Ling 566
Nov 11, 2023

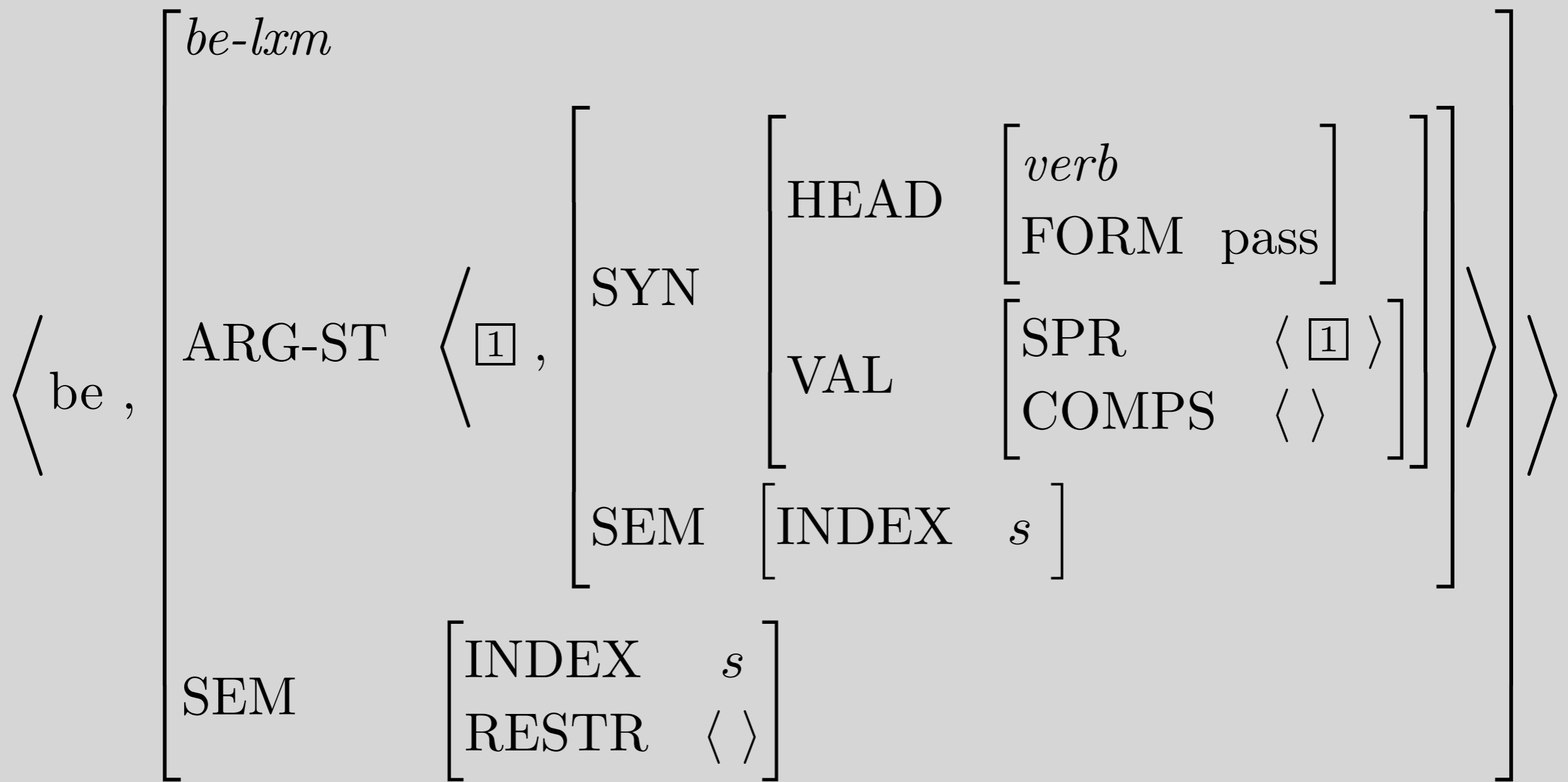## Non-referential NPs, Expletives, and Extraposition

Overview
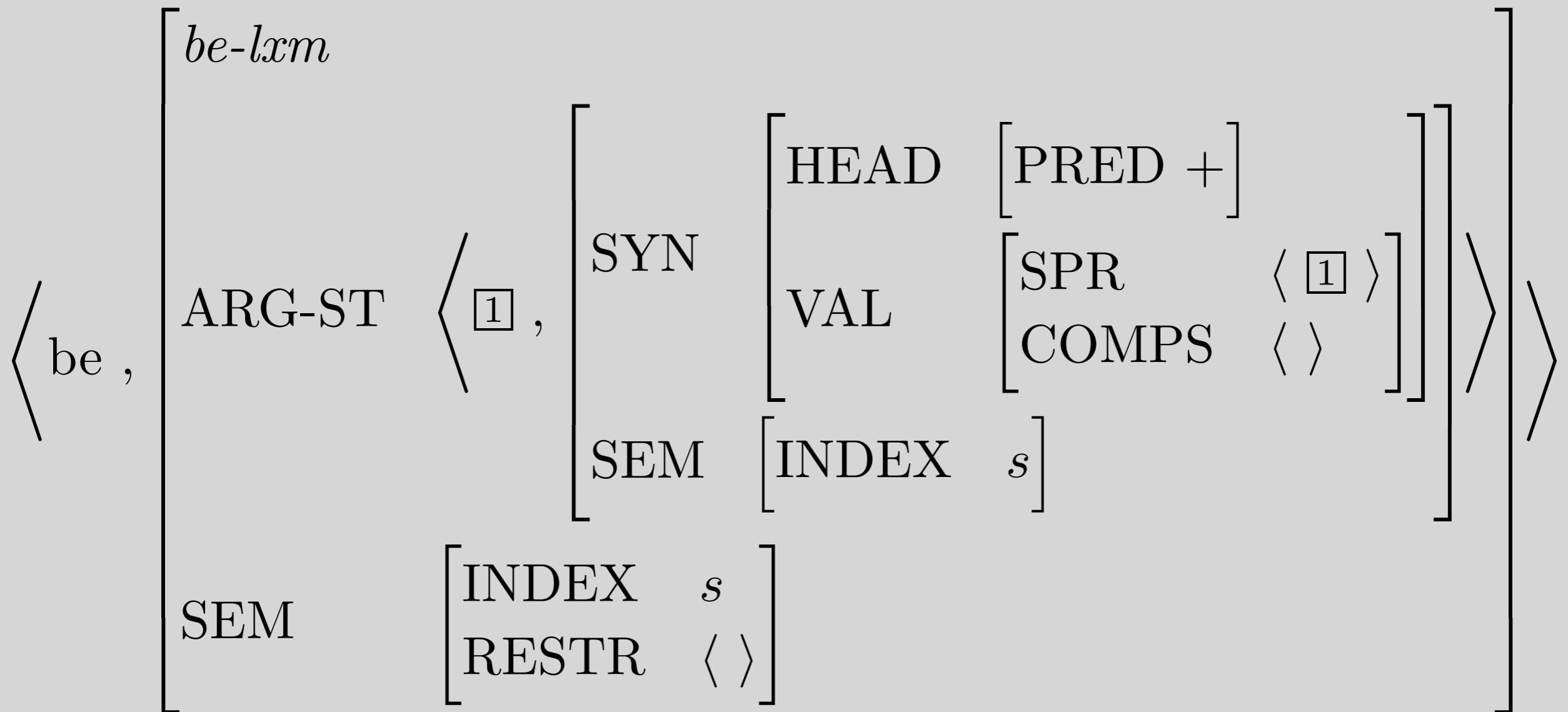
- Existentials

- Extraposition

- Idioms

Where We Are, and Where We're Going

- Last time, we met the passive *be*.

- Passive *be* is just a special case -- that *be* generally introduces [PRED **+**] constituents (next slide).

- Today, we'll start with another *be*, which occurs in existential sentences starting with *there*, e.g. *There is a monster in Loch Ness*.

- Then we'll look at this use of *there*.

- Which will lead us to a more general examination of NPs that don't refer, including some uses of *it* and certain idiomatic uses of NPs.
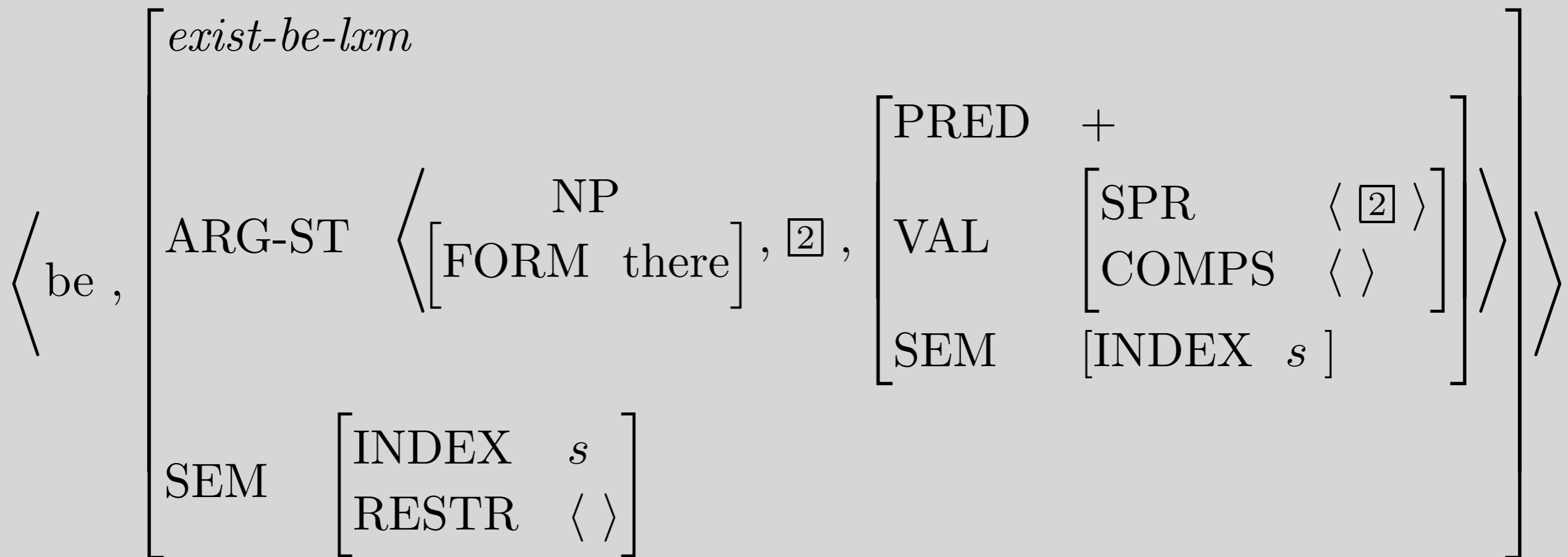
Chapter 10 entry for *be*

$$
\left\langle \text{be} \,,\;
\begin{bmatrix}
\textit{be-lxm} \\[4pt]
\text{ARG-ST} \quad
\left\langle
\boxed{1} \,,\;
\begin{bmatrix}
\text{SYN} \quad
\begin{bmatrix}
\text{HEAD} \quad
\begin{bmatrix}
\textit{verb} \\
\text{FORM} \quad \text{pass}
\end{bmatrix} \\[8pt]
\text{VAL} \quad
\begin{bmatrix}
\text{SPR} \quad \langle\, \boxed{1} \,\rangle \\
\text{COMPS} \quad \langle\,\rangle
\end{bmatrix}
\end{bmatrix} \\[18pt]
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad s
\end{bmatrix}
\end{bmatrix}
\right\rangle \\[24pt]
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad s \\
\text{RESTR} \quad \langle\,\rangle
\end{bmatrix}
\end{bmatrix}
\right\rangle
$$

Copula (generalized)

$$
\left\langle \text{be} , \begin{bmatrix} \textit{be-lxm} \\ \text{ARG-ST} \quad \left\langle \boxed{1} , \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \text{PRED} + \end{bmatrix} \\ \text{VAL} \begin{bmatrix} \text{SPR} \quad \langle \boxed{1} \rangle \\ \text{COMPS} \quad \langle \rangle \end{bmatrix} \end{bmatrix} \\ \text{SEM} \begin{bmatrix} \text{INDEX} \quad s \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{SEM} \begin{bmatrix} \text{INDEX} \quad s \\ \text{RESTR} \quad \langle \rangle \end{bmatrix} \end{bmatrix} \right\rangle
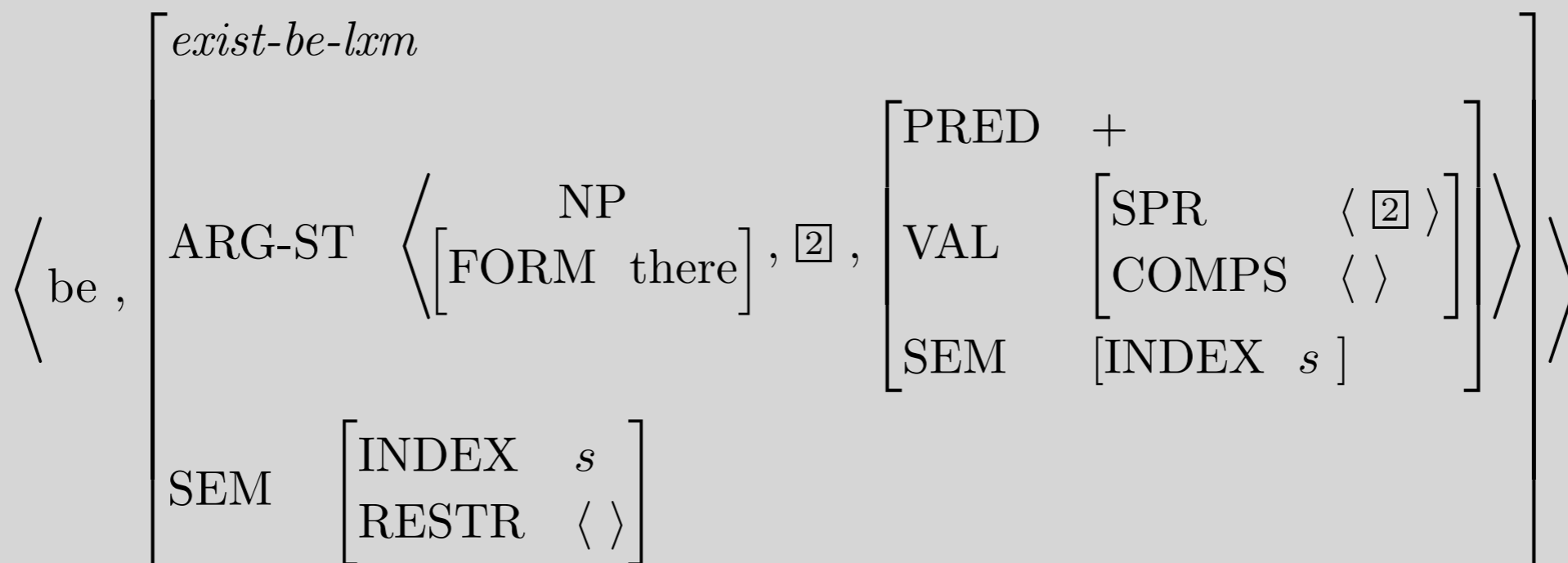$$

Existentials

- The *be* in *There is a page missing* cannot be the same *be* that occurs in sentences like *Pat is tall* or *A cat was chased by a dog*. Why not?

- So we need a separate lexical entry for this *be*, stipulating:

  - Its SPR must be *there*

  - It takes two complements, the first an NP and the second an AP, PP, or (certain kind of) VP.

  - The semantics should capture the relation between, e.g. *There is a page missing* and *A page is missing*.

Lexical Entry for the Existential *be*

$$
\left\langle \text{be} \ , \ \begin{bmatrix} \textit{exist-be-lxm} \\ \\ \text{ARG-ST} \ \left\langle \begin{bmatrix} \text{NP} \\ \text{FORM} \quad \text{there} \end{bmatrix} , \ \boxed{2} \ , \ \begin{bmatrix} \text{PRED} \quad + \\ \\ \text{VAL} \quad \begin{bmatrix} \text{SPR} \qquad \langle \ \boxed{2} \ \rangle \\ \text{COMPS} \quad \langle \ \rangle \end{bmatrix} \\ \text{SEM} \quad [\text{INDEX} \ \ s \ ] \end{bmatrix} \right\rangle \\ \\ \text{SEM} \quad \begin{bmatrix} \text{INDEX} \quad s \\ \text{RESTR} \quad \langle \ \rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

Questions About the Existential *be*

- What type of constituent is the third argument?
- Why is the third argument [PRED **+**]?
- Why is the second argument tagged as identical to the SPR of the third argument?
- What is the contribution of this *be* to the semantics of the sentences it occurs in?
- Can all [PRED **+**] predicates appear as the third argument in existentials?

$$
\left\langle \text{be} , \begin{bmatrix} \textit{exist-be-lxm} \\[4pt] \text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{NP} \\ \text{FORM} \quad \text{there} \end{bmatrix} , \boxed{2} , \begin{bmatrix} \text{PRED} & + \\[4pt] \text{VAL} & \begin{bmatrix} \text{SPR} & \langle\, \boxed{2}\, \rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \\[8pt] \text{SEM} & [\text{INDEX} \quad s] \end{bmatrix} \right\rangle \\[20pt] \text{SEM} \quad \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \langle\,\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

The Entry for Existential *there*

$$
\left\langle \text{there} \; , \; \begin{bmatrix} \textit{pron-lxm} \\[2ex] \text{SYN} \quad \begin{bmatrix} \text{HEAD} \quad \begin{bmatrix} \text{FORM} \quad \textit{there} \\[1ex] \text{AGR} \quad \begin{bmatrix} \text{PER} \quad \text{3rd} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\[4ex] \text{SEM} \quad \begin{bmatrix} \text{MODE} \quad \textit{none} \\ \text{INDEX} \quad \textit{none} \\ \text{RESTR} \quad \langle \; \rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

Questions About Existential *there*

- Why do we call it a pronoun?

- Why don't we give it a value for NUM?

- What does this entry claim is *there*'s contribution to the semantics of the sentences it appears in?
  Is this a correct claim?

$$\left\langle \text{there} \, , \begin{bmatrix} pron\text{-}lxm \\ \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{FORM} & \text{there} \\ \text{AGR} & \begin{bmatrix} \text{PER} & \text{3rd} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \\ \text{SEM} & \begin{bmatrix} \text{MODE} & \text{none} \\ \text{INDEX} & \text{none} \\ \text{RESTR} & \langle \, \rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

Sample tree for existential *be*
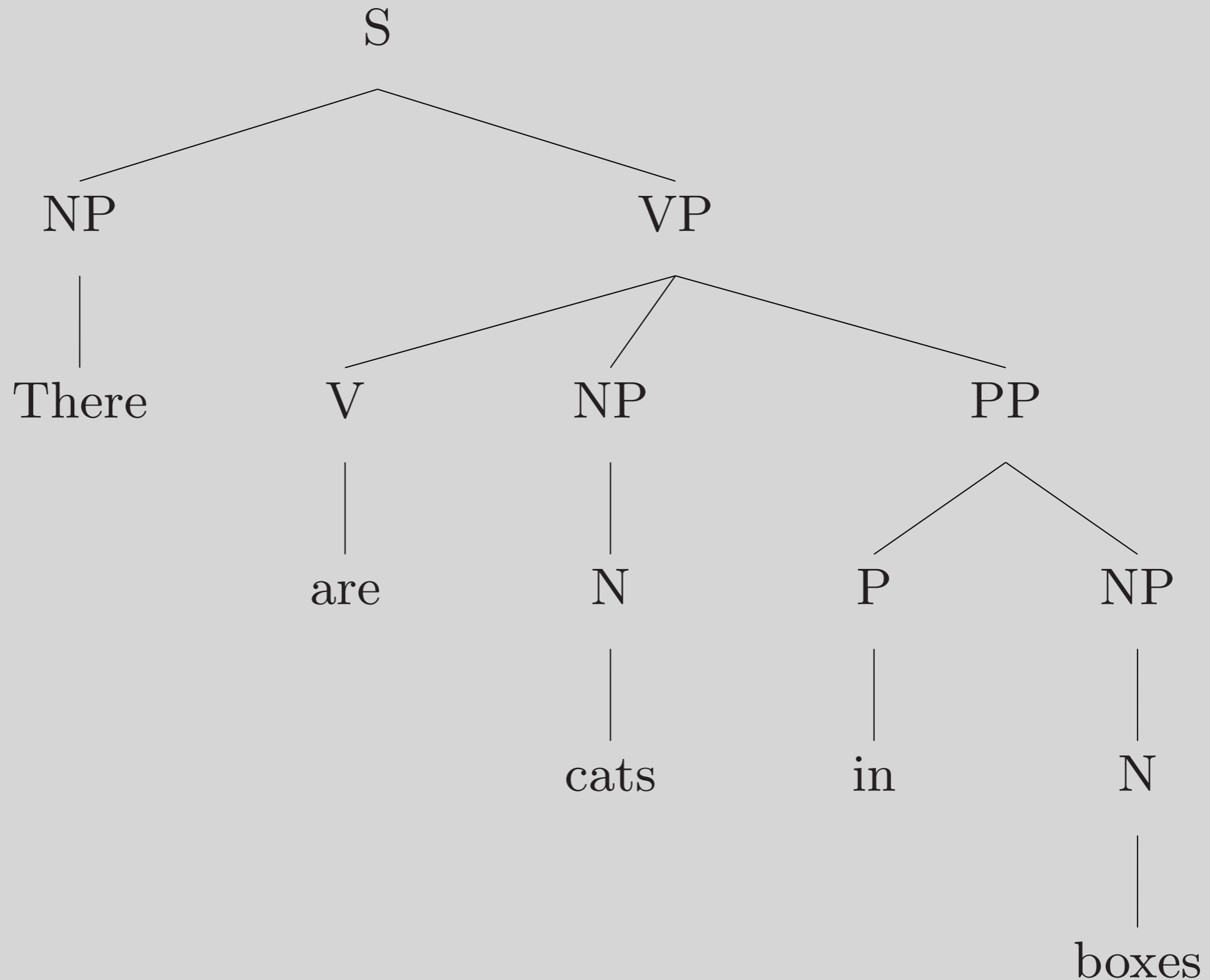
Other NPs that don't seem to refer

- *It sucks that the Rockies lost the series.*

- *It is raining.*

- *Andy took advantage of the opportunity.*

- *Lou kicked the bucket.*

What we need to deal with examples like *It follows that you are wrong*

- A lexical entry for this dummy *it*

- An analysis of this use of *that*

- Entries for verbs that take clausal subjects (as in *That you are wrong follows*)

- A rule to account for the relationship between pairs like *That you are wrong follows* and *It follows that you are wrong*

The Entry for Dummy *it*

$$\left\langle \text{it,} \begin{bmatrix} \textit{pron-lxm} \\ \\ \text{SYN} \quad \begin{bmatrix} \text{HEAD} \quad \begin{bmatrix} \text{FORM} & \text{it} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \end{bmatrix} \\ \\ \text{SEM} \quad \begin{bmatrix} \text{MODE} & \text{none} \\ \text{INDEX} & \text{none} \\ \text{RESTR} & \langle\,\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

Questions About Dummy *it*

- How does it differ from the entry for dummy *there*? Why do they differ in this way?

- Is this the only entry for *it*?

$$
\left\langle \text{it,} \begin{bmatrix} \textit{pron-lxm} \\ \\ \text{SYN} \quad \begin{bmatrix} \text{HEAD} \quad \begin{bmatrix} \text{FORM} & \text{it} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \end{bmatrix} \\ \\ \text{SEM} \quad \begin{bmatrix} \text{MODE} & \text{none} \\ \text{INDEX} & \text{none} \\ \text{RESTR} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$
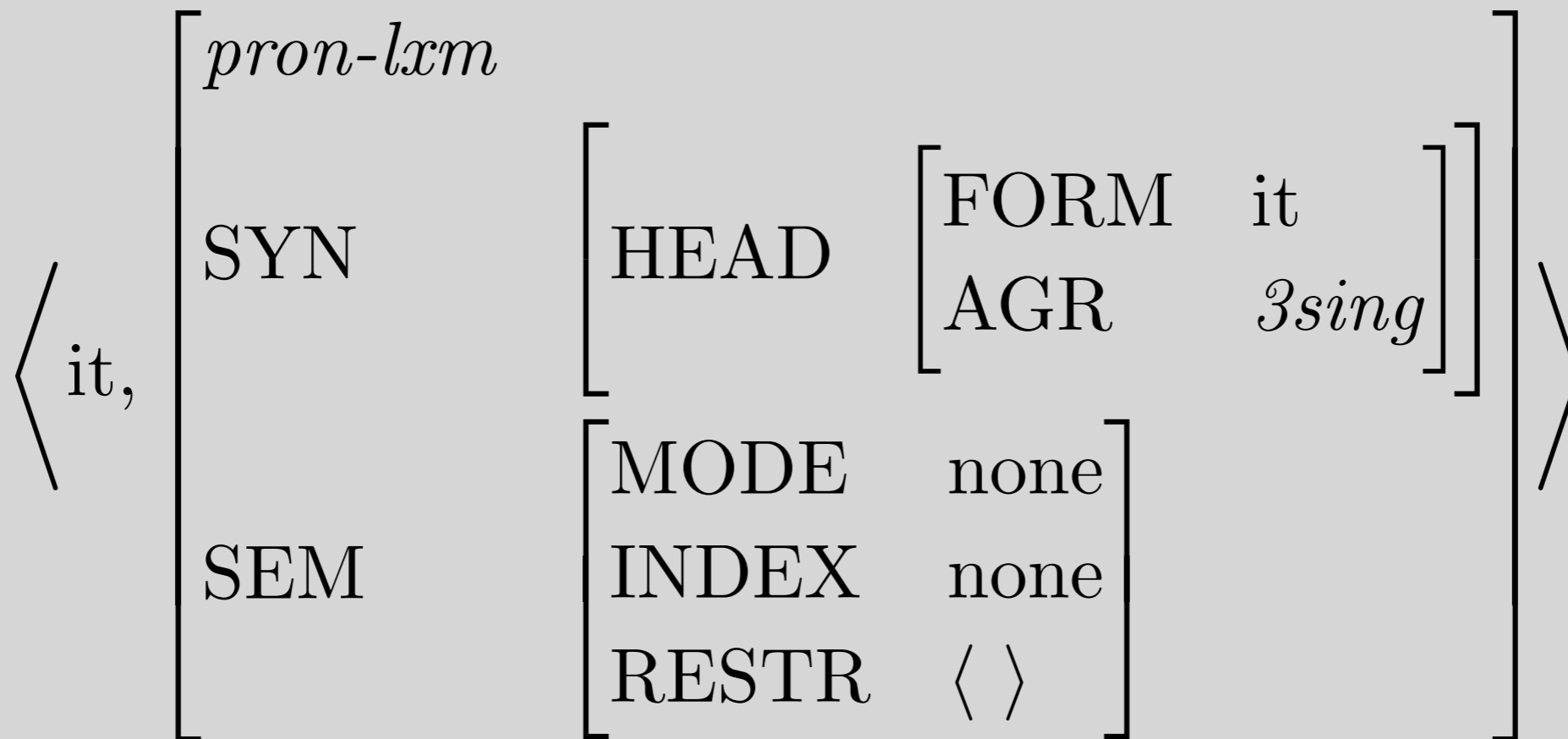
A New Type of Lexeme: Complementizers

$$
\textit{comp-lxm} : \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{comp} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle \ \rangle \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{S} \\ \text{INDEX} & s \end{bmatrix} \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \langle \ \rangle \end{bmatrix} \end{bmatrix}
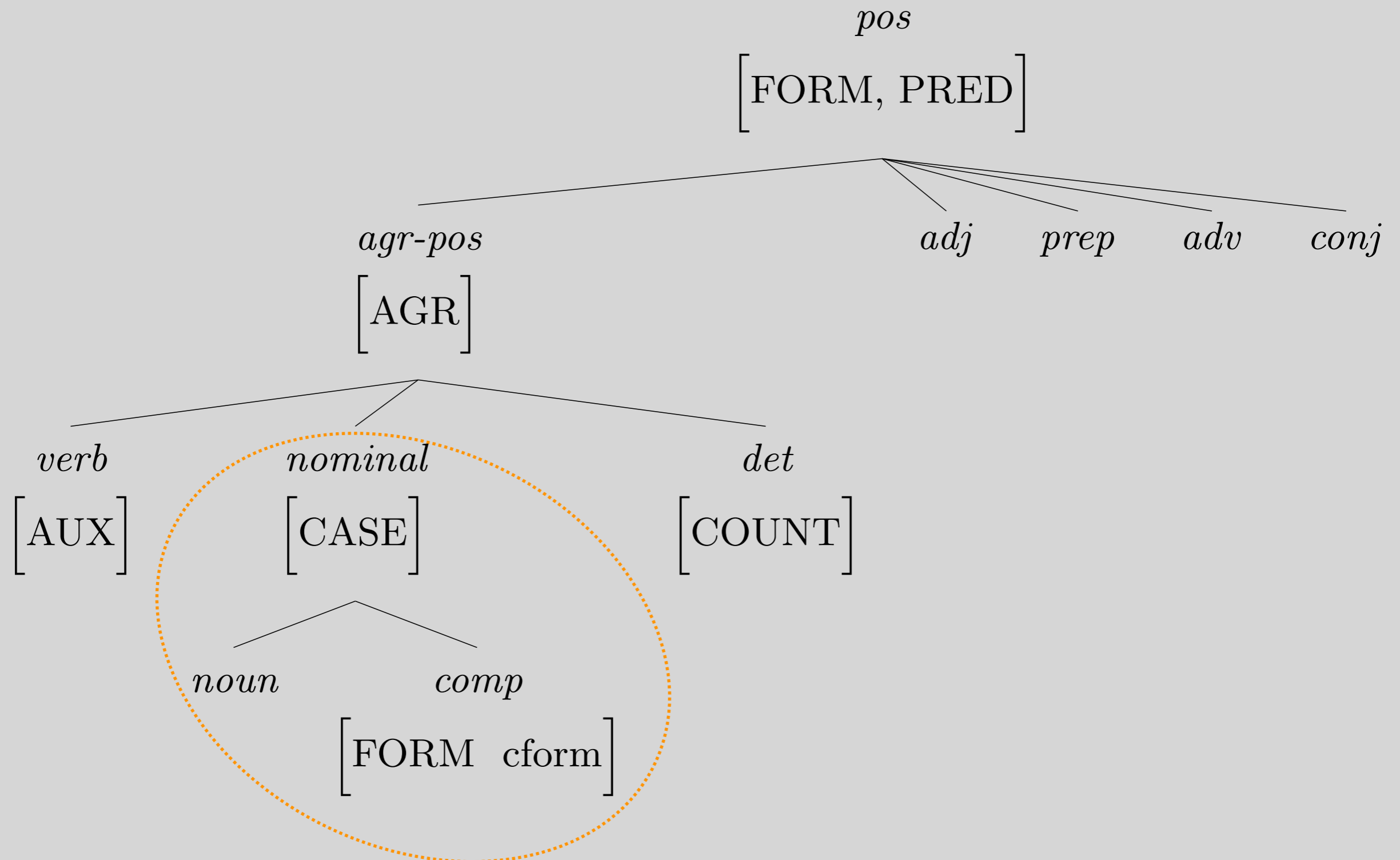$$

Questions About the Type *comp-lxm*

- Why does it stipulate values for both SPR and ARG-ST?

- Why is its INDEX value the same as its argument's?

- What is its semantic contribution?

$$
\textit{comp-lxm} : \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{comp} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle \, \rangle \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{array}{c} \text{S} \\ \begin{bmatrix} \text{INDEX} & s \end{bmatrix} \end{array} \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \langle \, \rangle \end{bmatrix} \end{bmatrix}
$$

The Type *comp*

pos

$$\left[\text{FORM, PRED} \right]$$

agr-pos

$$\left[\text{AGR} \right]$$

adj   *prep*   *adv*   *conj*

verb

$$\left[\text{AUX} \right]$$

nominal

$$\left[\text{CASE} \right]$$

det

$$\left[\text{COUNT} \right]$$

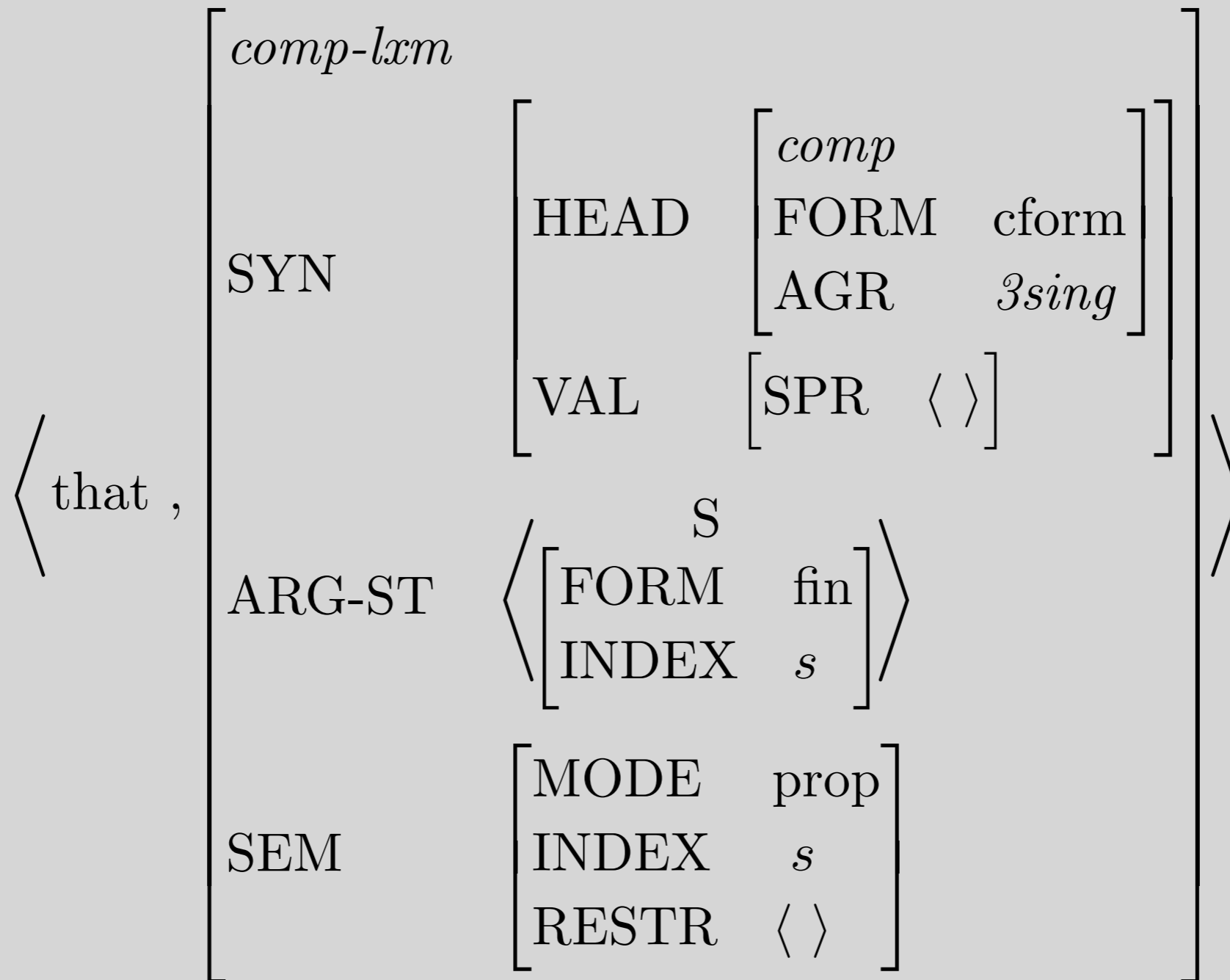*noun*   *comp*

$$\left[\text{FORM \quad cform} \right]$$

The Lexical Entry for Complementizer *that*

$$
\left\langle \text{that} \ , \
\begin{bmatrix}
\textit{comp-lxm} \\[4pt]
\text{ARG-ST} & \left\langle \begin{bmatrix} \text{FORM fin} \end{bmatrix} \right\rangle \\[8pt]
\text{SEM} & \begin{bmatrix} \text{MODE prop} \end{bmatrix}
\end{bmatrix}
\right\rangle
$$

…and with inherited information filled in

$$
\left\langle \text{that} , \begin{bmatrix} \textit{comp-lxm} \\[4pt] \text{SYN} \quad \begin{bmatrix} \text{HEAD} \quad \begin{bmatrix} \textit{comp} \\ \text{FORM} \quad \text{cform} \\ \text{AGR} \quad \textit{3sing} \end{bmatrix} \\[8pt] \text{VAL} \quad \begin{bmatrix} \text{SPR} \quad \langle\,\rangle \end{bmatrix} \end{bmatrix} \\[12pt] \text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{S} \\ \text{FORM} \quad \text{fin} \\ \text{INDEX} \quad s \end{bmatrix} \right\rangle \\[12pt] \text{SEM} \quad \begin{bmatrix} \text{MODE} \quad \text{prop} \\ \text{INDEX} \quad s \\ \text{RESTR} \quad \langle\,\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

Question: Where did [FORM cform] come from?

Structure of a Complementizer Phrase
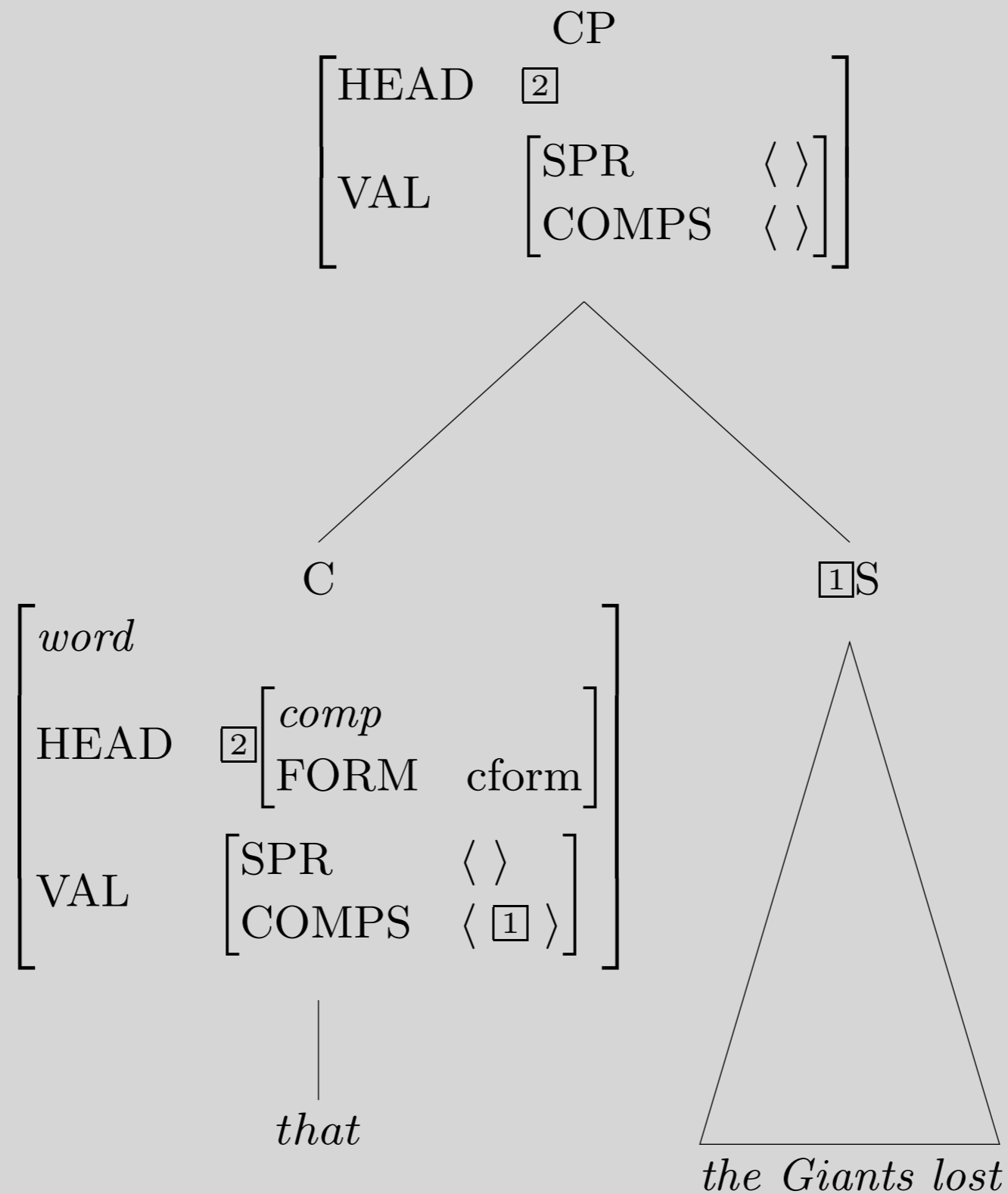
$$
\text{CP}
\begin{bmatrix}
\text{HEAD} & \boxed{2} \\
\text{VAL} & \begin{bmatrix} \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix}
\end{bmatrix}
$$

C
$$
\begin{bmatrix}
word \\
\text{HEAD} & \boxed{2}\begin{bmatrix} comp \\ \text{FORM} & \text{cform} \end{bmatrix} \\
\text{VAL} & \begin{bmatrix} \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\boxed{1}\,\rangle \end{bmatrix}
\end{bmatrix}
$$

$\boxed{1}$S

that

the Giants lost

What is your impression of treating complementizers as the head of constituents like this one?

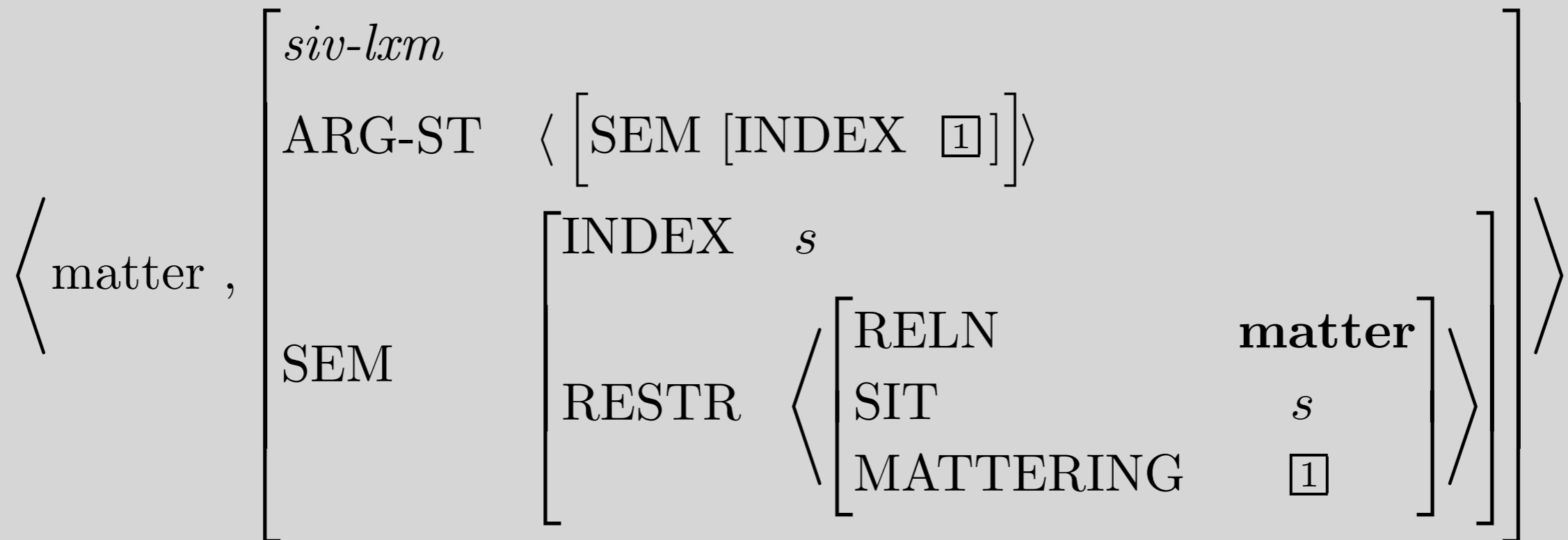Fine: I know CPs from another class

New to me but seems legit

Not sure

Seems strange, since the main part is the S

Total Results: 0

Sample Verb with a CP Subject

$$
\left\langle \text{matter} ,
\begin{bmatrix}
\textit{siv-lxm} \\
\text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{SEM [INDEX} \quad \boxed{1}] \end{bmatrix} \right\rangle \\
\text{SEM} \quad
\begin{bmatrix}
\text{INDEX} \quad s \\
\text{RESTR} \quad \left\langle
\begin{bmatrix}
\text{RELN} & \textbf{matter} \\
\text{SIT} & s \\
\text{MATTERING} & \boxed{1}
\end{bmatrix}
\right\rangle
\end{bmatrix}
\end{bmatrix}
\right\rangle
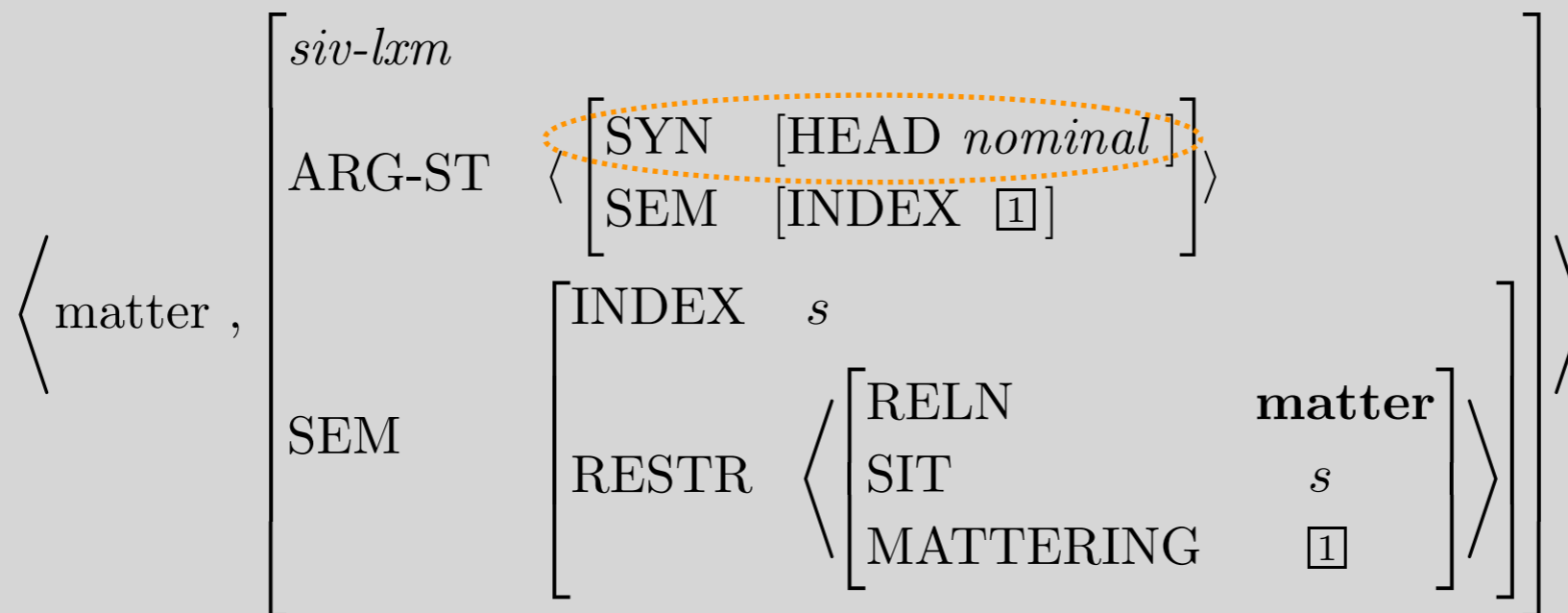$$

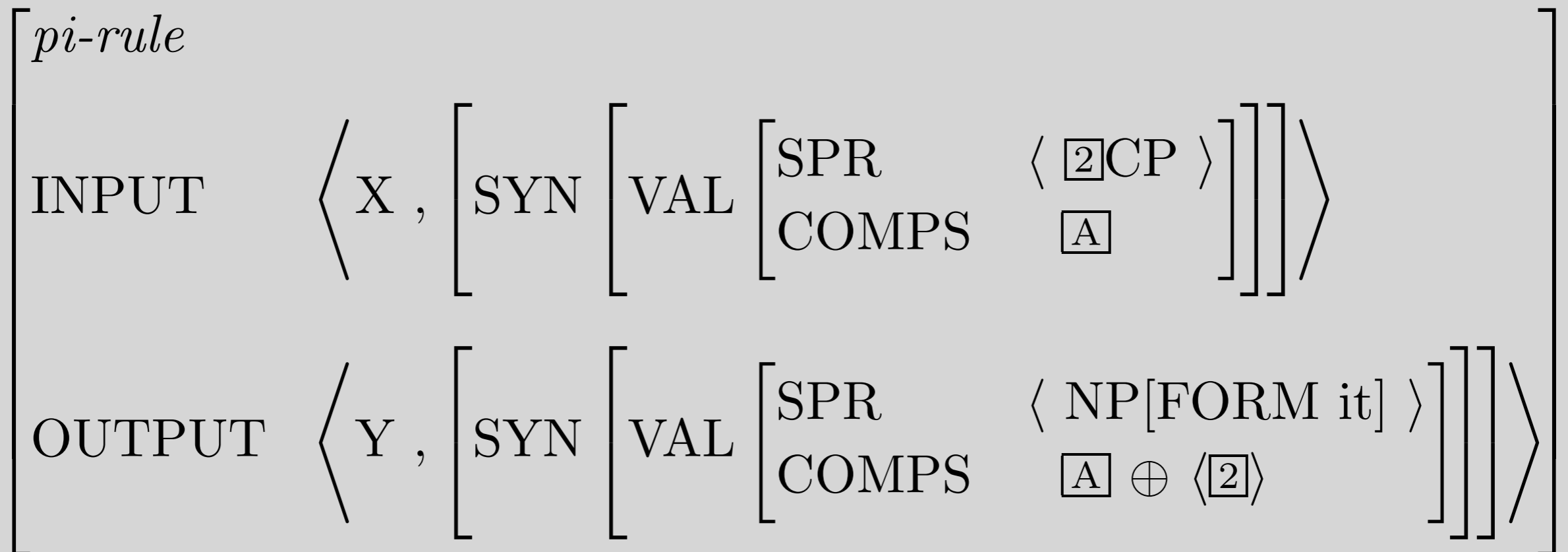Note: the only constraint on the first argument is semantic

A Problem

- We constrained the subject of *matter* only semantically. However...
  - CP and S are semantically identical, but we get:
    That Bush won matters vs. **Bush won matters*
  - Argument-marking PPs are semantically identical to their object NPs, but we get:
    The election mattered vs. **Of the election mattered*
- So we need to add a syntactic constraint.

$$
\left\langle \text{matter} , \begin{bmatrix} \textit{siv-lxm} \\ \\ \text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{SYN} & [\text{HEAD } \textit{nominal}] \\ \text{SEM} & [\text{INDEX } \boxed{1}] \end{bmatrix} \right\rangle \\ \\ \text{SEM} \quad \begin{bmatrix} \text{INDEX} & s \\ \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \textbf{matter} \\ \text{SIT} & s \\ \text{MATTERING} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

- S and PP subjects are generally impossible, so this constraint belongs on *verb-lxm*.

The Extraposition Lexical Rule

$$
\begin{bmatrix}
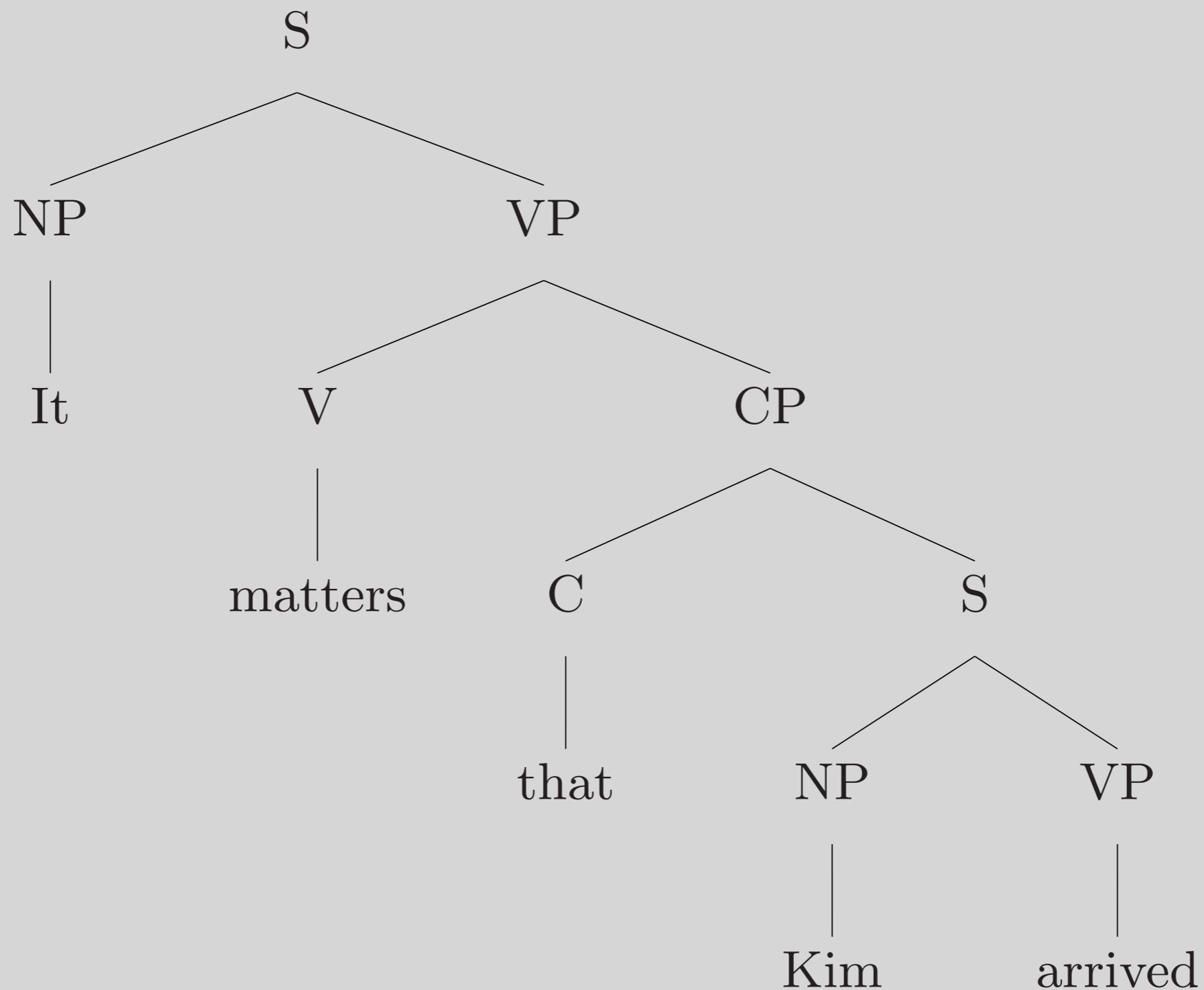\textit{pi-rule} \\[2ex]
\text{INPUT} \quad \left\langle \text{X}, \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{SPR} & \langle \boxed{2}\text{CP} \rangle \\ \text{COMPS} & \boxed{A} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\[4ex]
\text{OUTPUT} \quad \left\langle \text{Y}, \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{SPR} & \langle \text{NP[FORM it]} \rangle \\ \text{COMPS} & \boxed{A} \oplus \langle \boxed{2} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

- Why is the type *pi-rule*?
- Why doesn't it say anything about the semantics?
- Why is the COMPS on INPUT $\boxed{A}$, not < >?

Extraposition with Verbs whose COMPS Lists are Nonempty

- *It worries <u>me</u> that war is imminent.*

- *It occurred <u>to Pat</u> that Chris knew the answer.*

- *It endeared <u>you</u> <u>to Andy</u> that you wore a funny hat.*
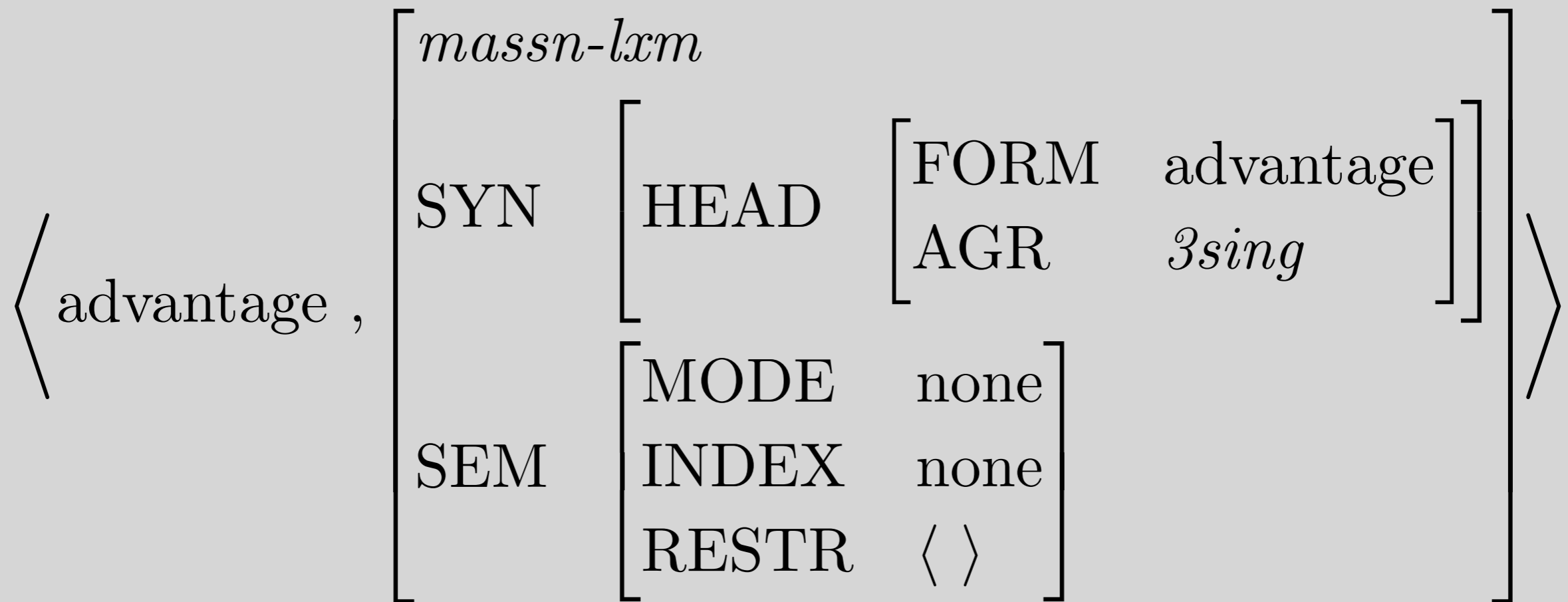
Sample tree with extraposition

```
                        S
              _____/  _____
            NP                     VP
            |              _____/    _____
            It           V                    CP
                         |             _____/  _____
                      matters        C                  S
                                     |            _____/  \_____
                                    that         NP            VP
                                                 |             |
                                                Kim          arrived
```
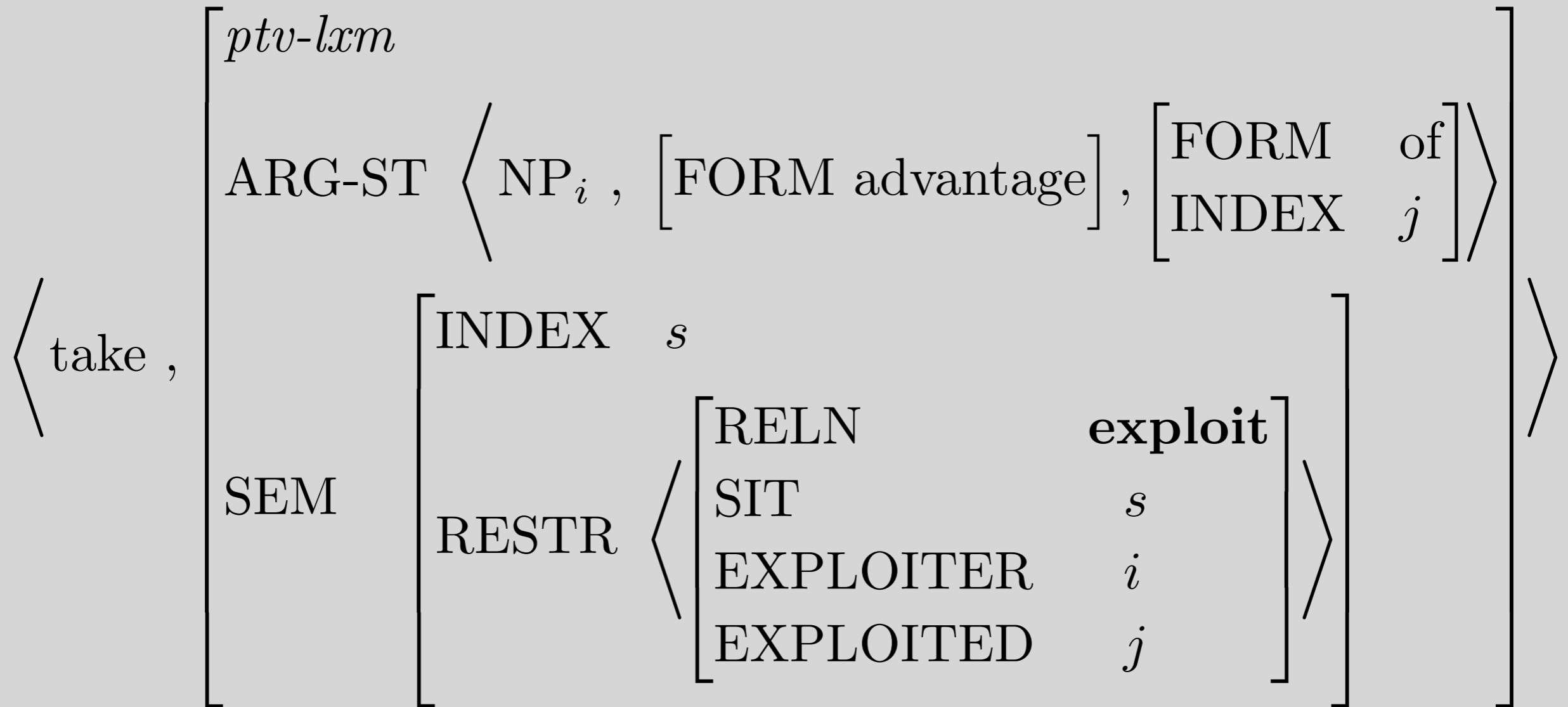
Your favorite (SFW) English idiom?

Total Results: 0

Another Nonreferential Noun

$$
\left\langle \text{advantage} \ , \ \begin{bmatrix} \textit{massn-lxm} \\[6pt] \text{SYN} \ \begin{bmatrix} \text{HEAD} \ \begin{bmatrix} \text{FORM} & \text{advantage} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \end{bmatrix} \\[18pt] \text{SEM} \ \begin{bmatrix} \text{MODE} & \text{none} \\ \text{INDEX} & \text{none} \\ \text{RESTR} & \langle \ \rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

The Verb that Selects *advantage*

$$
\left\langle \text{take} , \begin{bmatrix} \textit{ptv-lxm} \\[2pt] \text{ARG-ST} \left\langle \text{NP}_i , \begin{bmatrix} \text{FORM advantage} \end{bmatrix} , \begin{bmatrix} \text{FORM} & \text{of} \\ \text{INDEX} & j \end{bmatrix} \right\rangle \\[2pt] \text{SEM} \begin{bmatrix} \text{INDEX} & s \\[2pt] \text{RESTR} \left\langle \begin{bmatrix} \text{RELN} & \textbf{exploit} \\ \text{SIT} & s \\ \text{EXPLOITER} & i \\ \text{EXPLOITED} & j \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

Our analyses of idioms and passives interact...

- We generate

  Advantage was taken of the situation by many people.

  Tabs are kept on online activists.

- But not:

  Many people were taken advantage of.

- Why not?

Overview

- Existentials (*there*, *be*)

- Extraposition (*that*, *it*, *LR*)

- Idioms

RQs: *be*

- I'm unsure about how we arrived at the feature structure for be on p. 334 (*be-lxm*). Also why is it necessary to associate the index of the complement of be with its own index (the index of be)?

RQs: *be*

- In adding the existential be to our lexicon, how do we distinguish this entry from the be-lxm entry given on p.320. I understand they have different lexical types, but I thought that applying new types to a lexeme was the territory of Lexical Rules. Are we then considering these two "be"s to be separate words?

- For other verb lexemes that can take on different types through lexical rules, how do we determine which entry the root lexeme (that is to say, the original lexical entry before any alterations) is?

RQs: PRED

- Why are passive and present participles [PRED +]?

RQs: *pi-rule*

- Why do we need a new rule category (pi-rule) to accommodate extraposition? What does this kind of rule do that a d-rule couldn't?

RQs: FORM

- For (47) and (48)—I'm curious whether we can infer that a feature structure is a certain phrase type based on the FORM value given. For example, in 48b's ARG-ST, there is an element that is FORM advantage. Could we assume this is an NP, since FORM advantage is compatible with HEAD noun?

RQs: CP

- I remember spending a lot of time on the introduction of complementizers and complementizer phrases in a previous syntax class, so I feel as though it would help to go more in-depth in class/during lecture on the role they play in some other contexts (i.e. are there any places in our grammar where, up until Chapter 11, we might not have used CPs and something else that "worked for now"?

RQs: Idioms

- I understand that 50a and 50b can have idiomatic interpretations but how do we know whether the structure should be analyzed as an idiom or not since they are not grouped together as an idiom chunk as the previous examples (45a, 45b, 45c)?

- I don't quite understand why we need to consider idioms in the first place. It seems like a semantics concern. Why do we need to consider the difference in meaning between "kick the bucket" and "the bucket was kicked"?

RQs: Idioms

- How might a parser know that a particular idiom is not in fact being used for its idiom meaning if the sequence is exactly the same? I.e.:

- *My toddler always wants to kick the bucket of halloween candy over.*

- Can the idiom analysis be disallowed by adding restrictions on ARG-ST (for the idiom usage)?

RQs: Idioms

- The candidates take (unfair) advantage of the voters.

- What are the semantics of the modifier of idiomatic nouns such as unfair here? Since the INDEX of the modified is none, is its semantics similar to adverbial modifiers like today (that is, its MODE is none and RESTR includes only one predicate which includes RELN, SIT and ARG)?

RQs: Idioms

- I still don't understand why some idioms have an empty RESTR list but "kick the bucket" doesn't. Could you expand more on that?

RQs: Semantics

- Also, is there ever an intersection with HPSG and lambda calculus? (Or perhaps just a case where at ever feature structure, you might write out the lamba calculus representation next to it or similar). Thinking of the existential terms, I feel like a lot of SEM features built up could be straightforward-ly translated into lambda calculus structure.

RQs: Implementations

- How do existing implementations of HPSG handle the fact that new idioms are constantly being formed? I'd imagine it's not the same as new words being introduced, since those would be subject to our rules and constraints, while idioms feel more freeform and difficult to constrain. Do we see patterns in the syntactic structure of idioms and how they fit into sentences, or do we have to tackle each idiom individually?