Ling 566
Nov 10, 2022

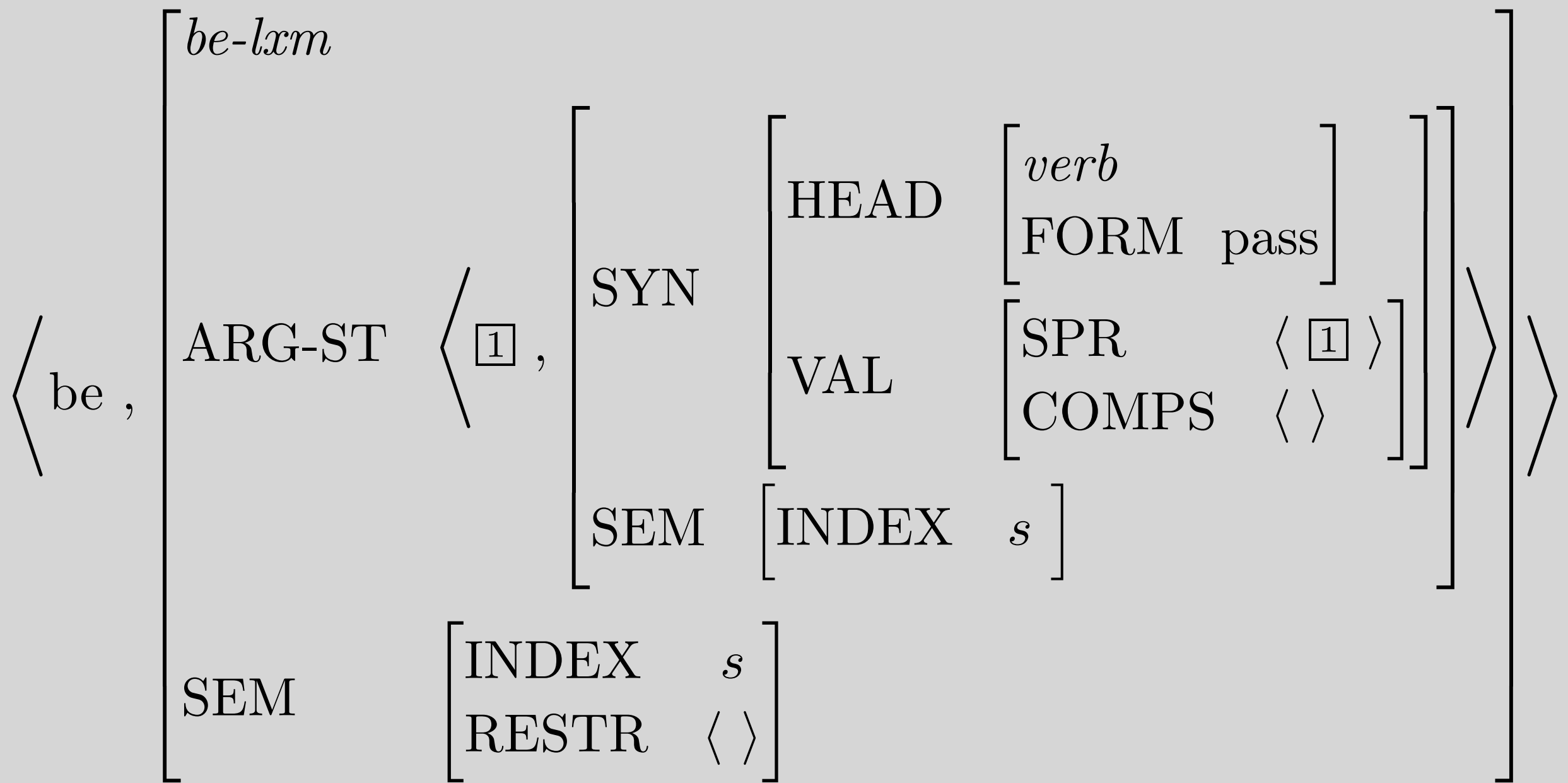Non-referential NPs, Expletives, and Extraposition

Overview

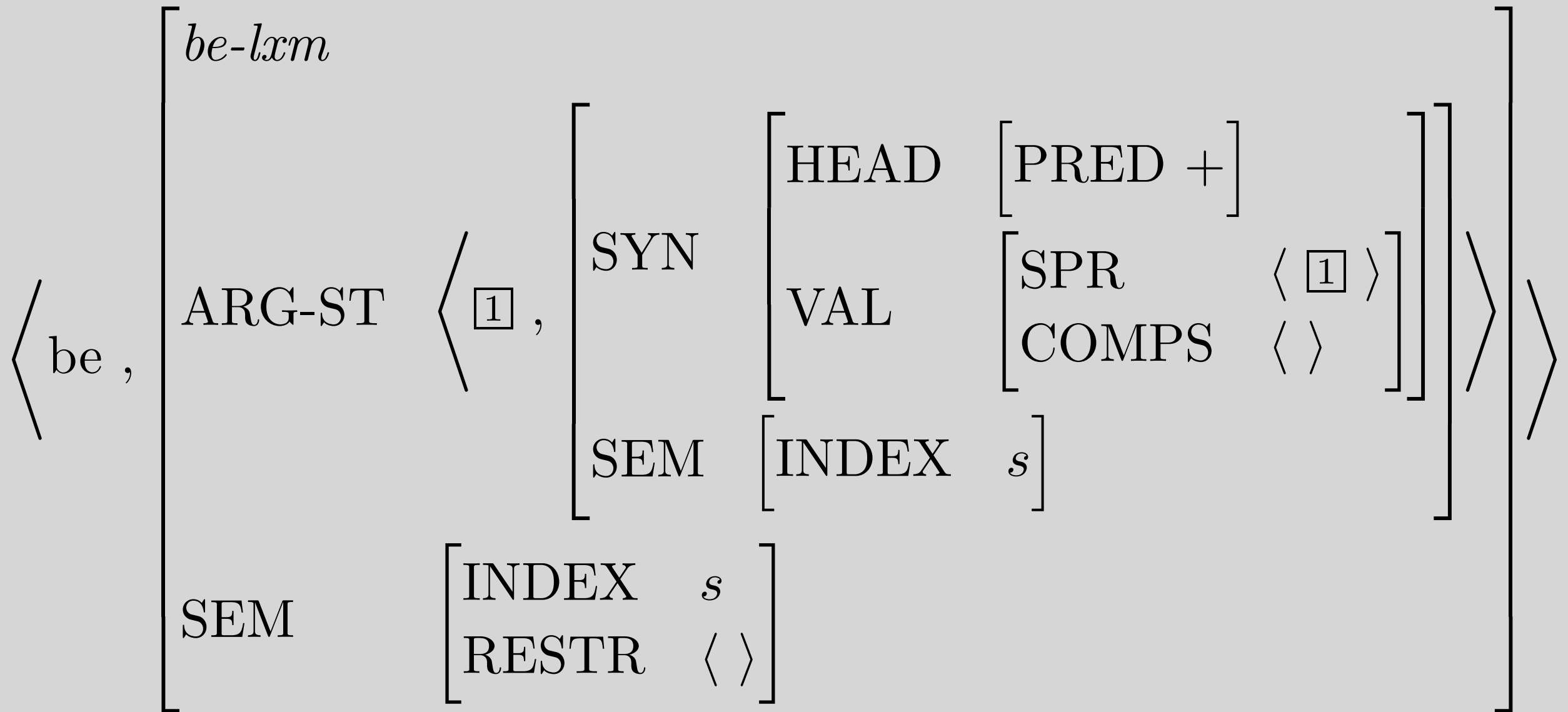- Existentials

- 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.
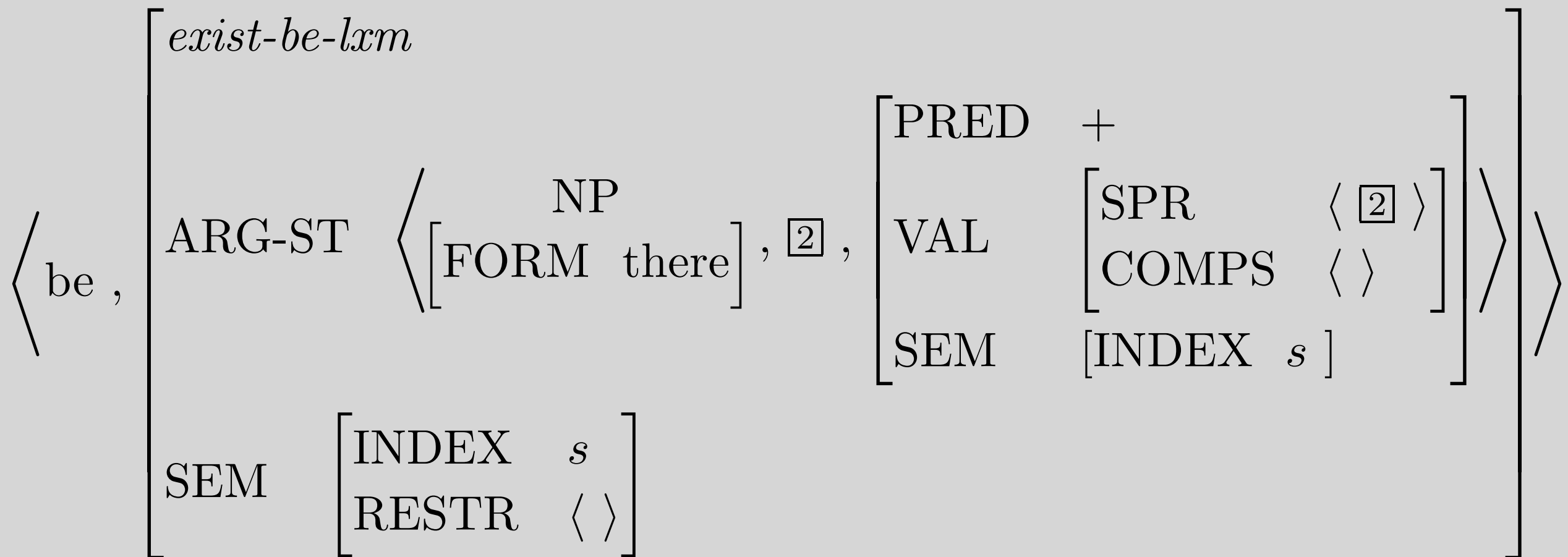
Chapter 10 entry for *be*

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

Copula (generalized)

$$
\left\langle \text{be} \, , \, \begin{bmatrix} \textit{be-lxm} \\[2em] \text{ARG-ST} \quad \left\langle \boxed{1} \, , \, \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{PRED} + \end{bmatrix} \\[1em] \text{VAL} & \begin{bmatrix} \text{SPR} & \langle \, \boxed{1} \, \rangle \\ \text{COMPS} & \langle \, \rangle \end{bmatrix} \end{bmatrix} \\[2em] \text{SEM} & \begin{bmatrix} \text{INDEX} & s \end{bmatrix} \end{bmatrix} \right\rangle \\[3em] \text{SEM} \quad \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \langle \, \rangle \end{bmatrix} \end{bmatrix} \right\rangle
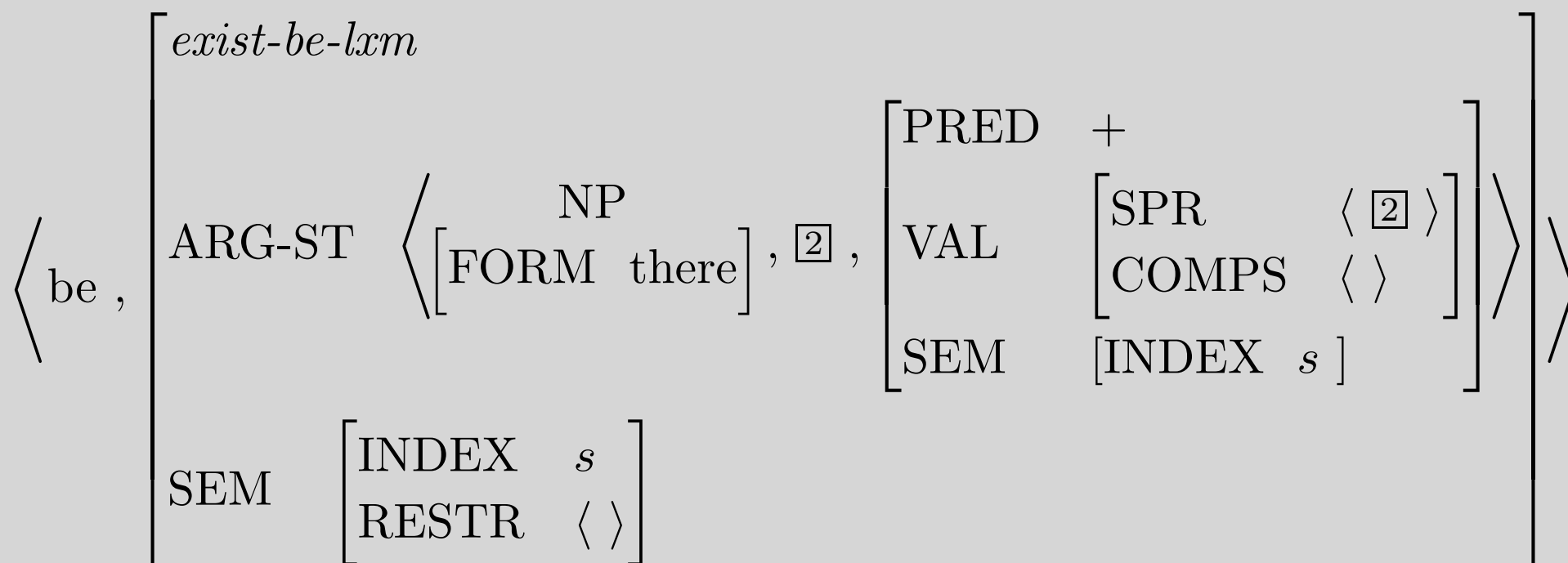$$

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*

$$
\left\langle \text{be} \;,\; \begin{bmatrix} \textit{exist-be-lxm} \\[2em] \text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{NP} \\ \text{FORM} \quad \text{there} \end{bmatrix} ,\; \boxed{2} \;,\; \begin{bmatrix} \text{PRED} \quad + \\[1em] \text{VAL} \quad \begin{bmatrix} \text{SPR} \quad\quad \langle\, \boxed{2} \,\rangle \\ \text{COMPS} \quad \langle\,\rangle \end{bmatrix} \\[1.5em] \text{SEM} \quad [\text{INDEX} \;\; s\,] \end{bmatrix} \right\rangle \\[3em] \text{SEM} \quad \begin{bmatrix} \text{INDEX} \quad s \\ \text{RESTR} \quad \langle\,\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

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?

$$
\left\langle \text{be} , \begin{bmatrix} \textit{exist-be-lxm} \\[4pt] \text{ARG-ST} \left\langle \begin{bmatrix} \text{NP} \\ \text{FORM} \quad \text{there} \end{bmatrix}, \boxed{2}, \begin{bmatrix} \text{PRED} \quad + \\ \text{VAL} \quad \begin{bmatrix} \text{SPR} \quad \langle\, \boxed{2}\, \rangle \\ \text{COMPS} \quad \langle\, \rangle \end{bmatrix} \\ \text{SEM} \quad [\text{INDEX} \quad s\,] \end{bmatrix} \right\rangle \\[12pt] \text{SEM} \quad \begin{bmatrix} \text{INDEX} \quad s \\ \text{RESTR} \quad \langle\, \rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

The Entry for Existential *there*

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

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?

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

Sample tree for existential *be*

```
                          S
                    ┌─────┴─────┐
                   NP           VP
                    │      ┌─────┼──────────┐
                 There     V    NP          PP
                           │     │      ┌────┴────┐
                          are    N      P        NP
                                 │      │         │
                                cats    in        N
                                                  │
                                                boxes
```
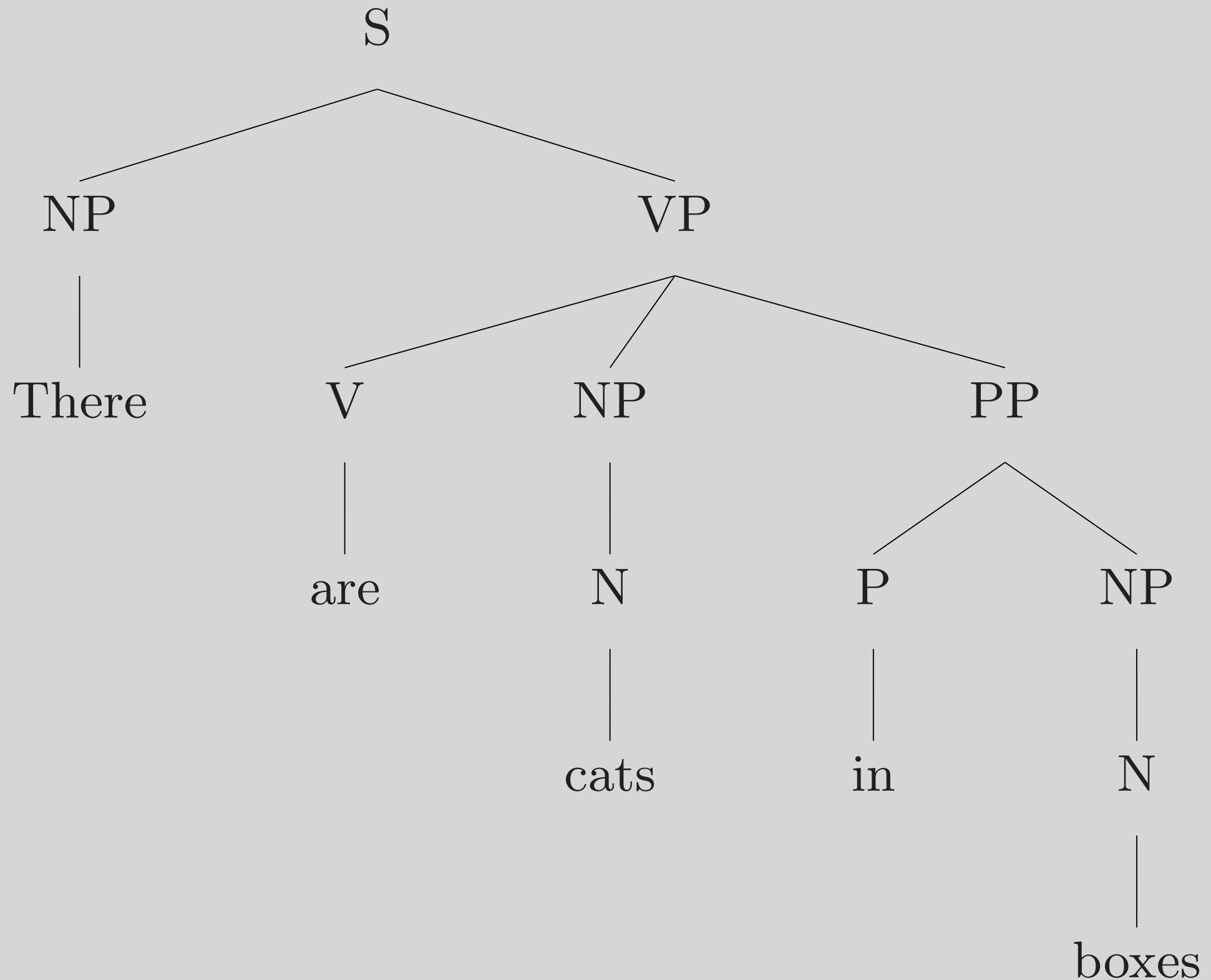
Other NPs that don't seem to refer

- *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*

$$\left\langle \text{it,} \begin{bmatrix} \textit{pron-lxm} \\[1em] \text{SYN} \quad \begin{bmatrix} \text{HEAD} \quad \begin{bmatrix} \text{FORM} & \text{it} \\ \text{AGR} & \textit{3sing} \end{bmatrix} \end{bmatrix} \\[2em] \text{SEM} \quad \begin{bmatrix} \text{MODE} & \text{none} \\ \text{INDEX} & \text{none} \\ \text{RESTR} & \langle\,\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

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*?

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

A New Type of Lexeme: Complementizers

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

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?

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

The Type *comp*

pos

$$\begin{bmatrix} \text{FORM, PRED} \end{bmatrix}$$

agr-pos

$$\begin{bmatrix} \text{AGR} \end{bmatrix}$$

adj   *prep*   *adv*   *conj*

verb

$$\begin{bmatrix} \text{AUX} \end{bmatrix}$$

nominal

$$\begin{bmatrix} \text{CASE} \end{bmatrix}$$

det

$$\begin{bmatrix} \text{COUNT} \end{bmatrix}$$

noun   *comp*

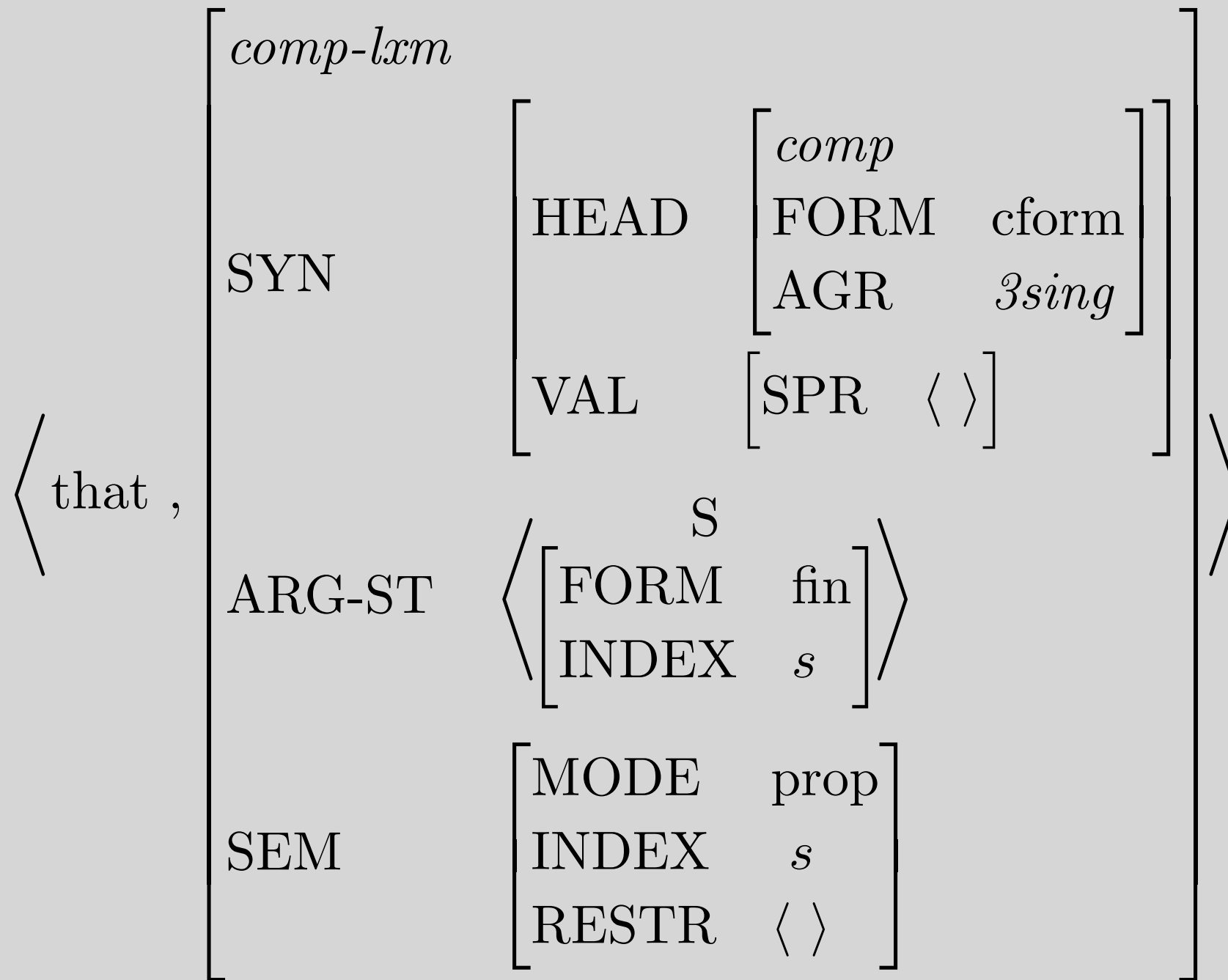$$\begin{bmatrix} \text{FORM} & \text{cform} \end{bmatrix}$$
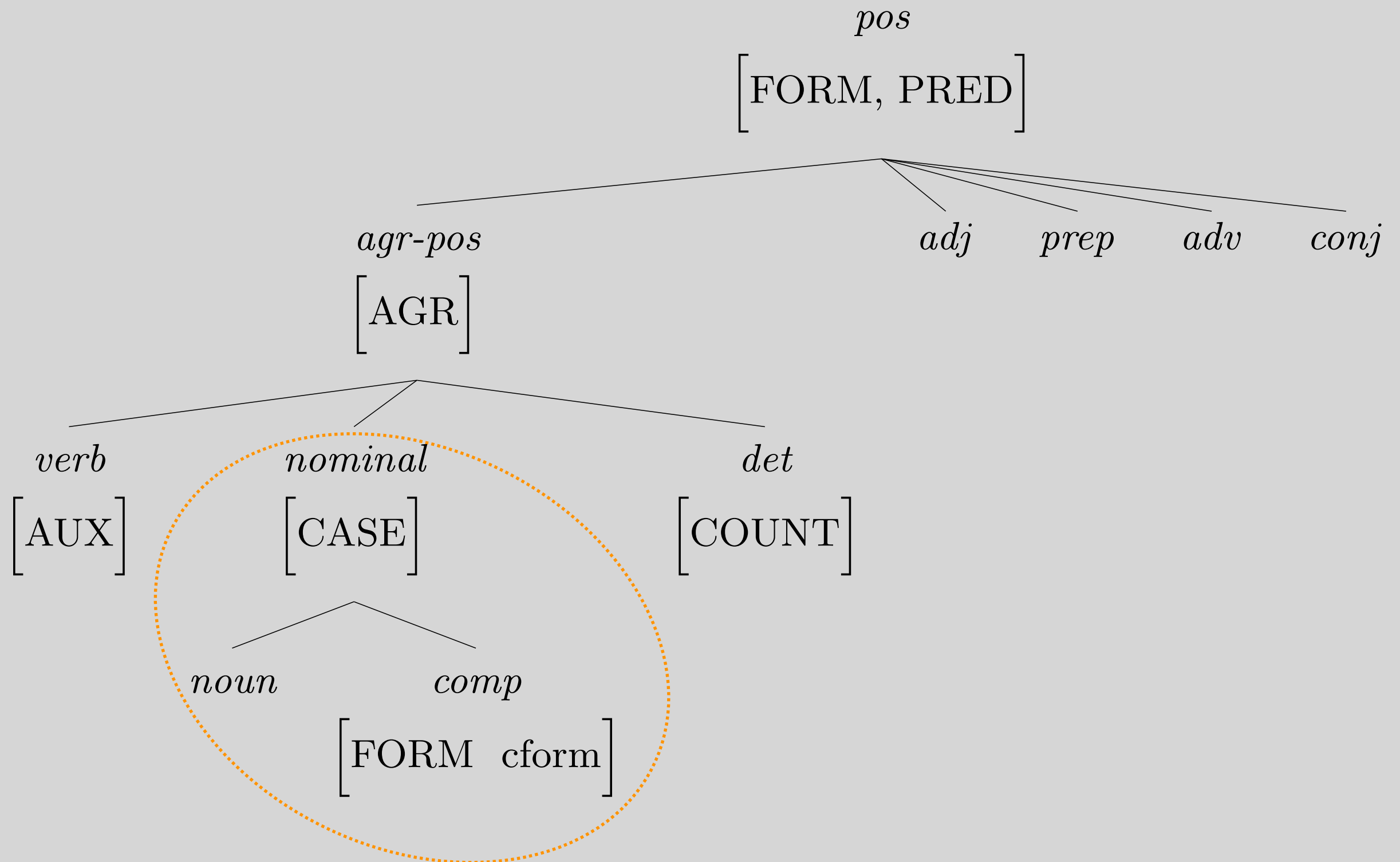
The Lexical Entry for Complementizer *that*

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

…and with inherited information filled in

$$
\left\langle \text{that} \ , \
\begin{bmatrix}
\textit{comp-lxm} \\[2pt]
\text{SYN} \quad
\begin{bmatrix}
\text{HEAD} \quad
\begin{bmatrix}
\textit{comp} \\
\text{FORM} \quad \text{cform} \\
\text{AGR} \quad \textit{3sing}
\end{bmatrix} \\[6pt]
\text{VAL} \quad
\begin{bmatrix}
\text{SPR} \quad \langle \ \rangle
\end{bmatrix}
\end{bmatrix} \\[10pt]
\text{ARG-ST} \quad
\left\langle
\begin{bmatrix}
\text{S} \\
\text{FORM} \quad \text{fin} \\
\text{INDEX} \quad s
\end{bmatrix}
\right\rangle \\[10pt]
\text{SEM} \quad
\begin{bmatrix}
\text{MODE} \quad \text{prop} \\
\text{INDEX} \quad s \\
\text{RESTR} \quad \langle \ \rangle
\end{bmatrix}
\end{bmatrix}
\right\rangle
$$

Question: Where did [FORM cform] come from?

Structure of a Complementizer Phrase
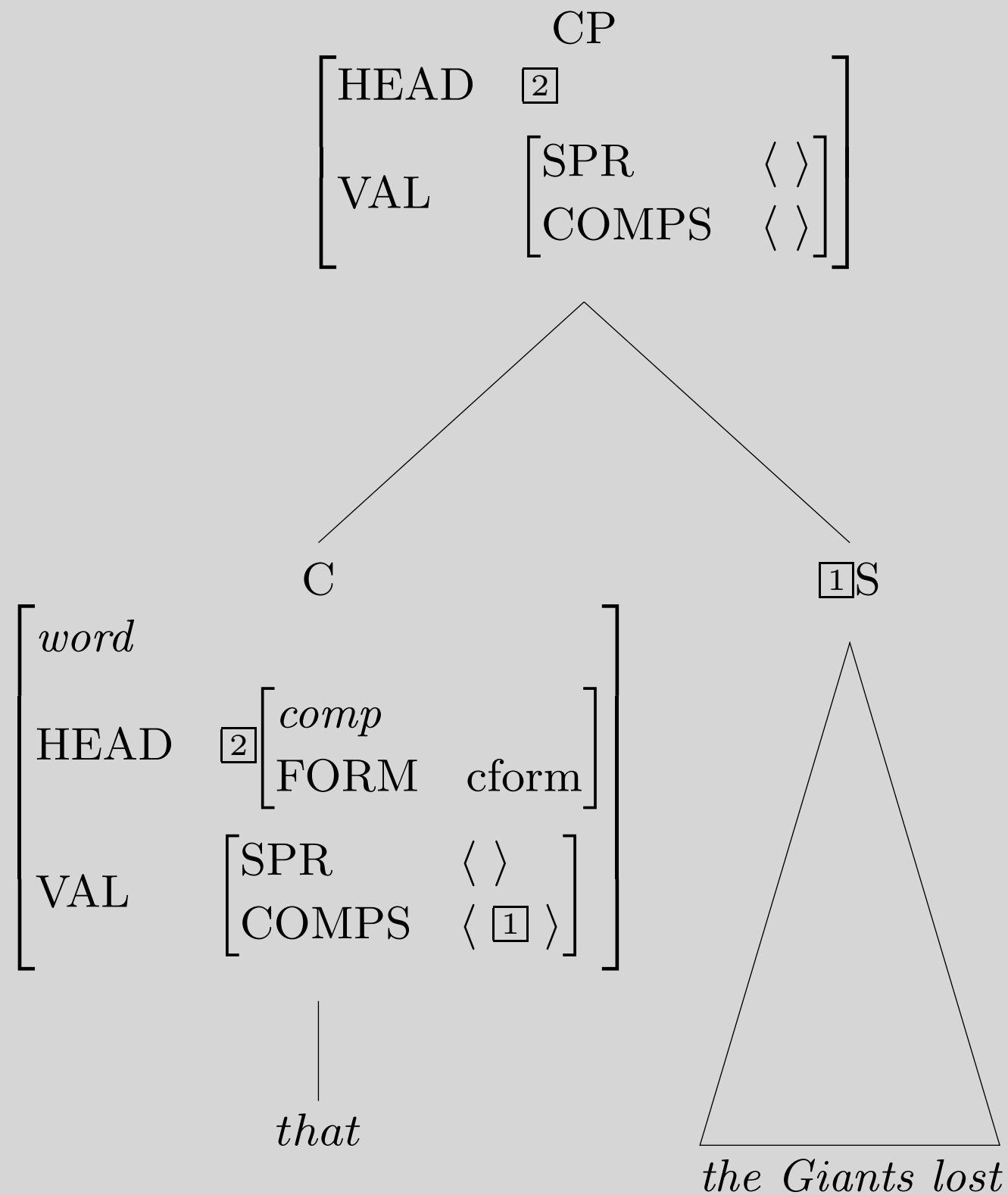
CP

$$\begin{bmatrix} \text{HEAD} & \boxed{2} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle\,\rangle \\ \text{COMPS} & \langle\,\rangle \end{bmatrix} \end{bmatrix}$$

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

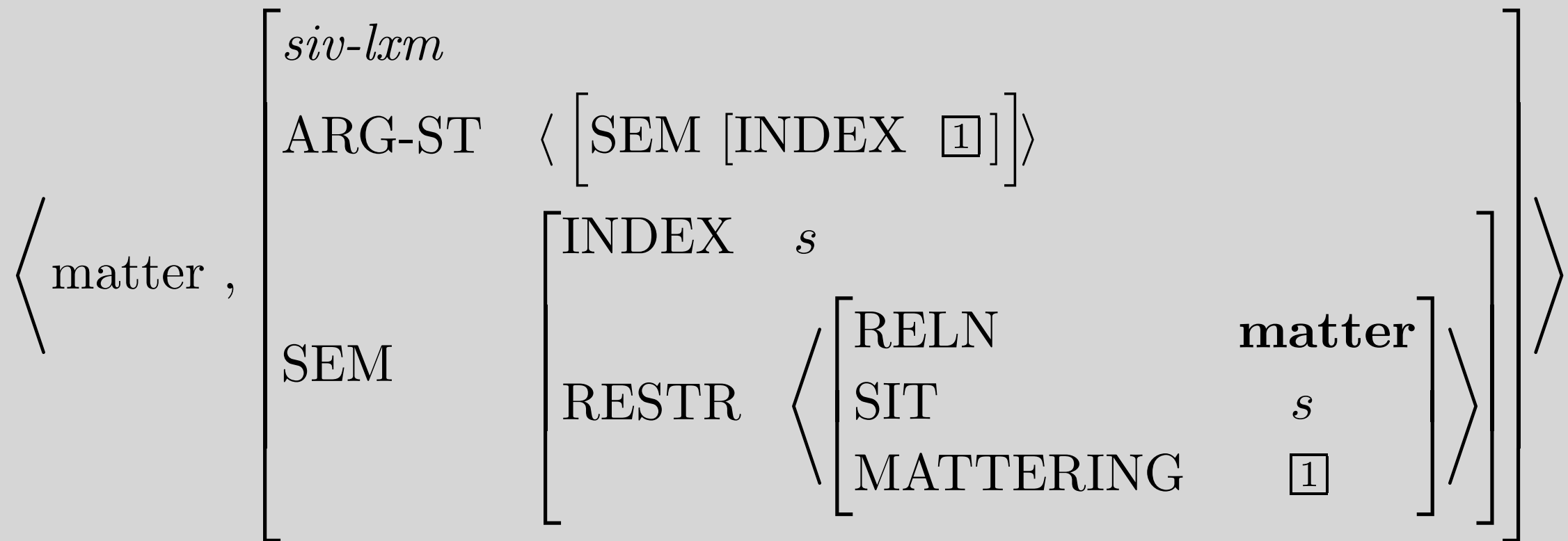# What is your impression of treating complementizers as the head of constituents like this one?

Fine: I know CPs from another class

New to me but seems legit

Not sure
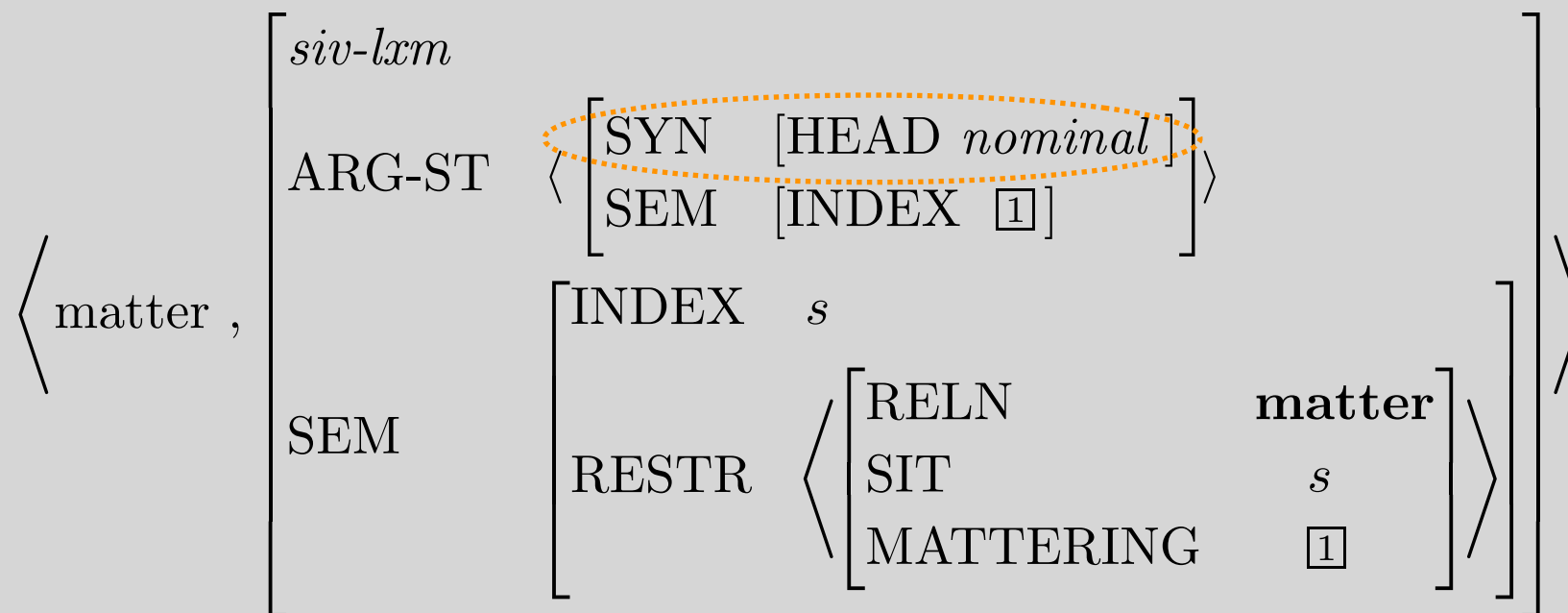
Seems strange, since the main part is the S

Sample Verb with a CP Subject

$$
\left\langle \text{matter} , \begin{bmatrix} \textit{siv-lxm} \\[4pt] \text{ARG-ST} \quad \left\langle \begin{bmatrix} \text{SEM [INDEX } \boxed{1}] \end{bmatrix} \right\rangle \\[12pt] \text{SEM} \quad \begin{bmatrix} \text{INDEX} \quad s \\[10pt] \text{RESTR} \quad \left\langle \begin{bmatrix} \text{RELN} & \textbf{matter} \\ \text{SIT} & s \\ \text{MATTERING} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

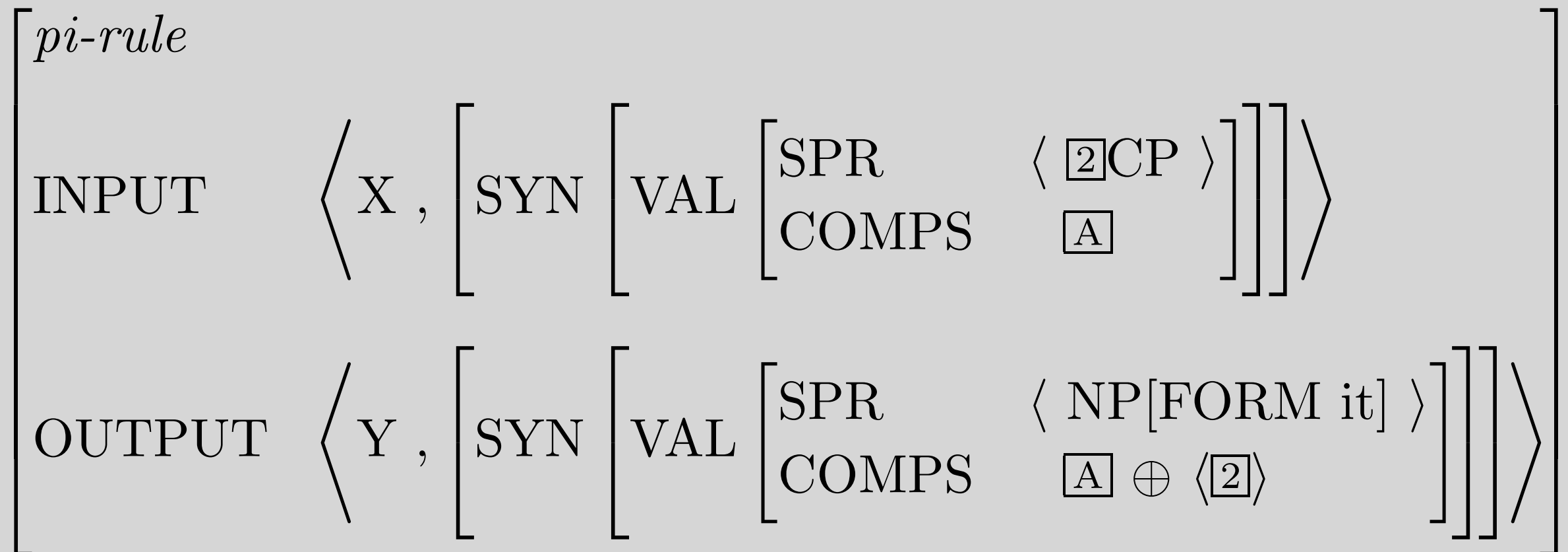Note: the only constraint on the first argument is semantic

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.

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

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

The Extraposition Lexical Rule

$$
\begin{bmatrix}
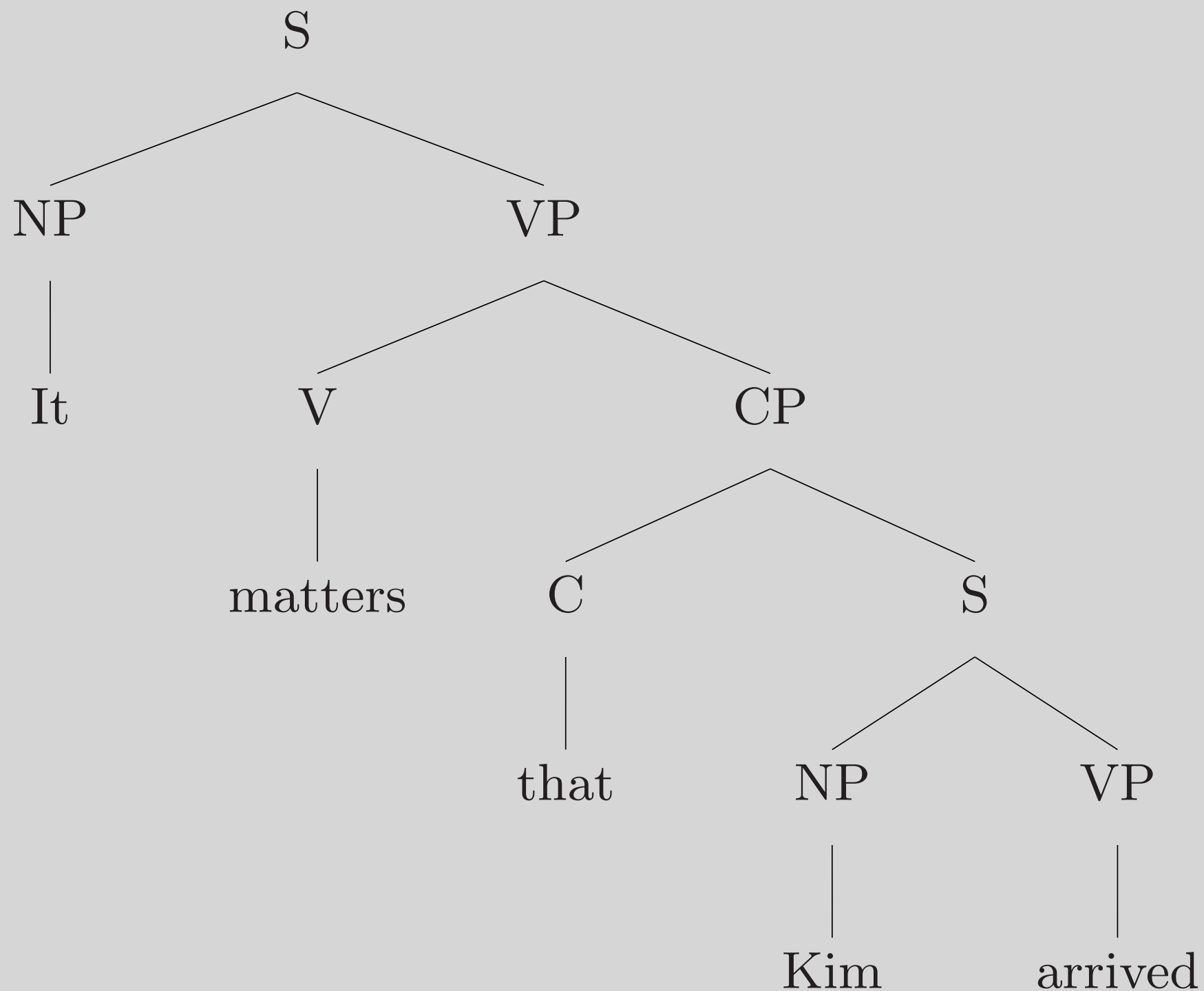\textit{pi-rule} \\[2em]
\text{INPUT} \quad \left\langle \text{X} , \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{SPR} & \left\langle \boxed{2}\text{CP} \right\rangle \\ \text{COMPS} & \boxed{A} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \\[3em]
\text{OUTPUT} \quad \left\langle \text{Y} , \begin{bmatrix} \text{SYN} \begin{bmatrix} \text{VAL} \begin{bmatrix} \text{SPR} & \left\langle \text{NP[FORM it]} \right\rangle \\ \text{COMPS} & \boxed{A} \oplus \left\langle \boxed{2} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

- 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.*
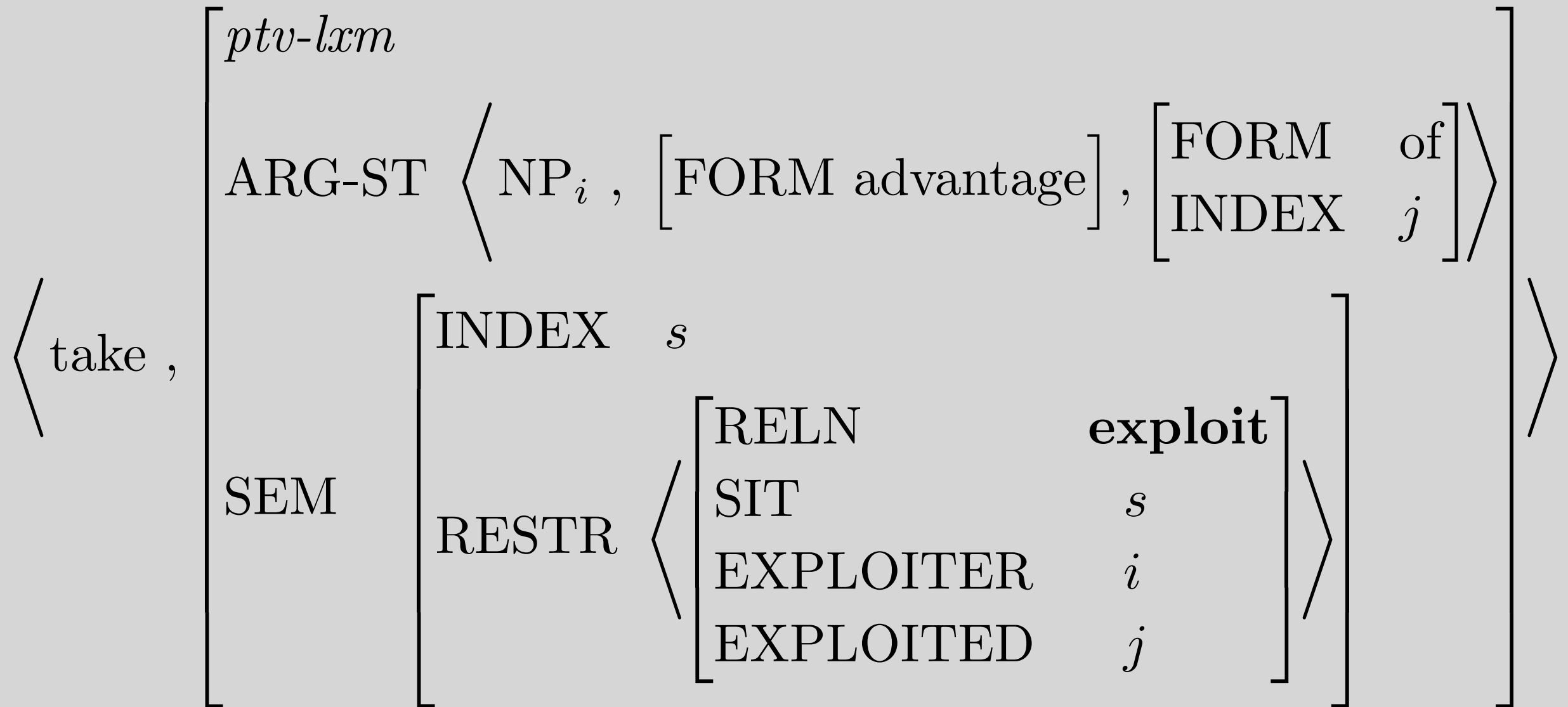
Sample tree with extraposition

```
                          S
                  _____/ _____
                 /               \
               NP                VP
                |          _____/ _____
               It        /               \
                        V                CP
                        |          _____/ _____
                     matters      /               \
                                 C                S
                                 |          _____/ \_____
                               that       /             \
                                         NP             VP
                                          |              |
                                         Kim          arrived
```

Your favorite (SFW) English idiom?

Another Nonreferential Noun

$$
\left\langle \text{advantage} \; , \;
\begin{bmatrix}
\textit{massn-lxm} \\[4pt]
\text{SYN} \quad
\begin{bmatrix}
\text{HEAD} \quad
\begin{bmatrix}
\text{FORM} & \text{advantage} \\
\text{AGR} & \textit{3sing}
\end{bmatrix}
\end{bmatrix} \\[12pt]
\text{SEM} \quad
\begin{bmatrix}
\text{MODE} & \text{none} \\
\text{INDEX} & \text{none} \\
\text{RESTR} & \langle \, \rangle
\end{bmatrix}
\end{bmatrix}
\right\rangle
$$
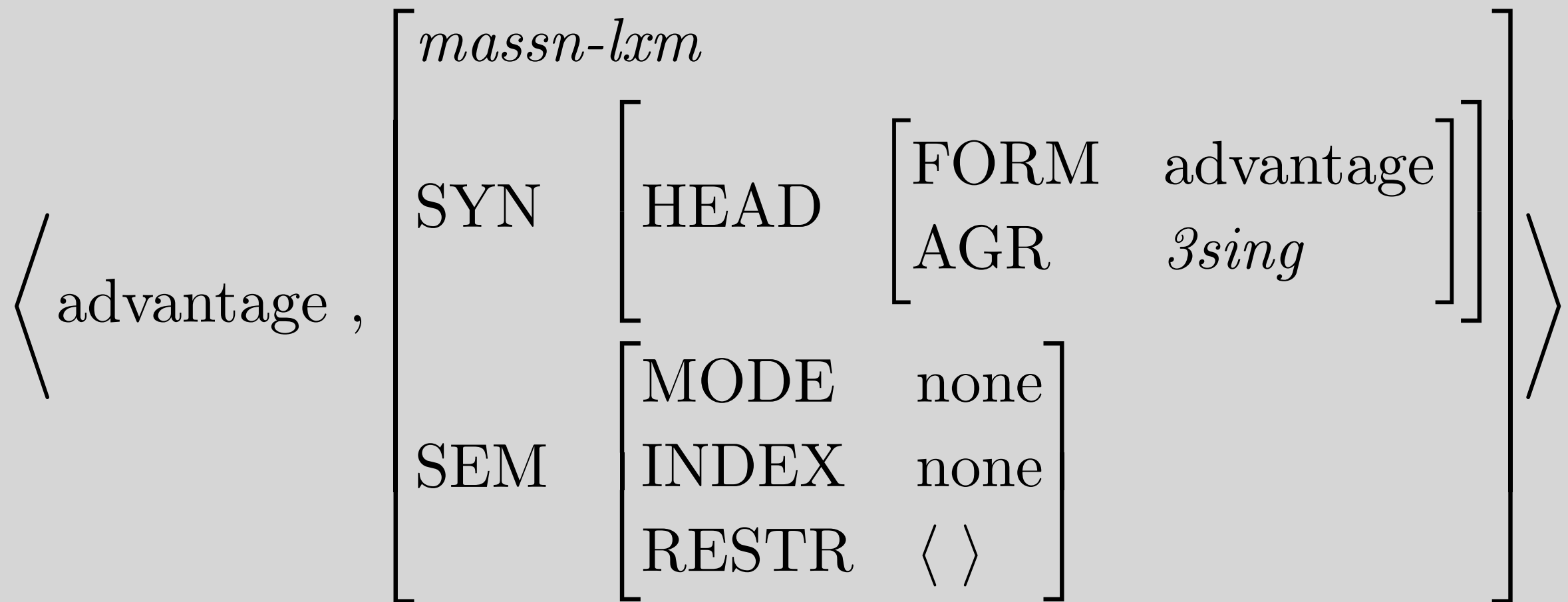
The Verb that Selects *advantage*

$$
\left\langle \text{take} , \begin{bmatrix} \textit{ptv-lxm} \\[2pt] \text{ARG-ST} \ \left\langle \text{NP}_i , \ \begin{bmatrix} \text{FORM advantage} \end{bmatrix} , \ \begin{bmatrix} \text{FORM} & \text{of} \\ \text{INDEX} & j \end{bmatrix} \right\rangle \\[10pt] \text{SEM} \ \begin{bmatrix} \text{INDEX} & s \\[4pt] \text{RESTR} \ \left\langle \begin{bmatrix} \text{RELN} & \textbf{exploit} \\ \text{SIT} & s \\ \text{EXPLOITER} & i \\ \text{EXPLOITED} & j \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle
$$

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

Reading Questions

- If be-lxm is a subtype of verb-lxm, where does exist-be-lxm fall on the lexeme hierarchy? Is it a subtype of be-lxm or does it stand alone?

Reading Questions

- INDEX of 'there' is set to be none is so smart!! It matches the meaninglessness and rules out 'there loved Sandy'. But can INDEX of placeholder 'it' be the same as CP? 'It' seems to bear the meaning of CP (or other constituents like infinitives). What problems are created if its INDEX is not none?

Reading Questions

- Why do some lexical rules specify for ARG-STR and not SPR and COMPS, while others do the opposite (specify for SPR and COMPS but not ARG-STR). Doesn't the ARP guarantee these values are the same? The ARP is a constraint on type word -- is it just that rules involving words don't need to specify for ARG-STR but all others do?

Reading Questions

- If we take the book's understanding of lexical rules seriously, i.e. they are not functions that "transform" an input into an output, but rather a kind of filter or sieve for finding paired tuples, is there any reason to not think of them as reversible? Is there a clear reason to select the output rather than the input? Which element of the tuple is selected seems an arbitrary, external decision. And if they are reversible, then the dative alternation rule, a passive rule, and perhaps others, needn't be thought of as "the passive rule" after all, it's reversible. We could equally call it the "active" rule. Is that right? This might just be a minor corollary to the idea that the grammar should be equally intelligible and operable "bottom up" or "top down."

Reading Questions

- How can you tell when a certain phenomena is a good candidate for a pi-rule?

Reading Questions

- Say there was a verb that could take all other complementizer clauses except for that-clauses. Would the lexical entry list all the acceptable complementizers or is there a succinct way to say in HPSG "all but X"?

Reading Questions

- The different values of FORM are becoming somewhat arbitrary to me. For example, nonreferential 'be' and 'there' are serving very similar semantic and syntactic roles, but 'be' is FORM be and 'there' is FORM there. Aren't we missing out on a linguistic generalization by getting so specific with FORM? I guess I'm also wondering if HPSG outside our snowglobe grammar fragment has a slightly different analysis? >>

Reading Questions

- I also wanted to ask about the different values of FORM. As of this point in the grammar that we've developed, FORM values of verbs seem to capture general phenomena (passive voice, participle form, etc.) that make more sense as categories, while others are either broad (like "nform" for nouns, "cform" for complementizers) or just certain words ("by" and "on" for prepositions, "it" and "there" for existentials). Is this uneven distribution of specificity simply due to what we have/haven't covered, or a property of the grammar?

Reading Questions

- Can we use the entire string of other phrasal verbs (fall apart, go out with, throw up) in a single lexical entry like the verb kick the bucket? Or would we use FORM like for take advantage? Why can't we have lexical entries for take advantage anyway? What's the benefit of using FORM?

Reading Questions

- I was just wondering with the increase in FORM values how do we interpret FORM when we see it in a tree? And how has the usage/definition of FORM changed from the previous chapters to this chapter?

Reading Questions

- What is the function of the NP on the SPR list of the PP? I know that in (8) the SPR allows it to fit the constraint for a be-lxm's ARG-ST but other than that, what is a SPR for a PP indicating? It could license a NP like "The book under the table", I suppose, but to me "under the table" is a modifier in this case and should combine with "The book" via the HMR not the HSR.

Reading Questions

- "You can make it in Hollywood." Would we treat "make it" as idiomatic expression meaning "succeed"?

- Is there any way to capture idioms like "keep tabs on" and "take advantage" more generally? It just feels very arbitrary to be able to say that all idioms have their own form.

Reading Questions

- Is there more than the ability to apply the passive rule motivating the decision to treat some idioms, like "keep tabs on" and "take advantage of", as a combination of verby lexical entries that take specific (semantically "empty") complements? (As opposed to treating the entire entity as a single verb, as we do with "kick the bucket".) If so, can you elaborate on some of the other motivations for representing idioms in these two different ways?

Reading Questions

- How do we deal with non-verby idioms? For example, "two cents" can be used with many different verbs to yield the same meaning. "I gave my two cents", "He put in his two cents", "They asked for your two cents", etc. With our current set up it seems like each of these verbs would have an idiomatic lexical entry with complements that have FORM two and FORM cents and semantics features with empty/none values, but the fact that this expression can occur with many different verbs makes it seem likely that the semantic information should be tied to the "two cents" part of the expression, not the verb.

Reading Questions

- What do we gain from not nesting predications when trying to notate semantic embedding? Is it simply a readability thing?

Reading Questions

- Chapter 11 somewhat reinforced my intuition that, despite the difficulties mentioned in class on Tuesday, being able to select for some kind of generic semantic value would be helpful. Hard-coding specific words values in FORM works in our syntax. But might it not miss some productive generalizations that this other approach might be able to capture?

Reading Questions

- Japanese has nominalizer suffixes no and koto which are appended to VPs and APs and allow us to treat them as nouns. These nominalizers have very similar behavior to the complementizer that, and I see that we have even placed the new type comp under a new IST nominal in the type hierarchy. What subtle differences are there between a nominalizer and a complementizer?