



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



HPSG as a Formal Linguistic Theory II

The grammar of HPSG'94

SS 2002

Fri, May 3rd 2002



Review

- general properties
- linguistic assumptions
- ID Schemata
 - SCHEMA 1 (HEAD-SUBJECT SCHEMA)
 - SCHEMA 2 (HEAD-COMPLEMENT SCHEMA)
 - SCHEMA 3 (HEAD-SUBJECT-COMPLEMENT SCHEMA)
 - SCHEMA 4 (HEAD-MARKER SCHEMA)
 - SCHEMA 5 (HEAD-ADJUNCT SCHEMA)
 - SCHEMA 6 (HEAD-FILLER SCHEMA)
- principles
 - The ID PRINCIPLE
 - HEAD FEATURE PRINCIPLE
 - SUBCATEGORIZATION PRINCIPLE
 - MARKING PRINCIPLE



1. SCHEMA 1 (HEAD-SUBJECT SCHEMA)

(a) “HEAD-DAUGHTER is a phrase”

(b) “SYNSEM | NONLOCAL | TO-BIND | SLASH value is {}”

2. SCHEMA 6 (HEAD-FILLER SCHEMA)



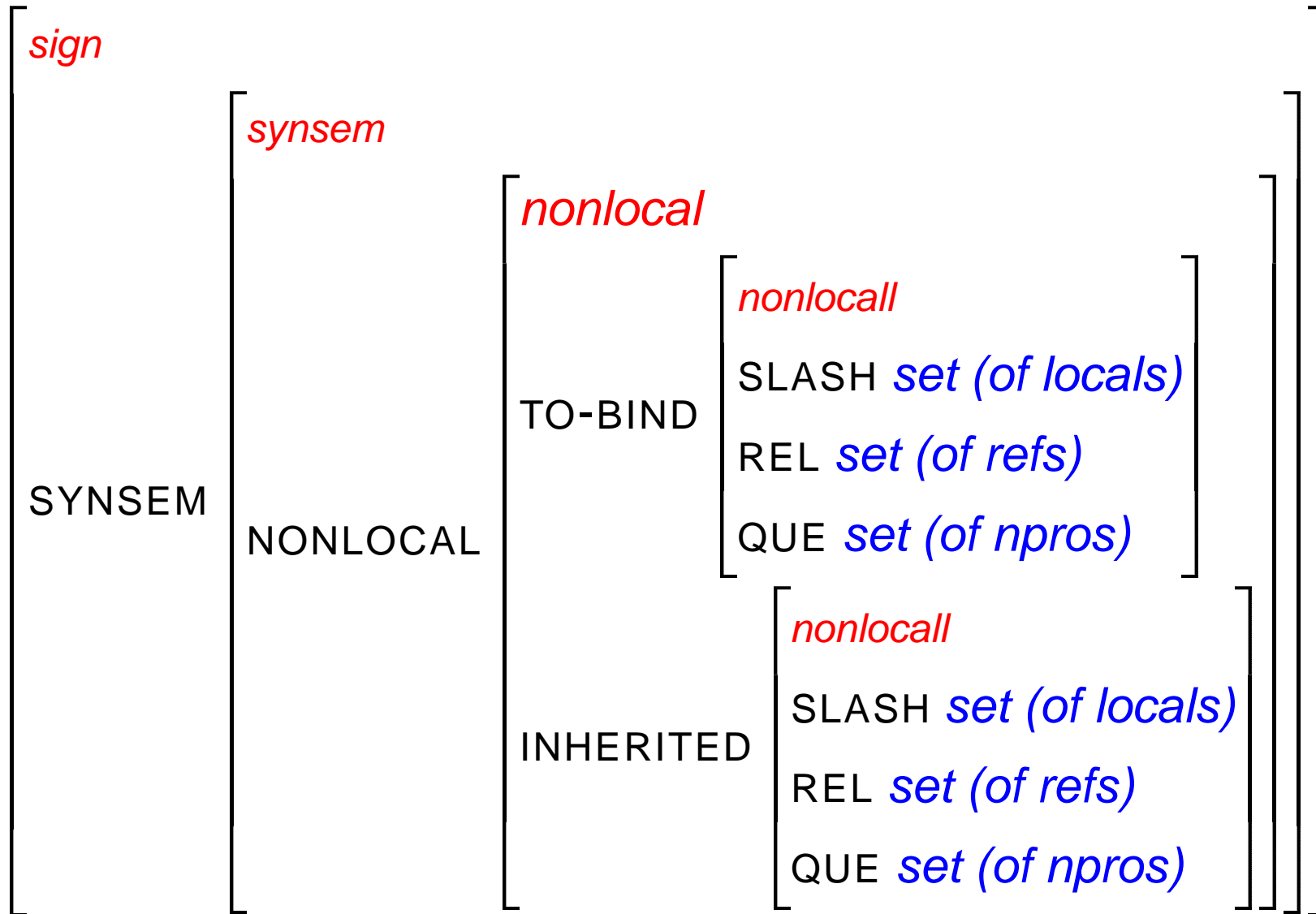
Explanation 1a

X-bar schema:

$$X^n \rightarrow \dots X^{n-1} \dots$$

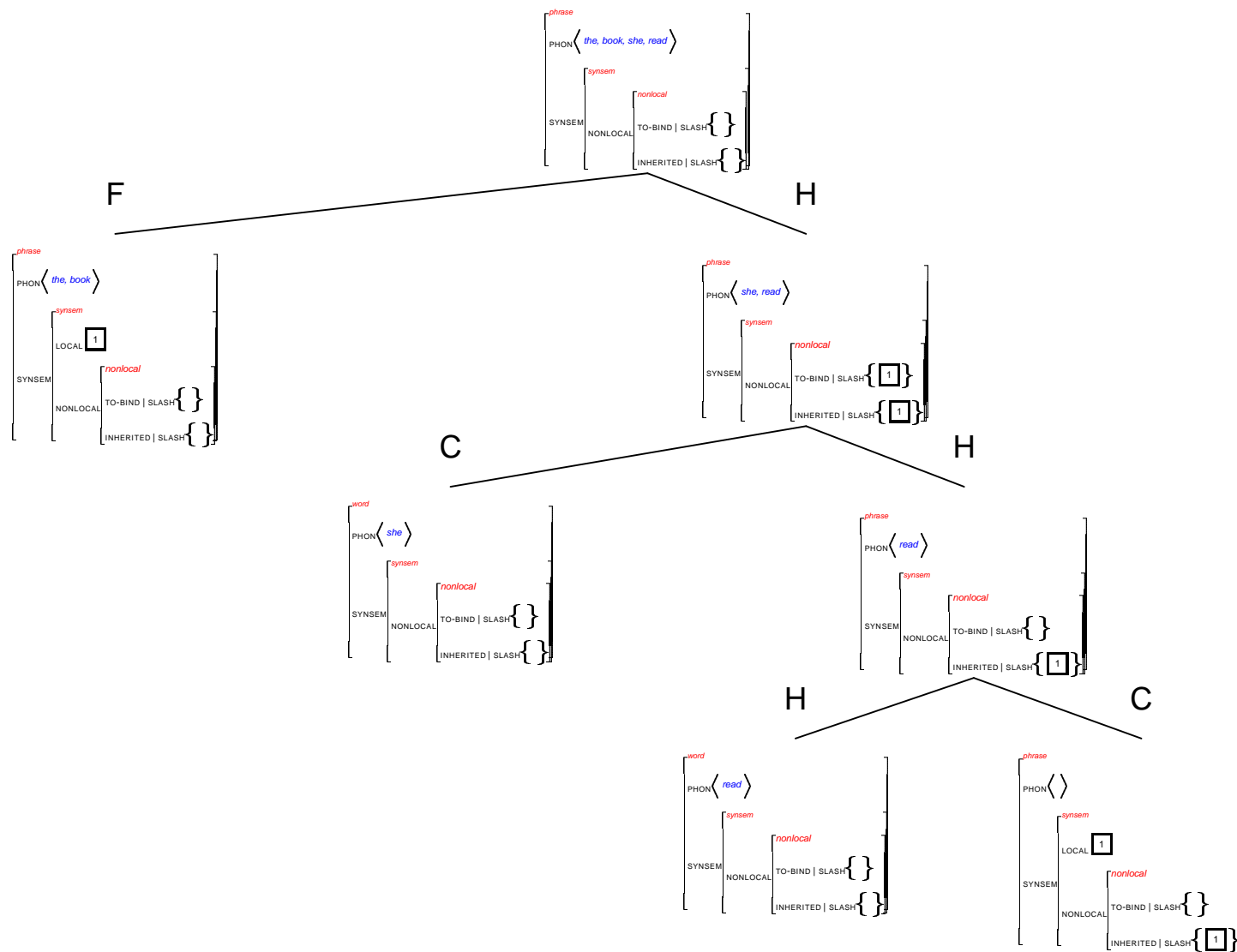


Explanation 1b





Explanation 1b





NONLOCAL FEATURE PRINCIPLE

In a headed phrase, for each nonlocal feature $F = \text{SLASH, QUE or REL}$, the value of $\text{SYNSEM} \mid \text{NONLOCAL} \mid \text{INHERITED} \mid F$ is the set difference of the union of the values on all the daughters, and the value of $\text{SYNSEM} \mid \text{NONLOCAL} \mid \text{TO-BIND} \mid F$ on the HEAD-DAUGHTER.

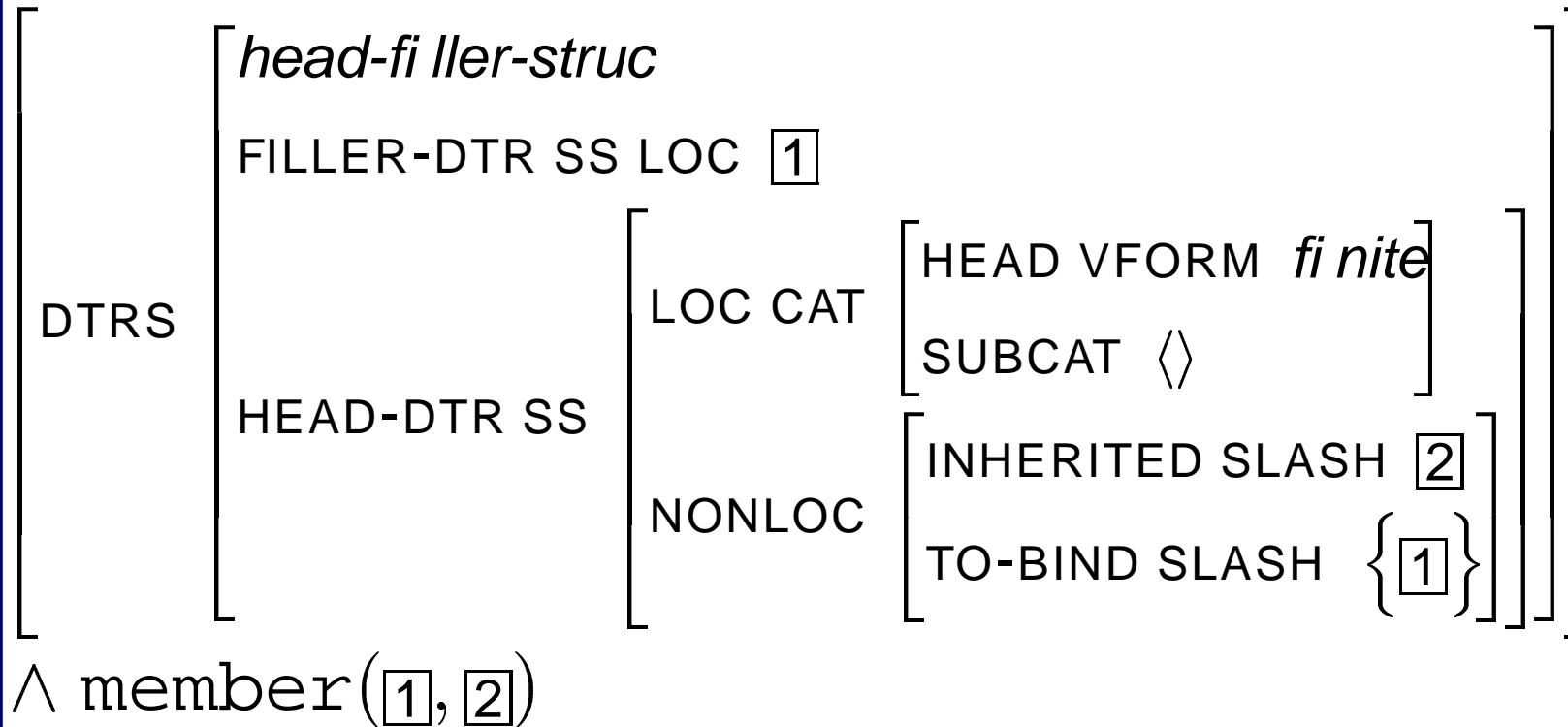


Explanation 2: SCHEMA 6 (HEAD-FILLER SCHEMA)

The DAUGHTERS value is an object of sort *head-filler-struct* 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.



Explanation 2: SCHEMA 6 (HEAD-FILLER SCHEMA) formalized





The Relation member

The relation $\text{member}(x,y)$ holds iff x is an element of the list y , that is

- either x is identical with y_{FIRST} or
- $\text{member}(x, y_{\text{REST}})$ holds.

Formalization:

$$\forall x \forall y (\text{member}(x,y) \leftrightarrow \exists \boxed{1} (y \left[\begin{array}{l} \text{FIRST } x \\ \text{REST } \boxed{1} \end{array} \right] \vee \text{member}(x, \boxed{1})))$$



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

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

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

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

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



The Relation *sign-to-synsem*

The Relation *sign-to-synsem* holds between a list x of *sign* objects and a list y of *synsem* objects iff

- x is an empty list and y is an empty list or
- the *SYNSEM* value of the first element of x is the first element of y and the relation *sign-to-synsem* holds between the lists x and y each shortened with respect to the first element.

Formalization:

$\forall x \forall y$

$(\text{sign-to-synsem}(x,y) \leftrightarrow$

$((x \text{ [elist]} \wedge y \text{ [elist]}) \vee$

$\exists \boxed{1} \exists \boxed{2} \exists \boxed{3}$

$((x \text{ [FIRST SYNSEM } \boxed{3}] \wedge y \text{ [FIRST } \boxed{3}] \wedge \text{sign-to-synsem}(\boxed{1}, \boxed{2})) \vee$
 $(x \text{ [REST } \boxed{1}] \wedge y \text{ [REST } \boxed{2}] \wedge \text{sign-to-synsem}(\boxed{1}, \boxed{2}))$))



The Relation `append`

The relation `append(x,y,z)` holds iff the list z is the concatenation of the list x and the list y .

Formalization:

$\forall x \forall y \forall z (\text{append}(x,y,z) \leftrightarrow$

$((x[\mathit{elist}] \wedge y = z) \vee$

$\left(\begin{array}{c} \exists \boxed{1} \exists \boxed{2} \exists \boxed{3} \\ x \left[\begin{array}{l} \text{FIRST } \boxed{1} \\ \text{REST } \boxed{2} \end{array} \right] \wedge z \left[\begin{array}{l} \text{FIRST } \boxed{1} \\ \text{REST } \boxed{3} \end{array} \right] \wedge \text{append}(\boxed{2},y,\boxed{3}) \end{array} \right) \vee \end{array} \right) \right)$



SCHEMA 4 (HEAD-MARKER SCHEMA)

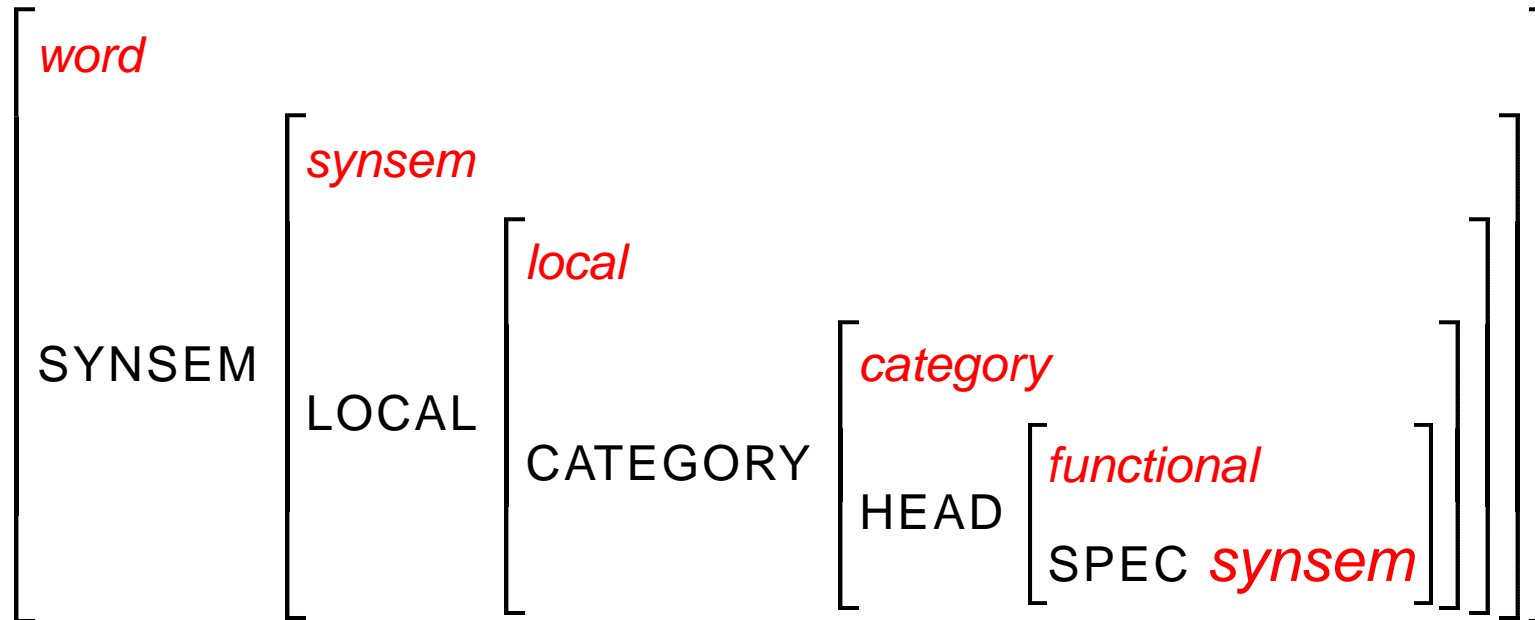
The DAUGHTERS value is an object of sort *head-marker-struct* whose HEAD-DAUGHTER | SYNSEM | NONLOCAL | TO-BIND | SLASH value is {}, and whose MARKER-DAUGHTER | SYNSEM | LOCAL | CATEGORY | HEAD value is of sort *marker*.

Formalization:

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



AVM Description of a Functional Part of Speech





In a headed phrase whose nonhead daughter (either the MARKER-DAUGHTER or COMPLEMENT-DAUGHTERS | FIRST) has a SYNSEM | LOCAL | CATEGORY | HEAD value of sort *functional*, the SPEC value of that value must be token-identical with the phrase's DAUGHTERS | HEAD-DAUGHTER | SYNSEM value.



SPEC PRINCIPLE formalized

$\forall \boxed{1} \forall \boxed{2}$

$$\left(\left(\left[\text{DTRS} \left[\left[\text{MARKER-DTR } \boxed{1} \right] \vee \left[\text{COMP-DTRS } \langle \boxed{1} \mid \textit{list} \rangle \right] \right] \right) \wedge \boxed{1} \left[\text{SS LOC CAT HEAD} \begin{array}{l} \textit{functional} \\ \text{SPEC } \boxed{2} \end{array} \right] \right) \rightarrow \left[\text{DTRS HEAD-DTR SS } \boxed{2} \right] \right)$$



The `SYNSEM` value of any trace must be a (noninitial) member of the `SUBCAT` list of a substantive word.

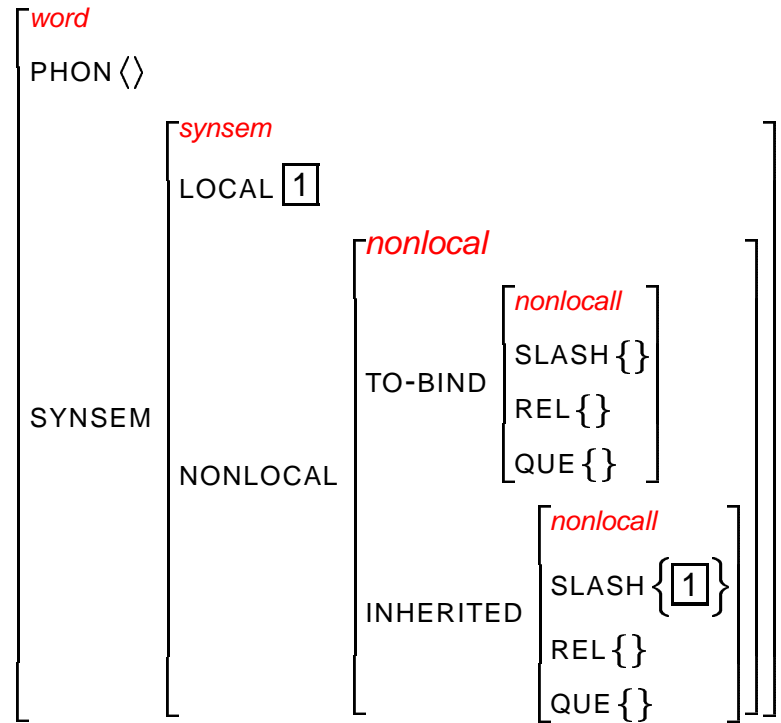


TRACE PRINCIPLE formalized

$$\forall x \forall_{[2]} \left(\begin{array}{l} \exists_{[1]} \\ \left(\begin{array}{l} \text{PHON } \textit{nelist} \\ \text{SS} \left[\begin{array}{l} \text{LOC CAT} \\ \text{NONLOC INHERITED} \end{array} \right] \left[\begin{array}{l} \text{HEAD VFORM } \textit{fi nite} \\ \text{SUBCAT } \langle \rangle \\ \text{QUE } \{ \} \\ \text{REL } \{ \} \\ \text{SLASH } \{ \} \end{array} \right] \\ \wedge \\ \text{sign} \\ \text{PHON } \langle \rangle \\ \text{X} \left[\begin{array}{l} \text{LOC } [1] \\ \text{NONLOC } [2] \end{array} \right] \left[\begin{array}{l} \text{INHERITED} \\ \text{TO-BIND} \end{array} \right] \left[\begin{array}{l} \text{QUE } \{ \} \\ \text{REL } \{ \} \\ \text{SLASH } [1] \\ \text{QUE } \{ \} \\ \text{REL } \{ \} \\ \text{SLASH } \{ \} \end{array} \right] \end{array} \right) \rightarrow \\ \exists y \exists_{[3]} \left(\begin{array}{l} \text{word} \\ \text{Y} \left[\begin{array}{l} \text{SS LOC CAT} \left[\begin{array}{l} \text{HEAD } \textit{substantive} \\ \text{SUBCAT } \langle \textit{object} [3] \rangle \end{array} \right] \end{array} \right] \wedge \text{member}([2], [3]) \end{array} \right) \end{array} \right)$$



AVM Description of Traces

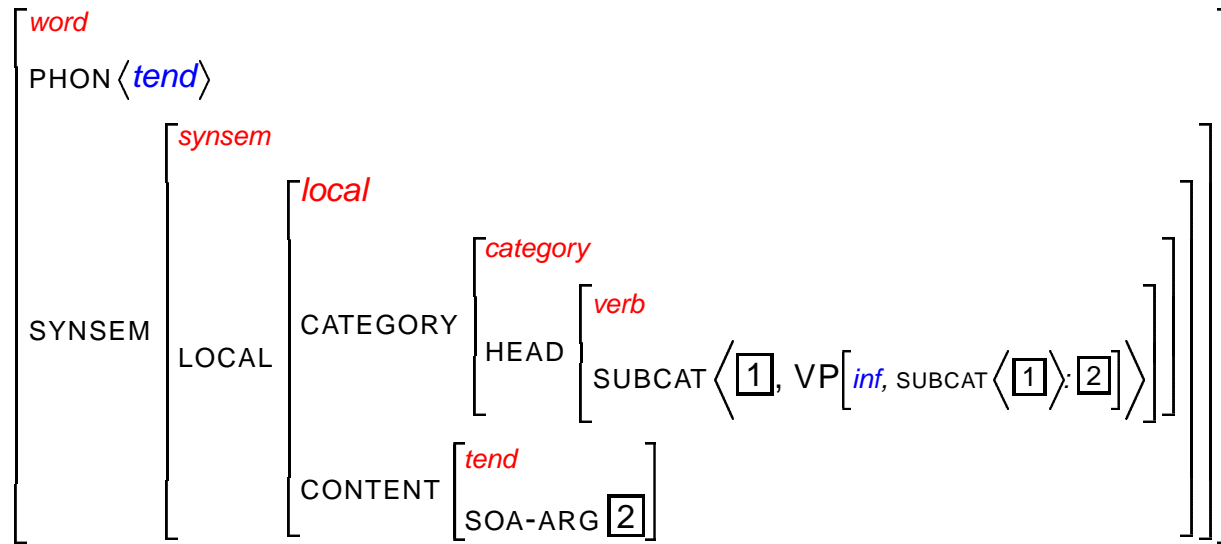




$$\left[\textit{word} \right] \rightarrow (LE_1 \vee \dots \vee LE_n)$$

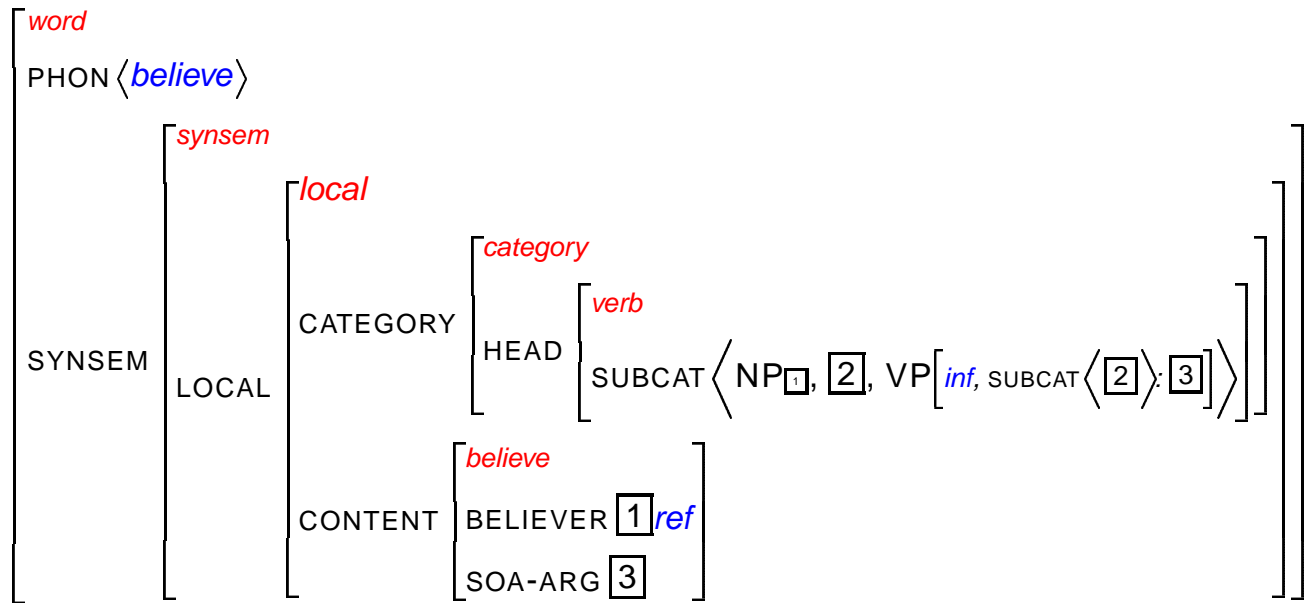


PARTIAL LEXICAL ENTRY OF A SUBJECT RAISING VERB



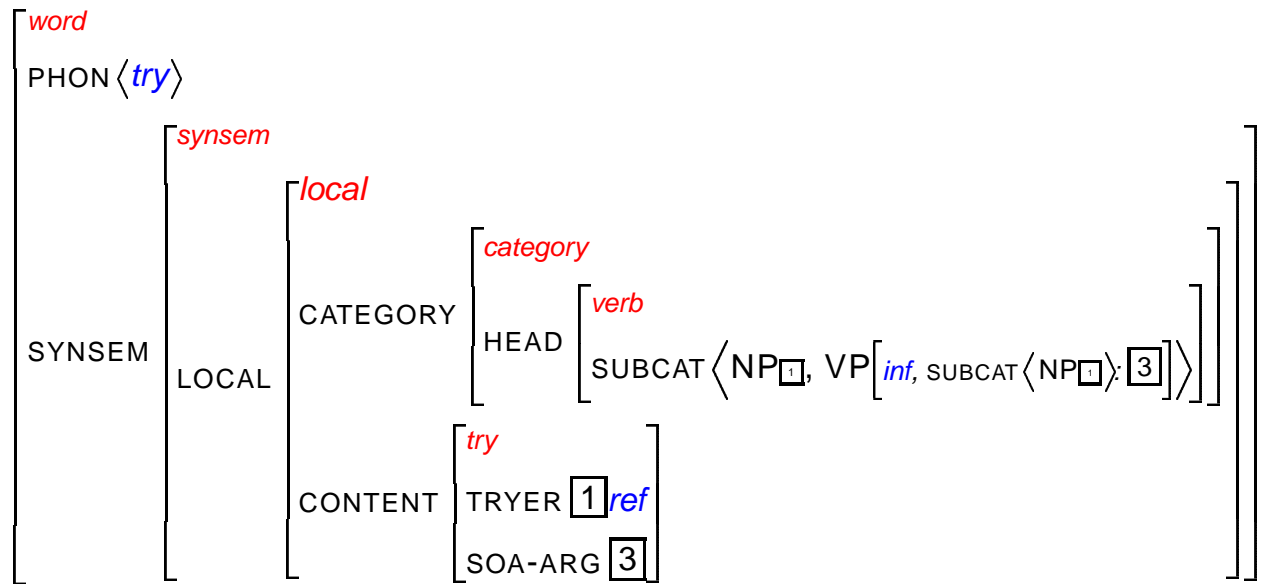


PARTIAL LEXICAL ENTRY OF AN OBJECT RAISING VERB



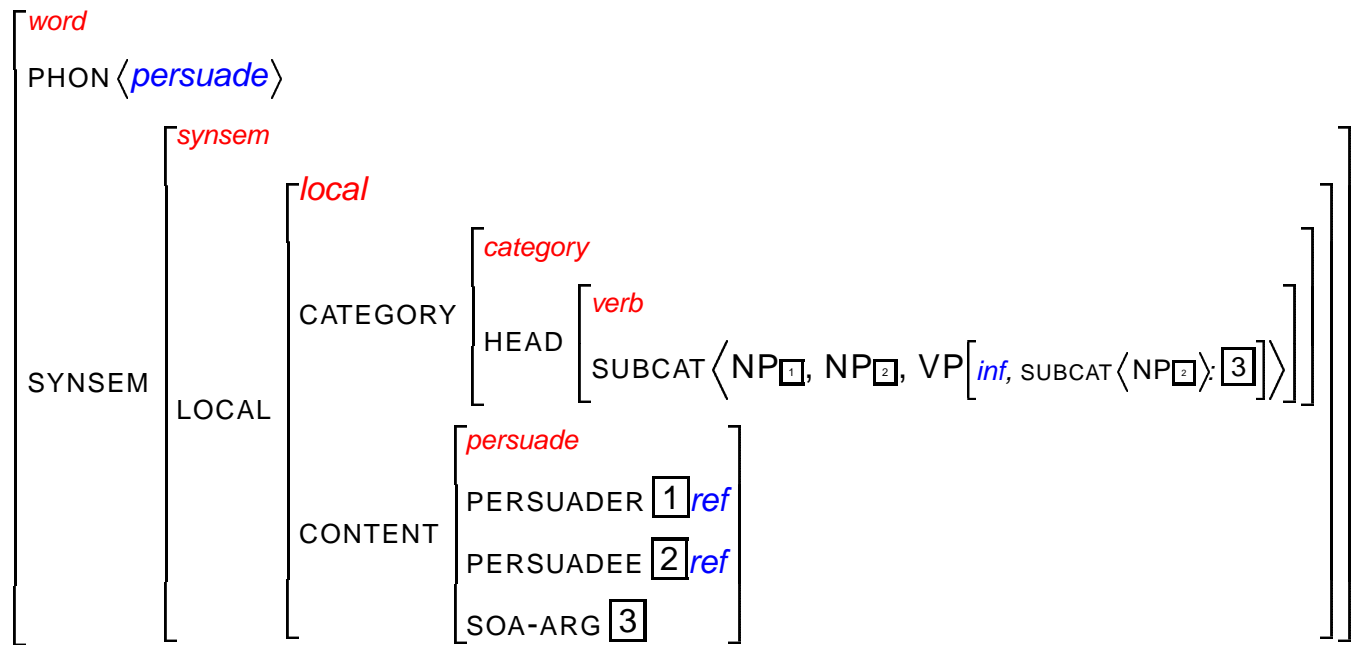


PARTIAL LEXICAL ENTRY OF A SUBJECT CONTROL VERB





PARTIAL LEXICAL ENTRY OF AN OBJECT CONTROL VERB





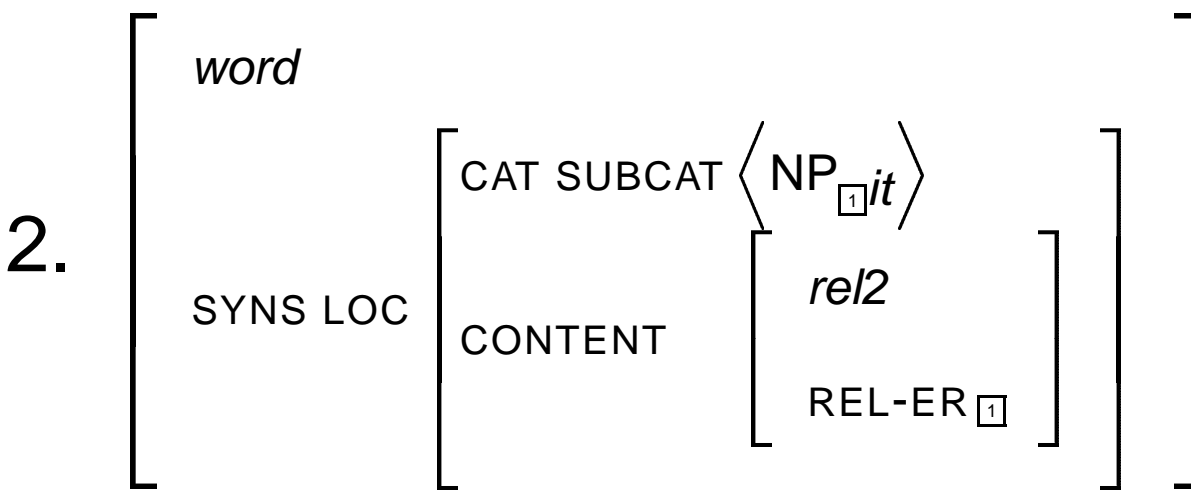
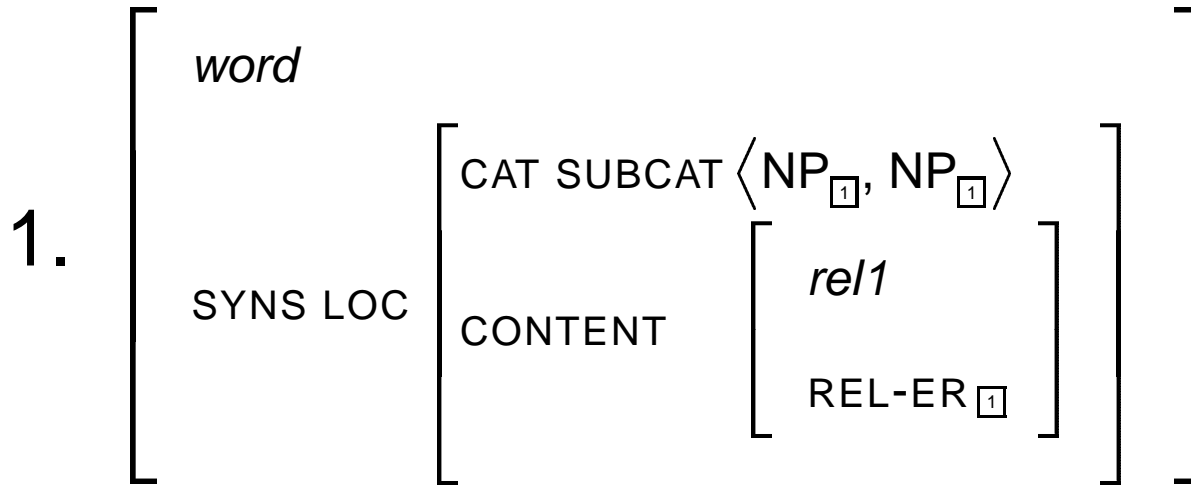
Let E be a lexical entry in which the (description of the) SUBCAT list L contains (a description corresponding to) a member x (of L) that is not explicitly described in E as an expletive. Then in (the description of) the CONTENT value, x is (described as) assigned no semantic role if and only if L (is described as if it) contains a non-subject whose own SUBCAT value is $\langle x \rangle$.

Logical structure of RAISING PRINCIPLE:

$$A \rightarrow (B \leftrightarrow C)$$

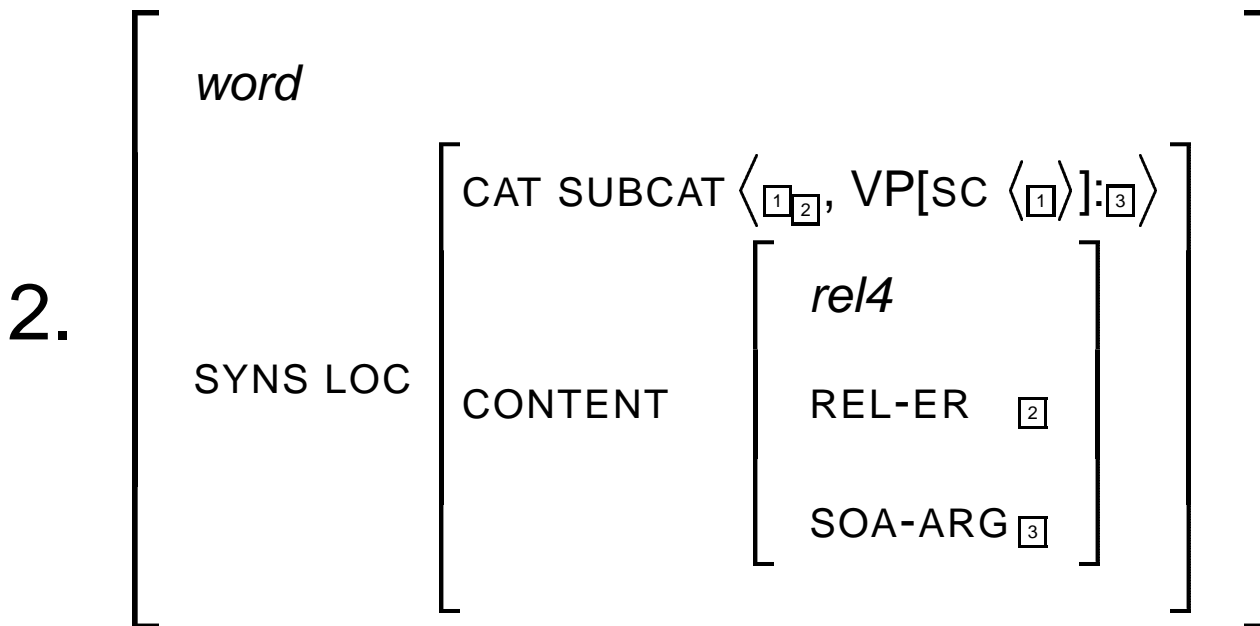
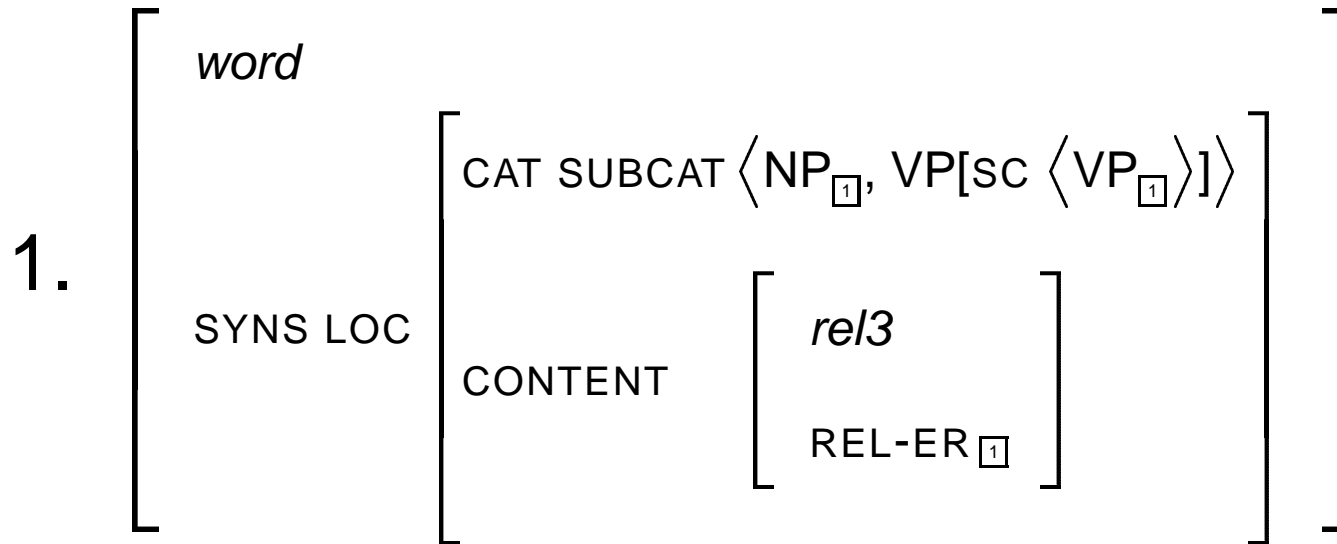


EXAMPLES I



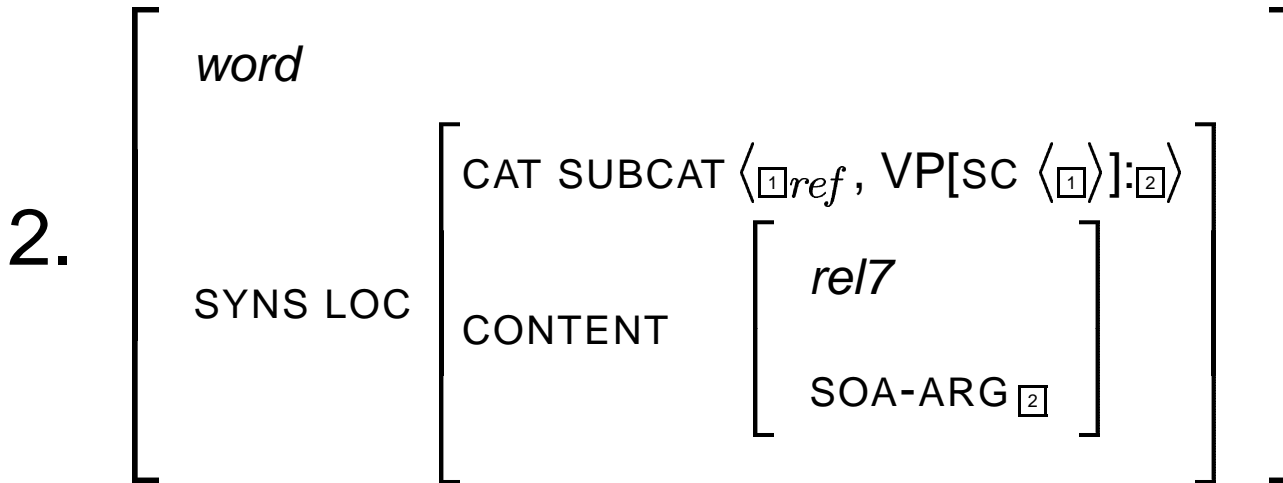
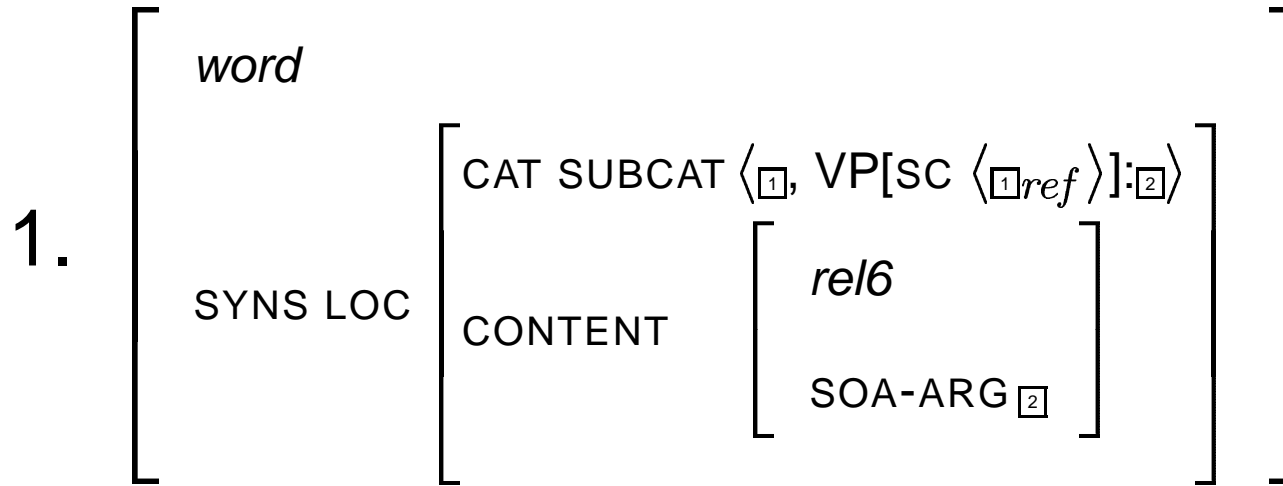


EXAMPLES II





EXAMPLES III





1. Principle A.
A locally o-commanded anaphor must be locally o-bound.
2. Principle B.
A personal pronoun must be locally o-free.
3. Principle C.
A nonpronoun must be o-free.



Principle A:

$$\forall x \left(\exists y \text{ loc-o-command}(y, x \left[\text{LOC CONT } ana \right]) \rightarrow \exists z \text{ loc-o-bind}(z, x) \right)$$

Principle B:

$$\forall x \left(x \left[\text{LOC CONT } ppro \right] \rightarrow \neg \exists y \text{ loc-o-bind}(y, x) \right)$$

Principle C:

$$\forall x \left(x \left[\text{LOC CONT } npro \right] \rightarrow \neg \exists y \text{ o-bind}(y, x) \right)$$



A referential *synsem* object *locally o-commands* another *synsem* object provided they have distinct `LOCAL` values and either

- (1) the second is more oblique than the first, or
- (2) the second is a member of the `SUBCAT` list of a *synsem* object that is more oblique than the first.



THE RELATION LOC-O-COMMAND FORMALIZED

$$\forall x \forall y \left(\text{loc-o-command}(x, y) \leftrightarrow \begin{array}{l} \exists s \exists \boxed{1} \exists \boxed{2} \exists \boxed{3} \\ \left(\left(x \left[\text{LOC} \boxed{1} \left[\text{CONT INDEX } \textit{ref} \right] \right] \wedge y \left[\text{LOC} \boxed{2} \right] \wedge \neg \boxed{1} = \boxed{2} \right) \wedge \right. \\ \left. \left(\text{more-oblique}(y, x) \vee \right. \right. \\ \left. \left(\text{more-oblique} \left(\begin{array}{l} s \\ \text{synsem} \\ \text{LOC CAT SUBCAT } \boxed{3} \end{array} \right), x \right) \right. \\ \left. \left. \wedge \text{member}(y, \boxed{3}) \right) \right) \end{array} \right)$$



One *synsem* object is *more oblique* than another provided it appears to the right of the other on the SUBCAT list of some word.

Formalization:

$$\forall x \forall y \left(\begin{array}{l} \text{more-oblique}(x, y) \leftrightarrow \\ \exists w \exists \boxed{1} \left(\begin{array}{l} \text{word} \\ \text{SS LOC CAT SUBCAT } \boxed{1} \end{array} \wedge \text{to-the-right}(x, y, \boxed{1}) \right) \end{array} \right)$$



THE RELATION TO-THE-RIGHT

The relation `to-the-right` holds for three objects x , y und z within a configuration of objects if y stands before x on the list z .

Formalization:

$\forall x \forall y \forall z$

$$\left(\text{to-the-right}(x, y, z) \leftrightarrow \left(\begin{array}{l} \exists \boxed{1} \left(z \begin{array}{l} \text{FIRST } y \\ \text{REST } \boxed{1} \end{array} \wedge \text{member}(x, \boxed{1}) \right) \\ \vee \exists \boxed{1} \left(z \begin{array}{l} \text{REST } \boxed{1} \end{array} \wedge \text{to-the-right}(x, y, \boxed{1}) \right) \end{array} \right) \right)$$



A referential *synsem* object *o-commands* another *synsem* object provided they have distinct LOCAL values and either (1) the second is more oblique than the first, (2) the second is a member of the SUBCAT list of a *synsem* object that is o-commanded by the first, or (3) the second has the same LOCAL | CATEGORY | HEAD value as a *synsem* object that is o-commanded by the first.



THE RELATION o-COMMAND formalized

$\forall x \forall y$

$\text{o-command}(x, y) \leftrightarrow$

$\exists s_1 \exists s_2 \exists_{\boxed{1}} \exists_{\boxed{2}} \exists_{\boxed{3}} \exists_{\boxed{4}}$

$\left(x_{\boxed{1}[\text{LOC } \boxed{1}[\text{CONT INDEX } \textit{ref}]]} \wedge y_{\boxed{2}[\text{LOC } \boxed{2}[\text{CAT HEAD } \boxed{4}]]} \wedge \neg_{\boxed{1}} =_{\boxed{2}} \right) \wedge$

$\left(\text{more-oblique}(y, x) \vee \right.$

$\left(\left(\text{o-command}(x, s_1_{\boxed{3}[\textit{synsem} \text{ LOC CAT SUBCAT } \boxed{3}]}]} \right) \wedge \text{member}(y, \boxed{3}) \right) \vee$

$\left(\text{o-command}(x, s_2) \wedge s_2_{\boxed{4}[\textit{synsem} \text{ LOC CAT HEAD } \boxed{4}]}]} \right) \right)$



THE RELATIONS LOC-O-BIND AND O-BIND

One referential *synsem* object (*locally*) *o-binds* another provided it (locally) *o-commands* and is coindexed with the other. A referential *synsem* object is (*locally*) *o-free* provided it is not (locally) *o-bound*. Two *synsem* objects are *coindexed* provided their LOCAL | CONTENT | INDEX values are token-identical.

$$\forall x \forall y \left(\begin{array}{l} \text{loc-o-bind}(x, y) \leftrightarrow \\ \exists_{[1]} \text{loc-o-command}(x_{[\text{LOC CONT INDEX } [1]}], y_{[\text{LOC CONT INDEX } [1]]}) \end{array} \right)$$

$$\forall x \forall y \left(\begin{array}{l} \text{o-bind}(x, y) \leftrightarrow \\ \exists_{[1]} \text{o-command}(x_{[\text{LOC CONT INDEX } [1]}], y_{[\text{LOC CONT INDEX } [1]]}) \end{array} \right)$$



Modified o-command Relation

$\forall x \forall y$

$\text{o-command}(x, y) \leftrightarrow$

$\exists s_1 \exists s_2 \exists z \exists_{\boxed{1}} \exists_{\boxed{2}} \exists_{\boxed{3}}$

$\left(x_{\text{LOC } \boxed{1} [\text{CONT INDEX } ref]} \wedge y_{\text{LOC } \boxed{2} [\text{CAT HEAD } \boxed{4}]} \wedge \neg_{\boxed{1}} = \boxed{2} \right) \wedge$

$\text{more-oblique}(y, x) \vee$

$\left(\text{o-command} \left(x, s_1 \left[\begin{array}{l} \text{synsem} \\ \text{LOC CAT SUBCAT } \boxed{3} \end{array} \right] \right) \wedge \text{member}(y, \boxed{3}) \right) \vee$

$\left(\text{o-command}(x, s_2) \wedge z_{\text{SYNSEM DTRS H-DTR SYNSEM } y} \right)$



SEMANTICS PRINCIPLE (clause (b))

If the semantic head's $\text{SYNSEM} \mid \text{LOCAL} \mid \text{CONTENT}$ value is of sort $psoa$, then the $\text{SYNSEM} \mid \text{LOCAL} \mid \text{CONTENT} \mid \text{NUCLEUS}$ value is token-identical with that of the semantic head, and the $\text{SYNSEM} \mid \text{LOCAL} \mid \text{CONTENT} \mid \text{QUANTS}$ value is the concatenation of the RETRIEVED value and the semantic head's $\text{SYNSEM} \mid \text{LOCAL} \mid \text{CONTENT} \mid \text{QUANTS}$ value; otherwise the RETRIEVED value is the empty list, and the $\text{SYNSEM} \mid \text{LOCAL} \mid \text{CONTENT}$ value is token-identical with that of the semantic head.



SEMANTICS PRINCIPLE formalized

