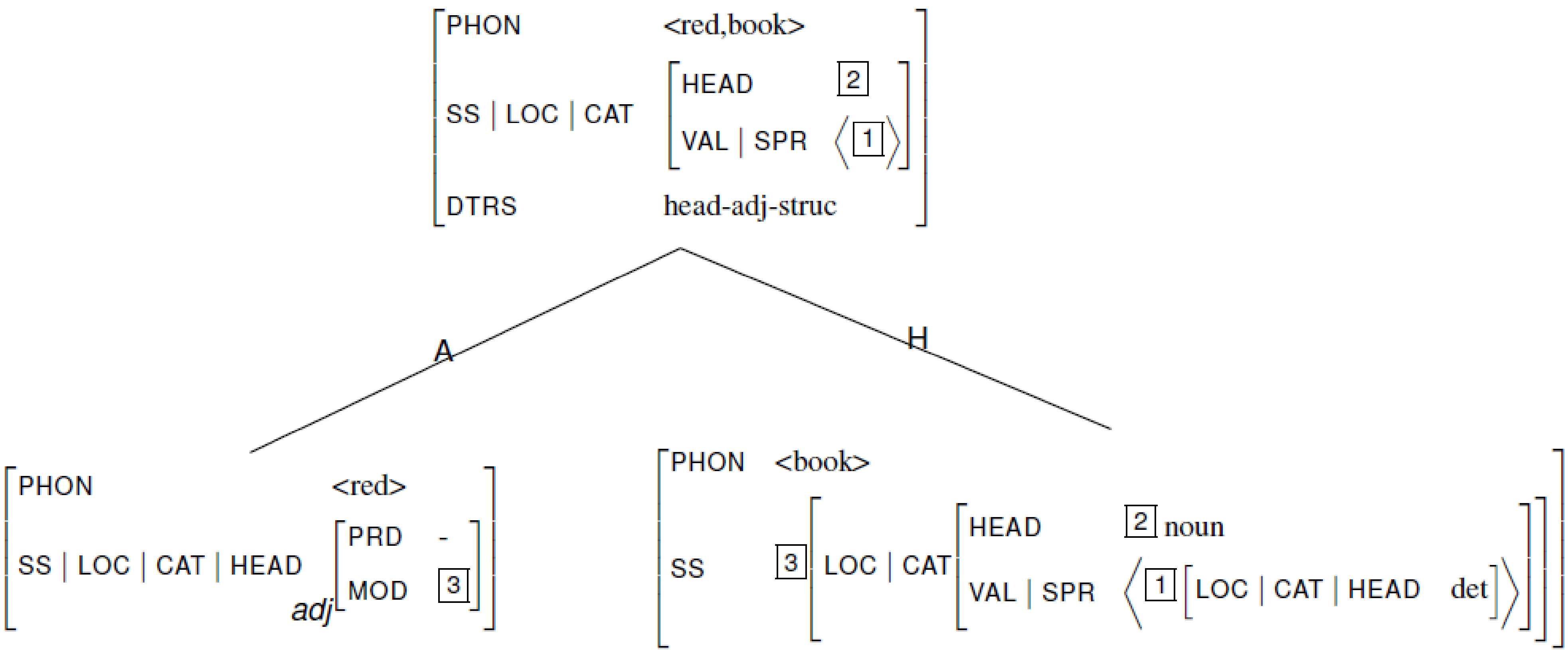
Immediate Dominance (ID) Principle

- 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 (ID) Schemata

$$
\begin{array}{lll}
\textit{phrase}\begin{bmatrix}\text{DTRS} & \text{head-struc}\end{bmatrix} \rightarrow & \begin{bmatrix}\text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{COMPS} & \langle\rangle \\ \text{DTRS} & \text{head-subj-struc}\end{bmatrix} & \text{(head-subject)} \\
\lor & \begin{bmatrix}\text{DTRS} & \text{head-comps-struc}\end{bmatrix} & \text{(head-complement)} \\
\lor & \begin{bmatrix}\text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} \mid \text{COMPS} & \langle\rangle \\ \text{DTRS} & \text{head-spr-struc}\end{bmatrix} & \text{(head-specifier)} \\
\lor & \begin{bmatrix}\text{DTRS} & \begin{bmatrix}\textit{head-marker-struc} & \\ \text{MARK-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \text{marker}\end{bmatrix}\end{bmatrix} & \text{(head-marker)} \\
\lor & \begin{bmatrix}\text{DTRS} & \begin{bmatrix}\textit{head-adj-struc} & \\ \text{ADJ-DTR} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \mid \text{MOD} & \boxed{1} \\ \text{HEAD-DTR} \mid \text{SS} & \boxed{1}\end{bmatrix}\end{bmatrix} & \text{(head-adjunct)} \\
\lor & \ldots &
\end{array}
$$

head-adjunct structure



Semantic principle

- The CONTENT value of a headed phrase is token identical to the CONTENT value of the semantic head daughter
- The semantic head daughter is identified as
  - The ADJ-DTR in a head-adjunct phrase
  - The HEAD-DTR in other headed phrases

$$
\mathit{phrase}\begin{bmatrix}\text{DTRS} & \text{head-struc}\end{bmatrix} \rightarrow
\begin{bmatrix}
\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONT} & \boxed{1} \\
\text{DTRS} & \begin{bmatrix}\mathit{head\text{-}adj\text{-}struc} \\ \text{ADJ-DTR}\,|\,\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONT} & \boxed{1}\end{bmatrix}
\end{bmatrix} \quad \text{(head-adjunct)}
$$

$$
\vee \begin{bmatrix}
\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONT} & \boxed{1} \\
\text{DTRS} & \begin{bmatrix}\neg\ \mathit{head\text{-}adj\text{-}struc} \\ \text{HEAD-DTR}\,|\,\text{SYNSEM}\,|\,\text{LOC}\,|\,\text{CONT} & \boxed{1}\end{bmatrix}
\end{bmatrix} \quad \text{(non-head-adjunct)}
$$

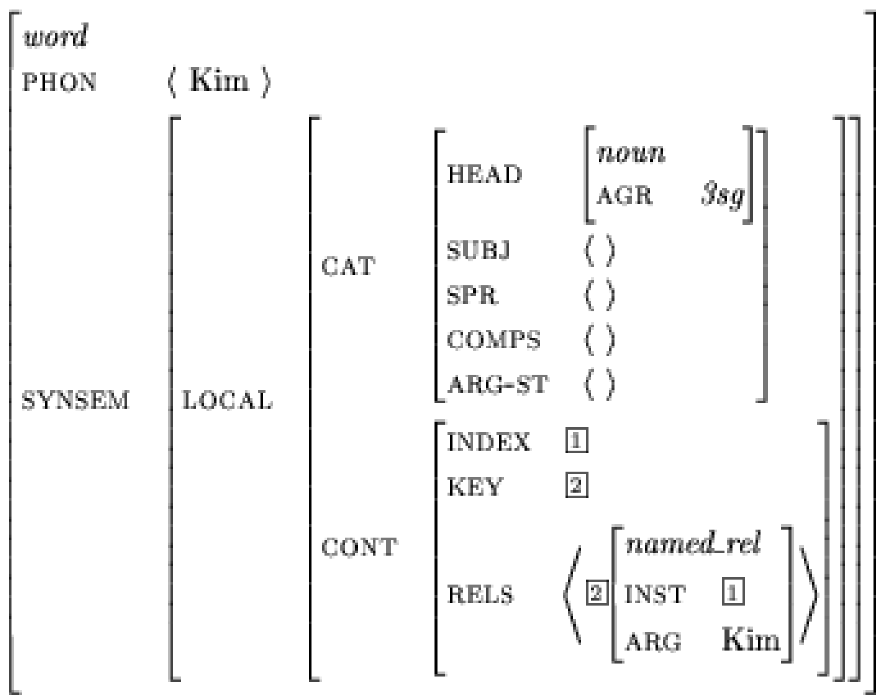
SPEC principle

- In a headed phrase whose non-head daughter (either the MARK-DTR or COMP-DTR|FIRST) has a SYNSEM|LOCAL|CATEGORY|HEAD value of type functional, the *spec* value of that value must be token-identical with the phrase's DTRS|HEAD-DTR|SYNSEM value

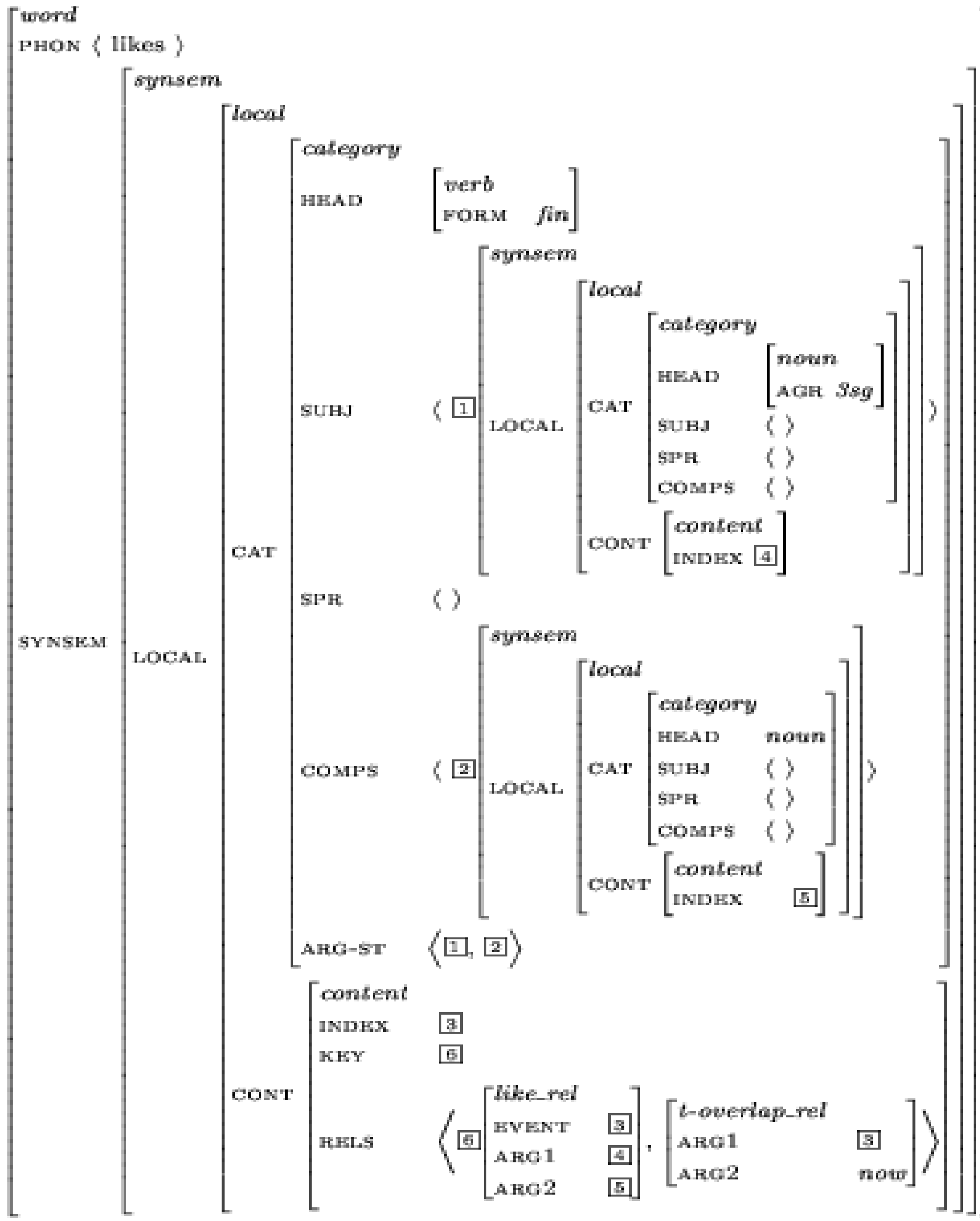
$$\left[_{phrase}\ \text{DTRS} \begin{bmatrix} \textit{head-struc} \\ \text{NONHEAD-DTR} \mid \text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{HEAD} \quad \text{functional} \end{bmatrix}\right] \rightarrow \left[\text{DTRS} \begin{bmatrix} \text{HEAD-DTR} \mid \text{SYNSEM} & \boxed{1} \\ \text{NONHEAD-DTR} \mid \text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{HEAD} \mid \text{SPEC} & \boxed{1} \end{bmatrix}\right]$$

Example 2

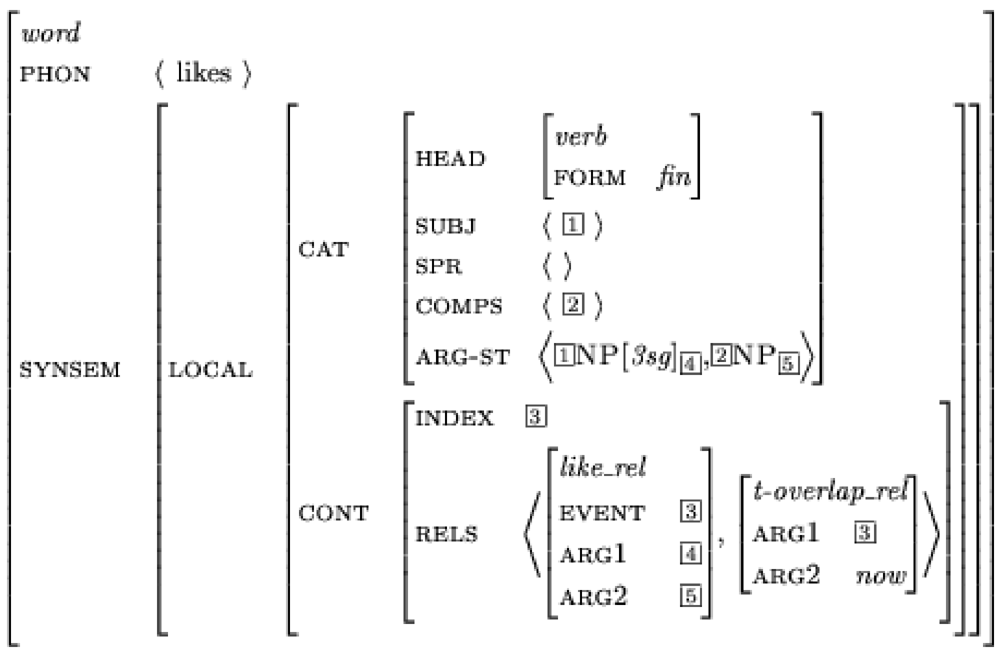
Kim *likes bagels*



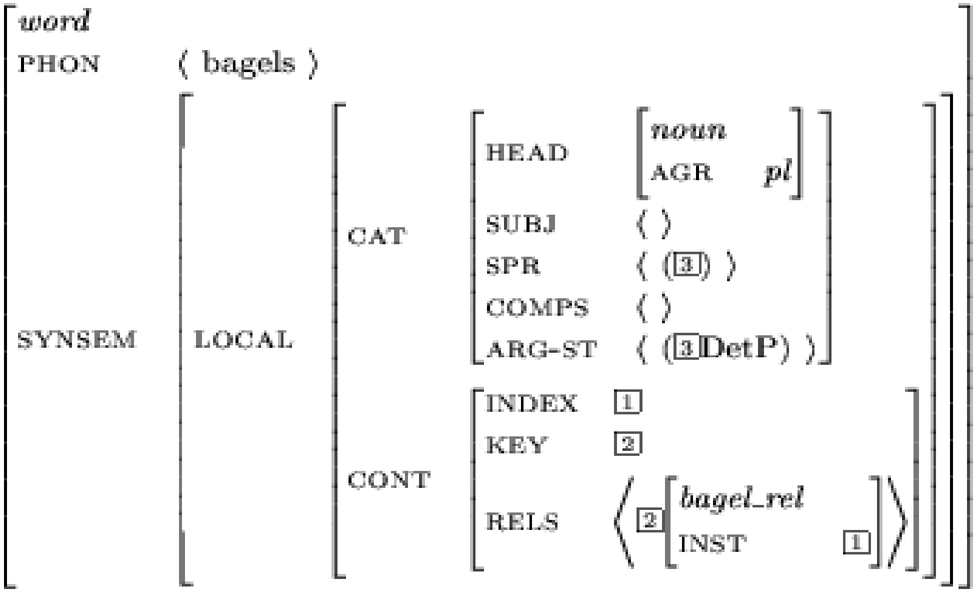
Kim ***likes(1)*** *bagels*



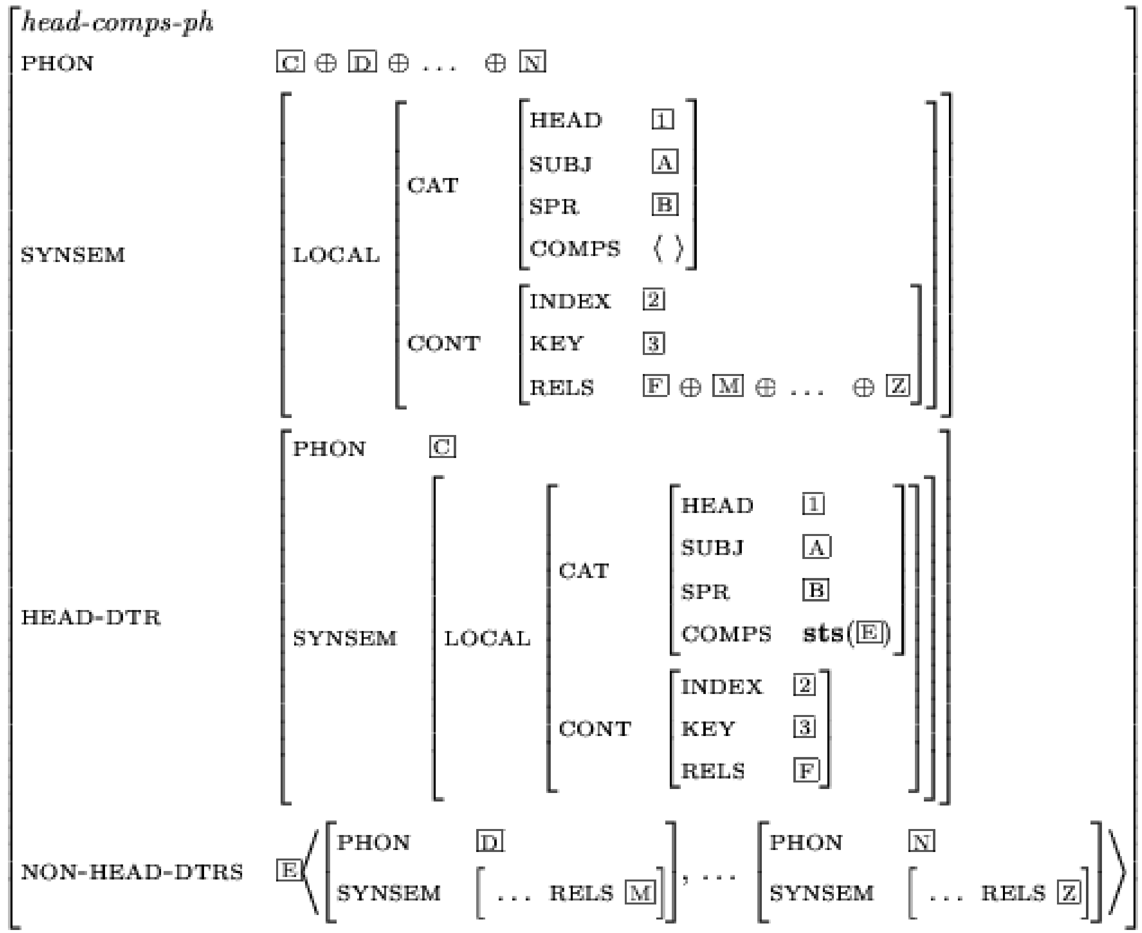
Kim ***likes(2)*** *bagels*



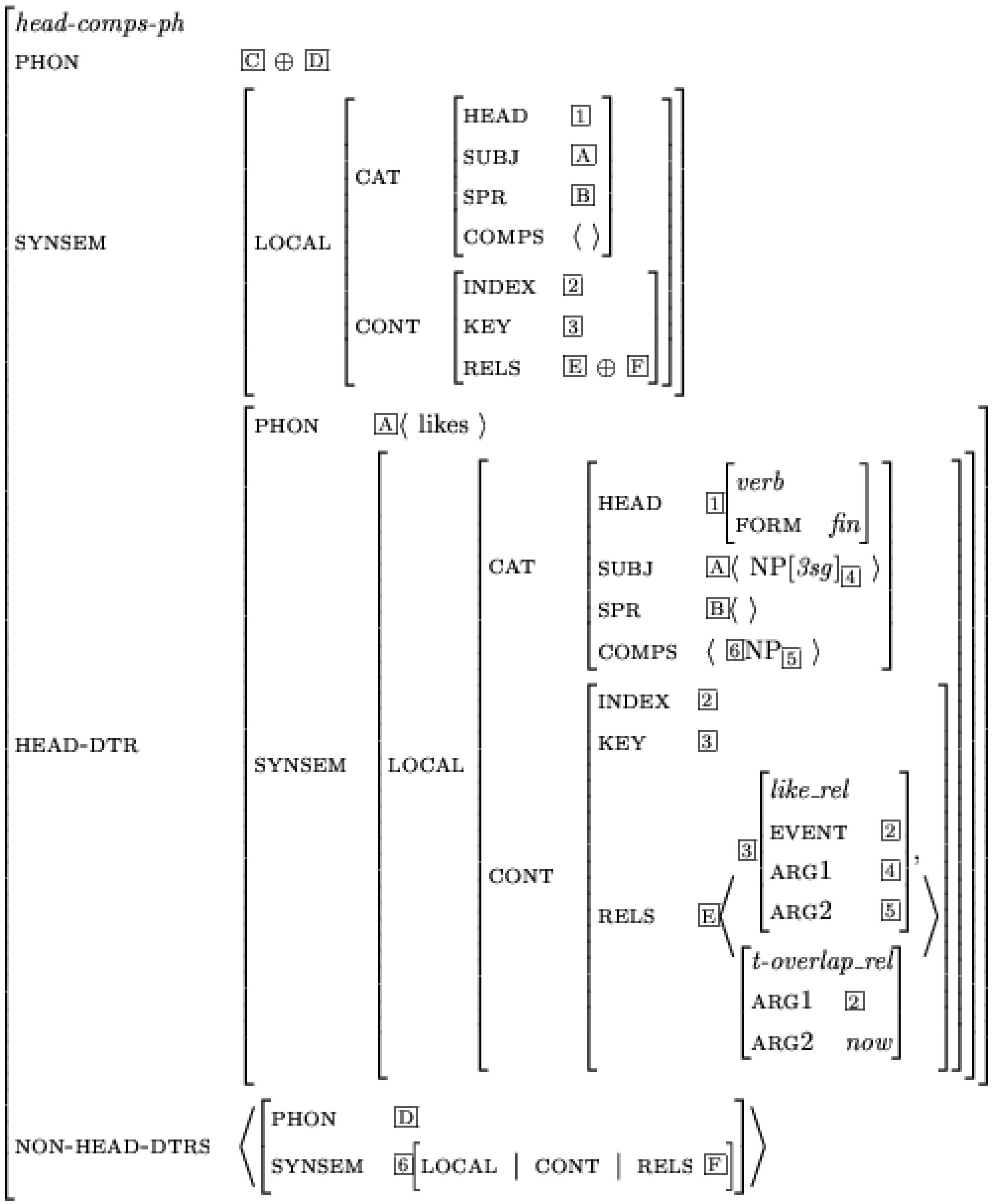
Kim likes ***bagels***



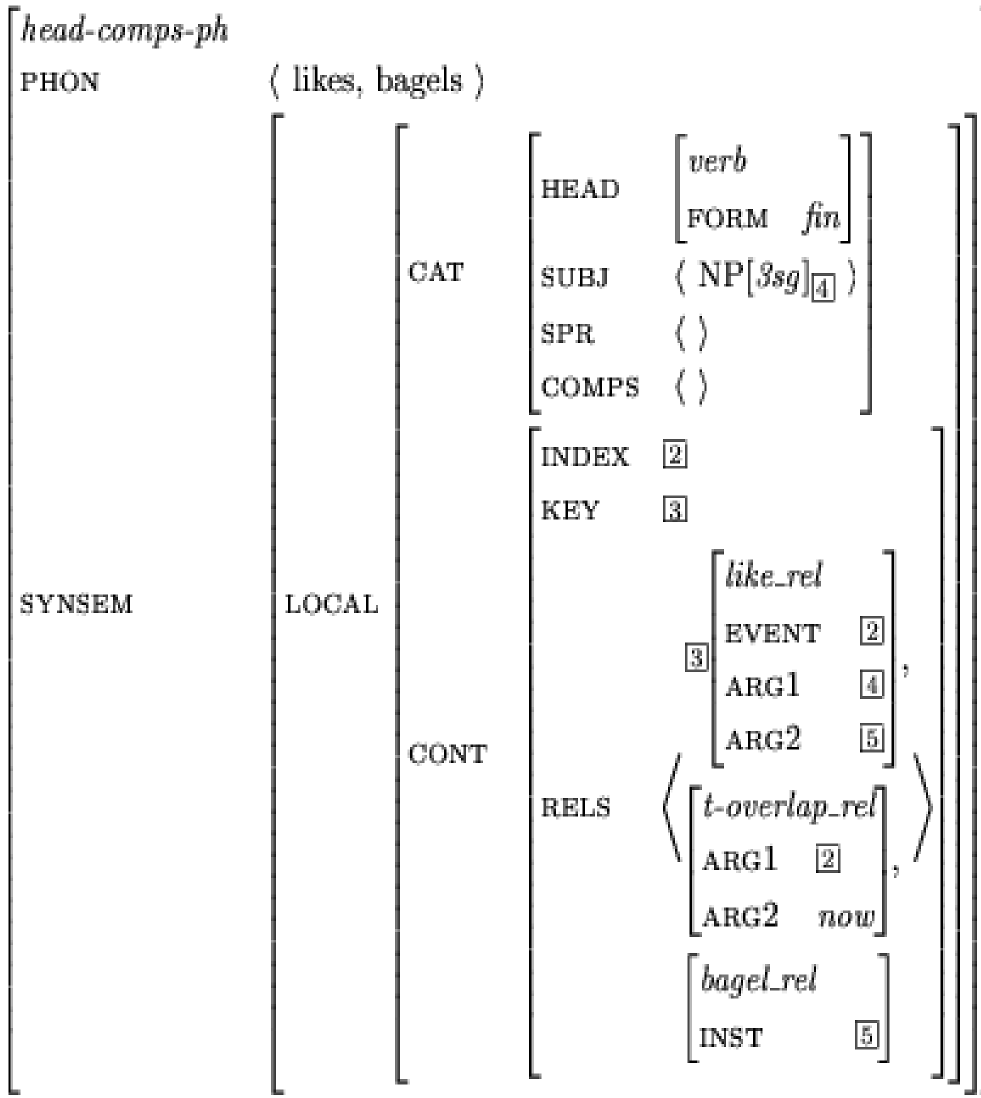
- *head-complement* schema



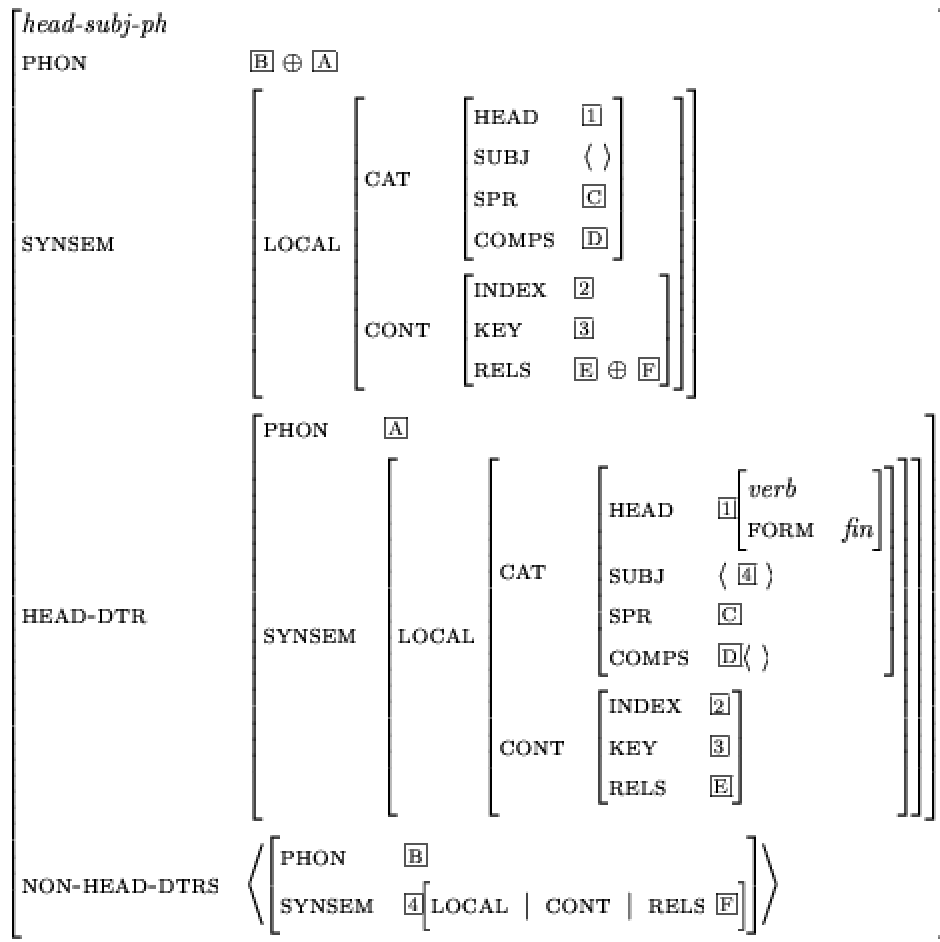
- *head-complement* schema headed by *likes*



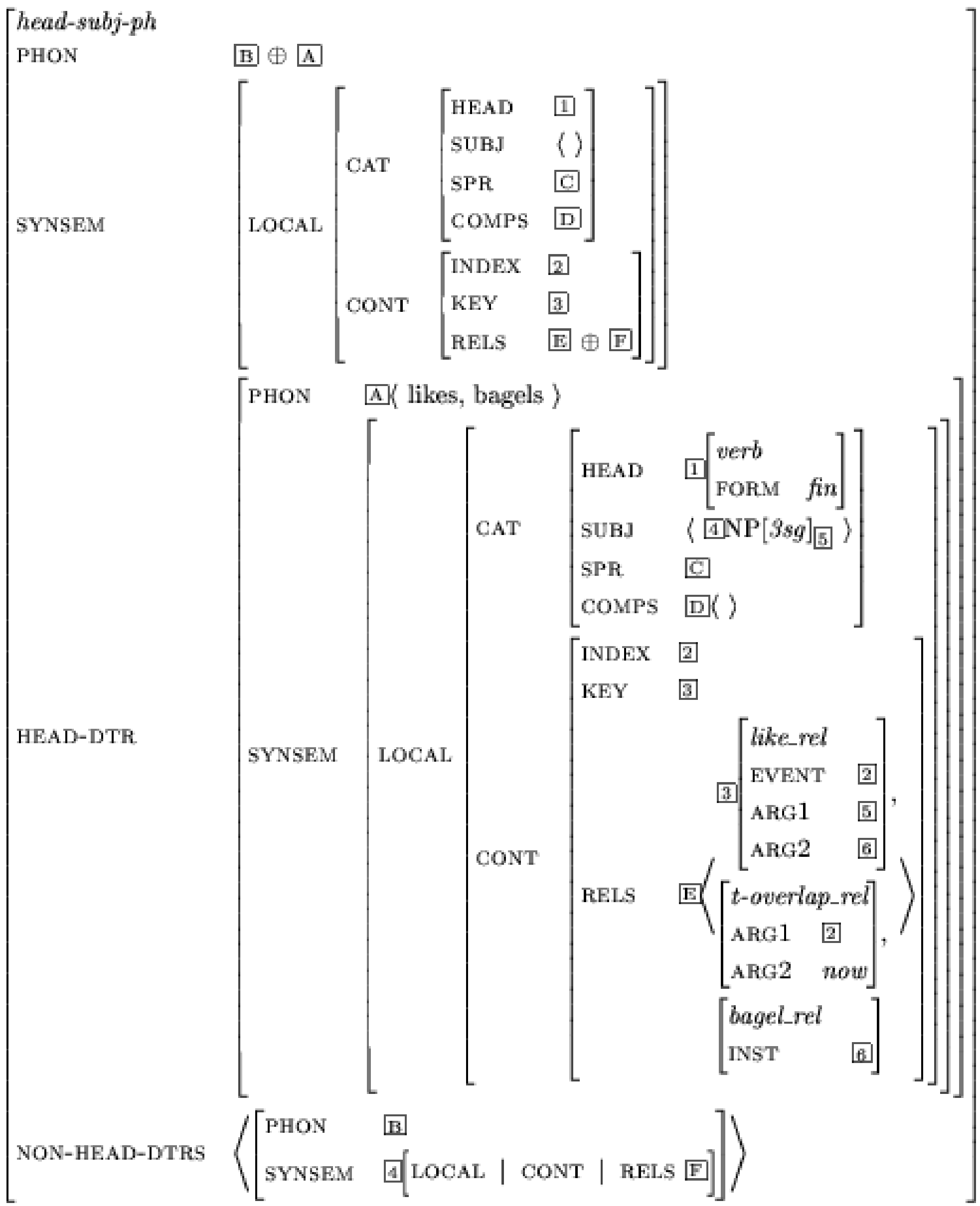
Kim ***likes bagels***



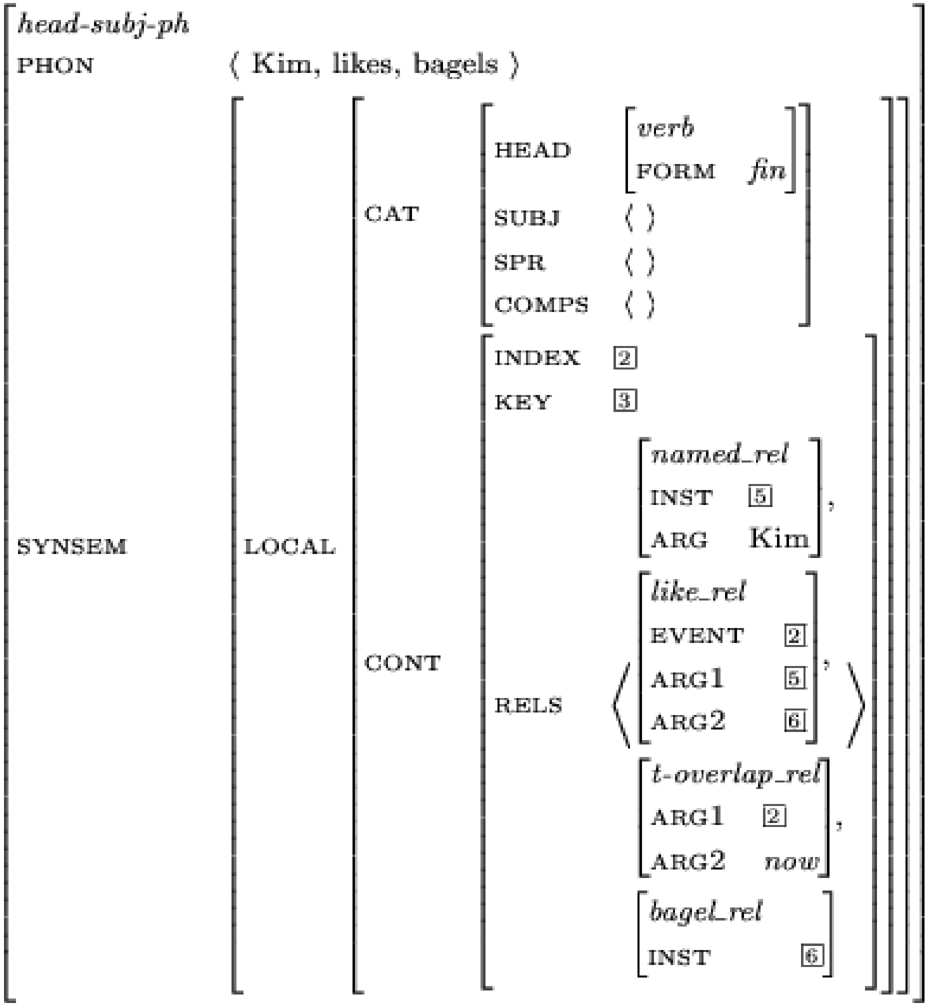
- *head-subject schema*



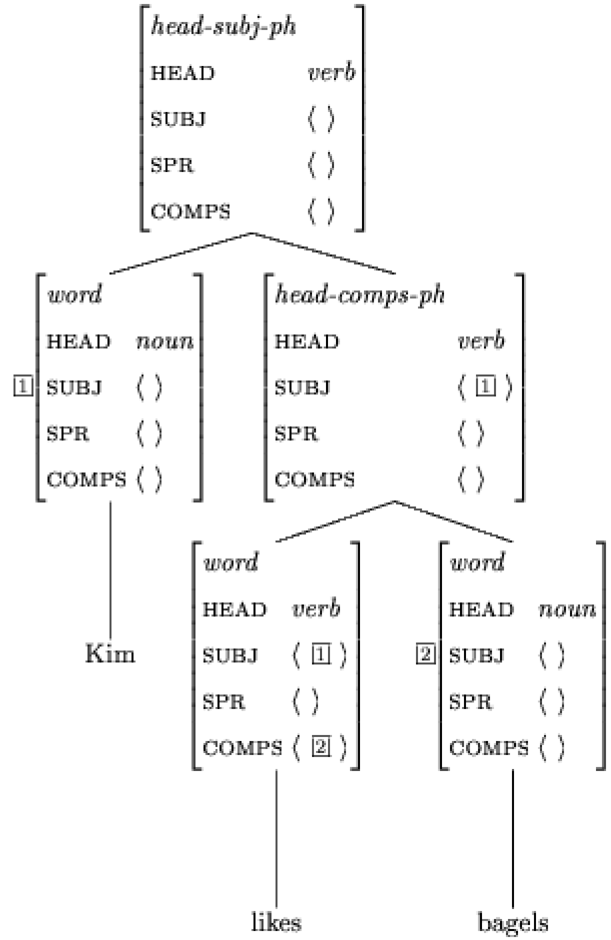
- *head-subject* schema headed by *likes bagels*



Kim likes bagels



Tree of *Kim likes bagels*



Compare HPSG to CFG

- Each sign or HPSG rule consists of SYNSEM, DTRS, and PHON parts.
- The SYNSEM part specifies how the syntax and semantics of the phrase (or word) are constrained. It corresponds roughly to the **left-hand side of CFG rules** but contains much more information.
- The DTRS part specifies the constituents that make up the phrase (if it is a phrase). (Each of these constituents is a complete sign.) This corresponds to part of the information on the **right-hand side of CFG rules**, but not to ordering information.
- The PHON part specifies the ordering of the constituents in DTRS (where this is constrained) and the pronunciation of these (if this is specifiable). This corresponds to the the ordering information on the **right-hand side of CFG rules**.

Simulation of Bottom-up parsing algorithm in HPSG

- Unify input lexical-signs with lexical-signs in the lexicon.
- Until no more such unifications are possible
  - Unify instantiated signs with the daughters of instantiated phrasal signs or with phrasal signs in the grammar.

if

all instantiated signs but one saturated one (S) are associated with daughters of other instantiated signs and the PHON value of all instantiated signs is completely specified

return the complete S structure

else fail.

Example 2: processing of unification

Kim walks

The words in the sentence specify only their pronunciations and their positions.

1 [PHON ((0 1 kim))]

2 [PHON ((1 2 walks))]

STEP 1: Unifying 1 with the lexical entry for *Kim* gives

3 [PHON ((0 1 kim))
SYNSEM [CAT [HEAD noun SUBCAT ()]
CONTENT [INDEX 1 [PER 3rd NUM sing]]
CONTEXT [BACKGR {[RELN naming BEARER 1 NAME Kim]}]]]

We now know something about the meaning of *Kim* (it refers to somebody named *Kim*) and something about its syntactic properties (it is third person singular).

1 [PHON ((0 1 kim))]
2 [PHON ((1 2 walks))]

STEP 2: Unifying 2 with the lexical entry for *walks* gives

4 [PHON ((1 2 walks))
SYNSEM [CAT [HEAD [VFORM fin]
SUBCAT ([CAT [HEAD noun SUBCAT ()]
CONTENT [INDEX 1 [PER 3rd NUM sing]]])]
CONTENT [RELN walk WALKER 1]]]

We know that *walks* refers to walking and that it requires a subject noun phrase which refers to the walker but doesn't require any object.

```
HEAD-DTR rule

[SYNSEM [CAT [HEAD 1 SUBCAT (2)]
         CONTENT 4]
 DTRS [HEAD-DTR [SYNSEM [CAT [HEAD 1 SUBCAT (2)]
                         CONTENT 4]
                 PHON 3]
       SUBJ-DTRS ()]
 PHON 3]
```

STEP 3: Unifying 4 with the HEAD-DTR of this rule gives

```
5 [SYNSEM [CAT [HEAD [VFORM fin]
                SUBCAT 2([CAT [HEAD noun SUBCAT ()]
                          CONTENT [INDEX 1 [PER 3rd NUM sing]]]])]
           CONTENT 4[RELN walk WALKER 1]]
  DTRS [HEAD-DTR [SYNSEM [CAT [HEAD [VFORM fin] SUBCAT (2)]]
                          CONTENT [4]]
                          PHON 3((1  2 walks))]
           SUBJ-DTRS ()]
  PHON 3((1  2 walks))]
```

Now we have a VP with the transitive verb *walks* as its head (and only constituent).

```
HEAD-DTR rule

6 [SYNSEM [CAT [HEAD 1 SUBCAT ()]
           CONTENT 4]
   DTRS [HEAD-DTR [SYNSEM [CAT [HEAD 1 SUBCAT (2)]
                           CONTENT 4]
                   PHON 3]
         SUBJ-DTRS ([PHON 5
                     SYNSEM 2]))]
   PHON (5 < 3)]
```

STEP 4: Unifying 5 with the HEAD-DTR of this rule gives

```
7 [SYNSEM [CAT [HEAD 1[VFORM fin SUBCAT ()]]
           CONTENT 4[RELN walk WALKER ]]
   DTRS [HEAD-DTR [SYNSEM [CAT [HEAD 1[VFORM fin]
                                SUBCAT 2([CAT [HEAD noun SUBCAT ()]
                                          CONTENT [INDEX
                                                   [PER 3rd NUM sing]]]])]
                           CONTENT [RELN walk WALKER 4]]
                   PHON 3((1 2 walks))]
         SUBJ-DTRS ([PHON 5
                     SYNSEM 2[CAT [HEAD noun SUBCAT ()]
                              CONTENT [INDEX ]]]])]
   PHON (5 < 3((1  2 walks)))]
```

STEP 5: Unifying 3 with the SUBJ-DTR of 7 gives

```
8 [SYNSEM [CAT [HEAD [VFORM fin SUBCAT ()]]
                     CONTENT [RELN walk WALKER [PER 3rd NUM sing]]]
   DTRS [HEAD-DTR [SYNSEM [CAT [HEAD [VFORM fin]
                                                   SUBCAT ([CAT [HEAD noun SUBCAT ()]
                                                                   CONTENT [INDEX [PER 3rd NUM sing]]]])
                                          CONTENT [RELN walk WALKER [PER 3rd NUM sing]]]
                            PHON ((1 2 walks))]
           SUBJ-DTRS ([PHON ((0 1 kim))
                                SYNSEM [CAT [HEAD noun SUBCAT ()]
                                               CONTENT [INDEX [PER 3rd NUM sing]]]])]
    PHON ((0 1 kim) (1 2 walks))]
```

Now the subject of the sentence is pronounceable, and we're done.

Phenomena covered by HPSG parsers

- Case assignment
- Word order : scrambling
- Long distance dependency
- Coordination
- Scope of adverbs and negation
- Topic drop
- Agreement
- Relative clause
- ...

Example 3:
unbounded dependency construction

- An unbounded dependency construction
  - involves constituents with different functions
  - involves constituents of different categories
  - is in principle unbounded

- Two kind of unbounded dependency constructions (UDCs)
  - Strong UDCs
  - Weak UDCs

Strong UDCs

- An overt constituent occurs in a non-argument position:
  - **Topicalization:**

    Kim$_i$, Sandy loves_$_i$.
  - **Wh-questions:**

    I wonder [who$_i$ Sandy loves_ $_i$].
  - **Wh-relative clauses:**

    This is the politician [who$_i$ Sandy loves_ $_i$].
  - **It -clefts:**

    It is Kim$_i$ [who$_i$ Sandy loves_ $_i$].
  - **Pseudoclefts:**

    [What$_i$ Sandy loves_ $_i$] is Kim$_i$.

Weak UDCs

- No overt constituent in a non-argument position:
  - Purpose infinitive (for -to clauses):

    I bought it $_i$ *for Sandy to eat_* $_i$ *.*
  - Tough movement:

    Sandy $_i$ *is hard to love_* $_i$ *.*
  - Relative clause without overt relative pronoun:

    This is [the politician] $_i$ *[Sandy loves_* $_i$ *].*
  - It -clefts without overt relative pronoun:

    It is Kim $_i$ *[Sandy loves_* $_i$ *].*

Using the feature SLASH

- To account for UDCs, we will use the feature SLASH (so-named because it comes from notation like S/NP to mean an S missing an NP)
- This is a non-local feature which originates with a trace, gets passed up the tree, and is finally bound by a filler

The bottom of a UDC: Traces

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle\rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOCAL} & \boxed{1} \\ \text{NONLOC} & \begin{bmatrix} \text{INHERITED}|\text{SLASH} & \{\boxed{1}\} \\ \text{TO-BIND}|\text{SLASH} & \{\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

- phonologically null, but structure-shares local and slash values

Traces

- Because the *local* value of a trace is structure-shared with the *slash* value, constraints on the trace will be constraints on the filler.
  - For example, *hates* specifies that its object be accusative, and this case information is local
  - So, the trace has [synsem|local|cat|head|case acc] as part of its entry, and thus the filler will also have to be accusative

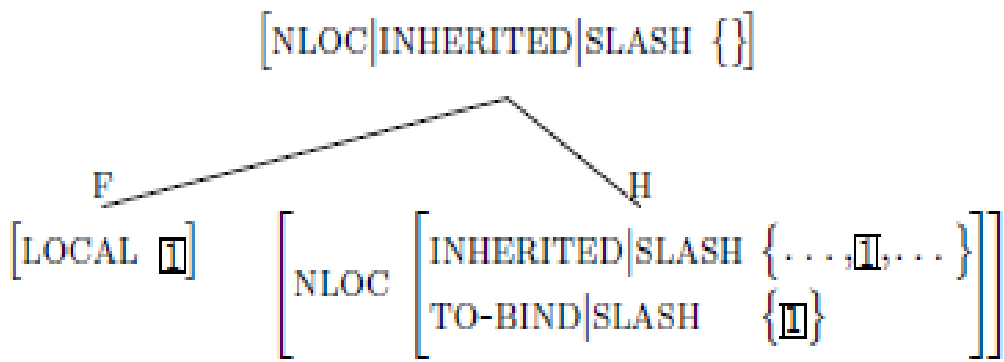
    $*He_i$/Him_i, John likes_ $_i$

The middle of a UDC: The Nonlocal Feature Principle (NFP)

- For each NON-LOCAL feature, the *inherited* value on the mother is the union of the *inherited* values on the daughter minus the *to-bind* value on the head daughter.
- In other words, the slash information (which is part of inherited) percolates “up” the tree
- This allows the all the local information of a trace to “move up” to the filler

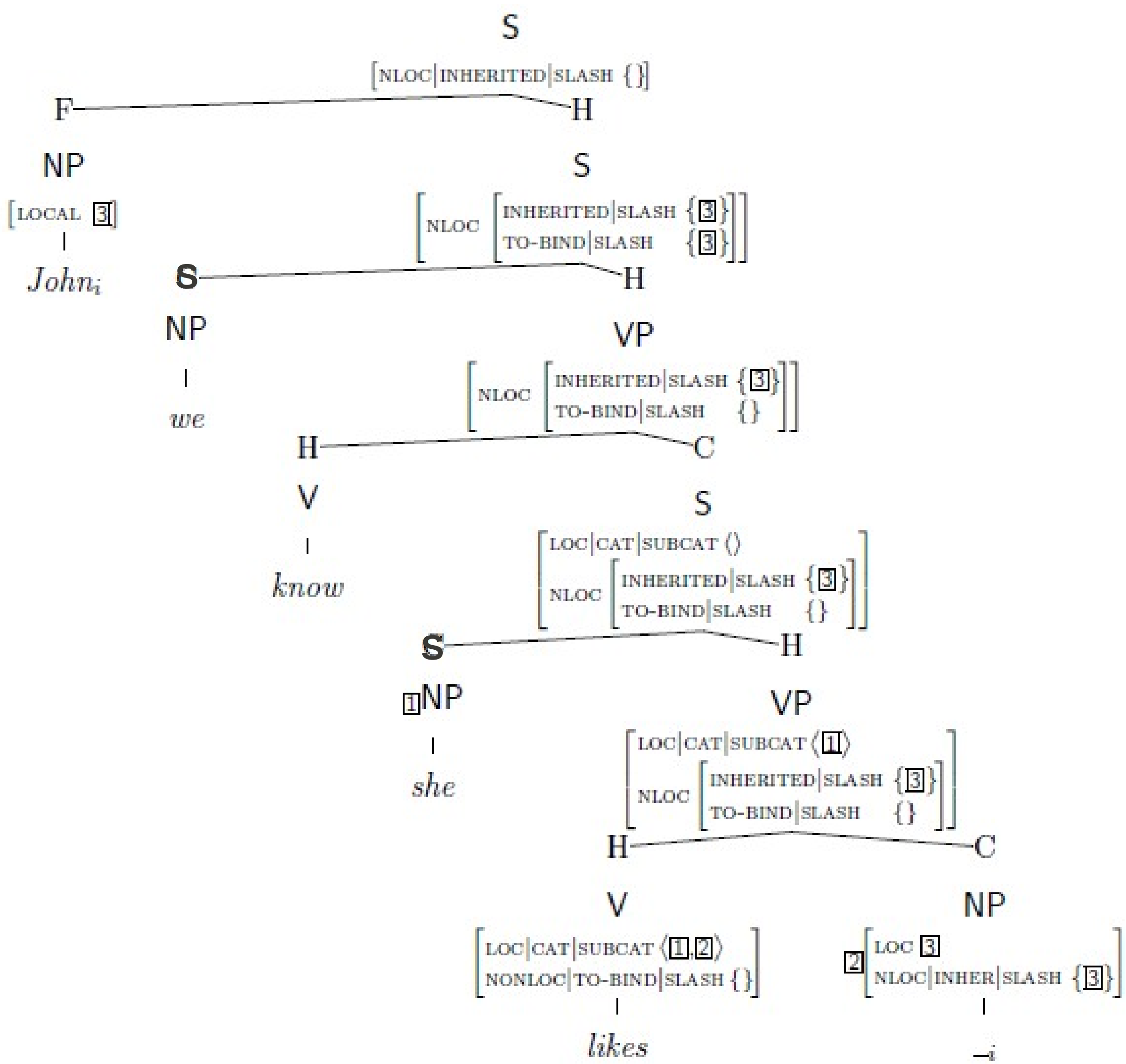
- The top of a UDC: *filler-head* structures

  Example for a structure licensed by the *filler-head* schema



- The analysis of the UDC example

John$_i$ we know She likes_$_i$

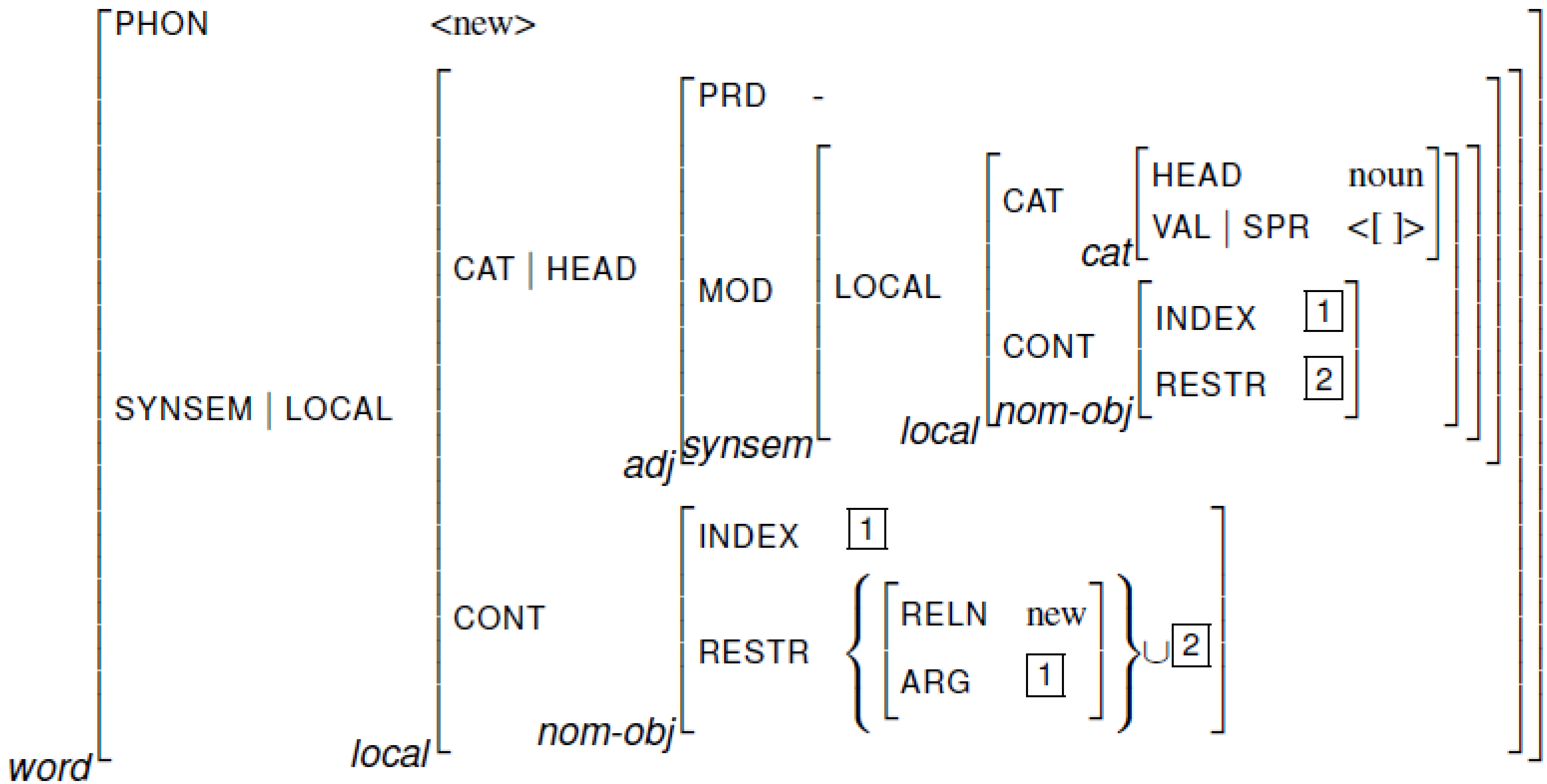


Example 4

John ***reads*** *a new book*

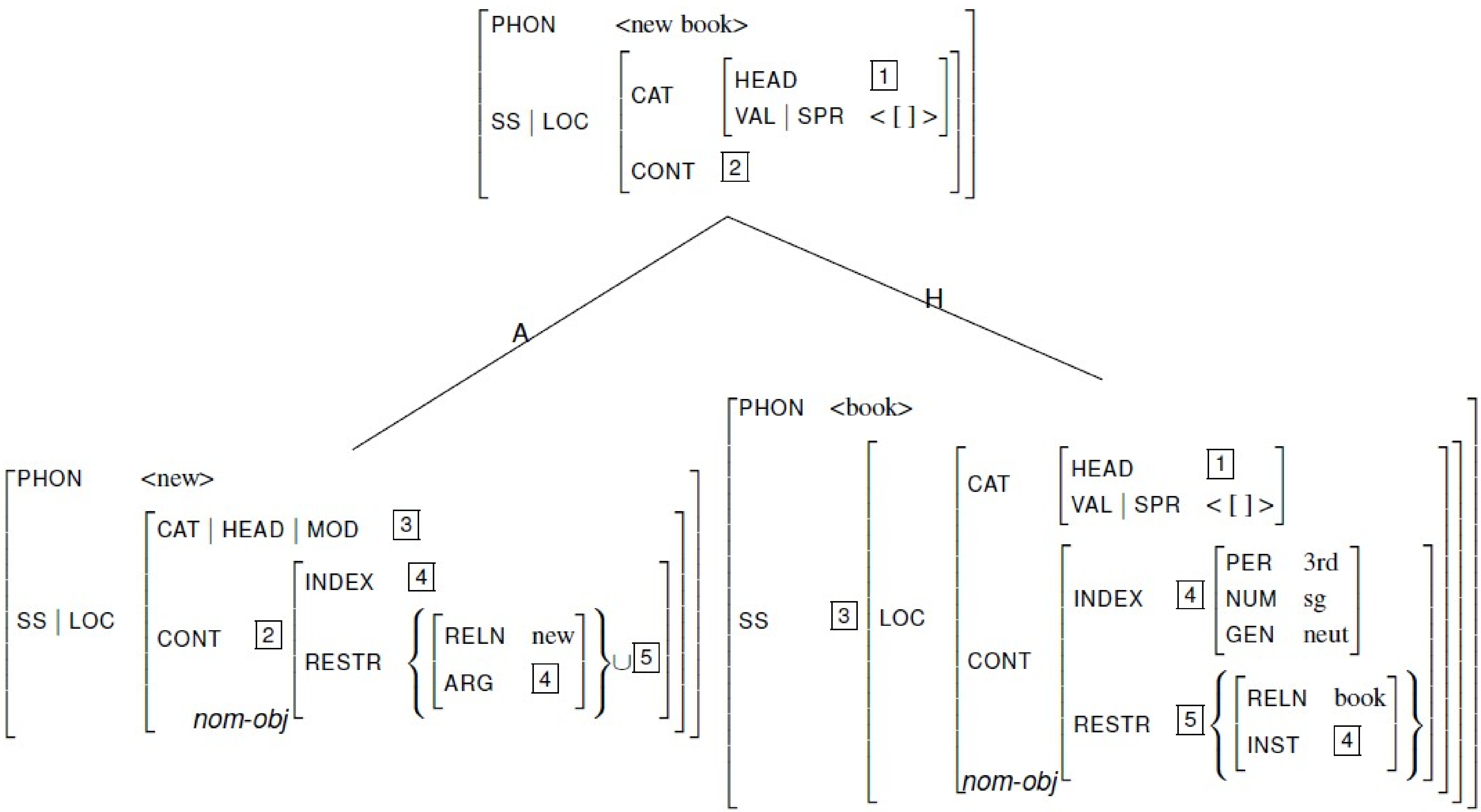
$$
{}_{word}\begin{bmatrix} \text{PHON} & \langle \text{reads} \rangle \\ \text{SYNSEM} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & {}_{verb}\begin{bmatrix} \text{VFORM} & \text{fin} \\ \text{AUX} & \text{bool} \\ \text{INV} & \text{bool} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \text{NP}_{\boxed{1}[3rd,sg]}[\text{nom,-PRD}] \rangle \\ \text{COMPS} & \langle \text{NP}_{\boxed{2}}[\text{acc,-PRD}] \rangle \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} & {}_{read}\begin{bmatrix} \text{READER} & \boxed{1} \\ \text{READEE} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

John reads a ***new*** *book*

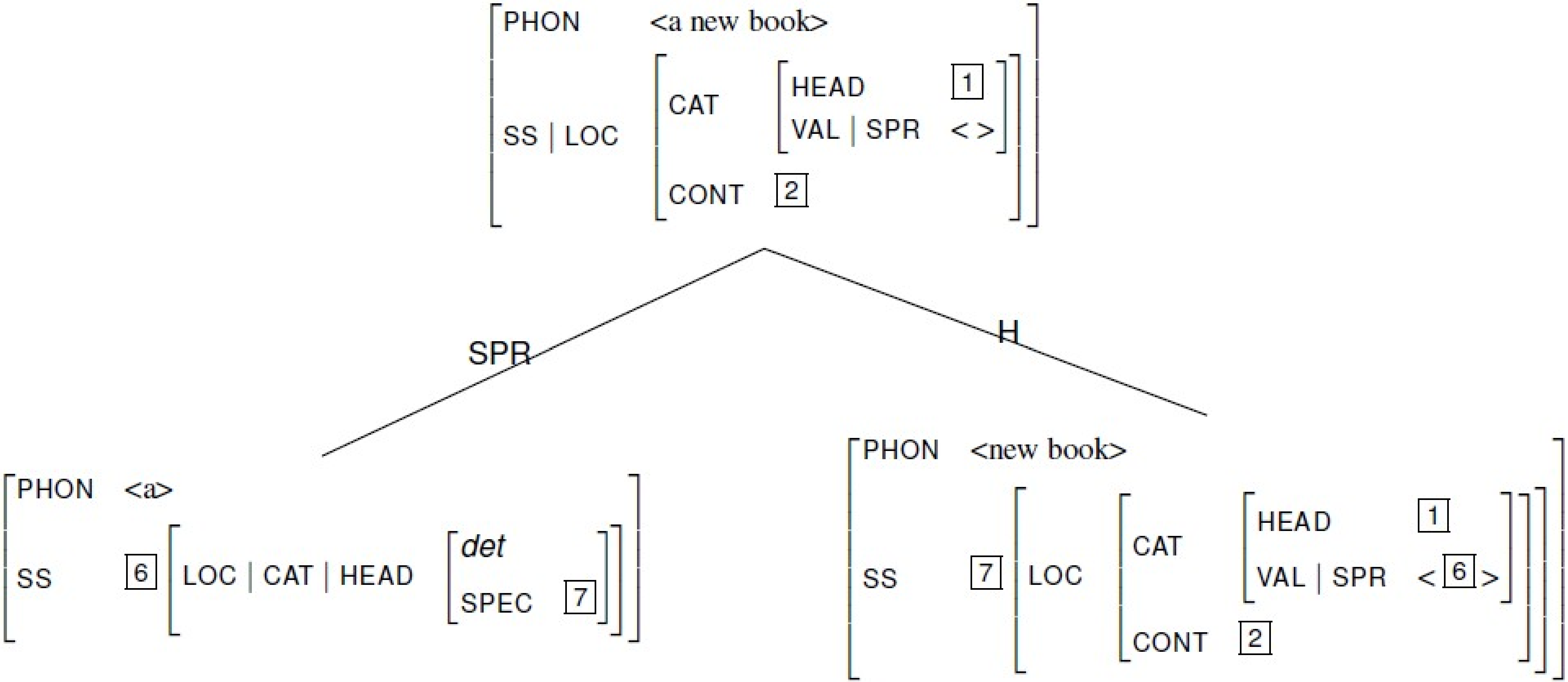


John read a new book

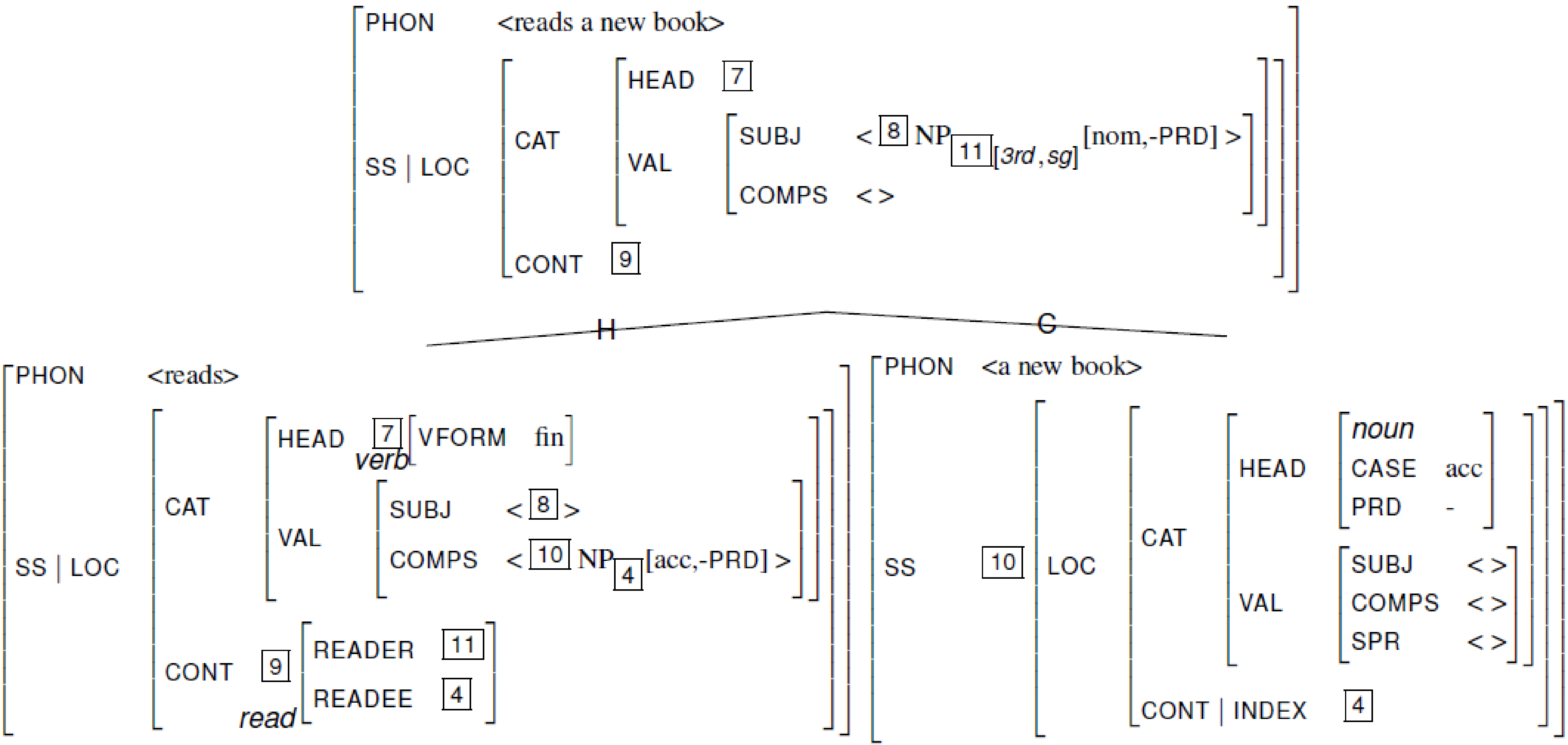
- Note: apply head-adjunct schema



John reads a new book



John reads a new book



John reads a new book -completed analysis