Ling 566
Oct 13, 2015

Valence, Agreement

Overview

- Review: pizza, feature structures, well-formed trees, HFP
- A problem with the Chapter 3 grammar
- Generalize COMPS and SPR
- The Valence Principle
- Agreement
- The SHAC
- Reading Questions

Pizza review

- Unification is an operation for combing constraints from different sources.
- What are those sources in the pizza example?
- Why do we need to combine information from different sources in our grammars?

Reminder: Where We Are

- Attempting to model English with CFG led to problems with the granularity of categories, e.g.
  - Need to distinguish various subtypes of verbs
  - Need to identify properties common to all verbs
- So we broke categories down into feature structures and began constructing a hierarchy of types of feature structures.
- This allows us to schematize rules and state cross-categorial generalizations, while still making fine distinctions.

A Tree is Well-Formed if …

- It and each subtree are licensed by a grammar rule or lexical entry
- All general principles (like the HFP) are satisfied.
- NB: Trees are part of our model of the language, so all their features have values (even though we will often be lazy and leave out the values irrelevant to our current point).

The Head Feature Principle

- Intuitive idea: Key properties of phrases are shared with their heads

$$\begin{bmatrix} \textit{phrase} & & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} & \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} & \end{bmatrix}$$

$$\begin{bmatrix} \textit{phrase} & & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} & \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} & \end{bmatrix} \qquad \begin{bmatrix} \textit{phrase} & & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} & \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} & \end{bmatrix}$$

$$\begin{bmatrix} \textit{word} & & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} & \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} & \end{bmatrix} \qquad \begin{bmatrix} \textit{word} & & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} & \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} & \end{bmatrix}$$

they swim

But it's still not quite right…

- There's still too much redundancy in the rules.
- The rules and features encode the same information in different ways.

Head-Complement Rule 1:

$$\begin{bmatrix} \textit{phrase} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textit{word} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

Head Complement Rule 2:

$$\begin{bmatrix} \textit{phrase} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textit{word} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \text{ NP}$$

Head Complement Rule 3:

$$\begin{bmatrix} \textit{phrase} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textit{word} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \text{dtr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \text{ NP NP}$$

Solution:
More Elaborate Valence Feature Values

- The rules just say that heads combine with whatever their lexical entries say they can (or must) combine with.
- The information about what a word can or must combine with is encoded in list-valued valence features.
  - The elements of the lists are themselves feature structures
  - The elements are "cancelled" off the lists once heads combine with their complements and specifiers.

Complements

Head-Complement Rule:

$$\begin{bmatrix} phrase \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} word \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \langle \boxed{1}, \ldots, \boxed{n} \rangle \end{bmatrix} \end{bmatrix} \boxed{1}, \ldots, \boxed{n}$$

- This allows for arbitrary numbers of complements, but only applies when there is at least one.
  - Heads in English probably never have more than 3 or 4 complements
  - This doesn't apply where Head-Complement Rule 1 would. (Why?)
- This covers lots of cases not covered by the old Head-Complement Rules 1-3. (Examples?)

Specifiers

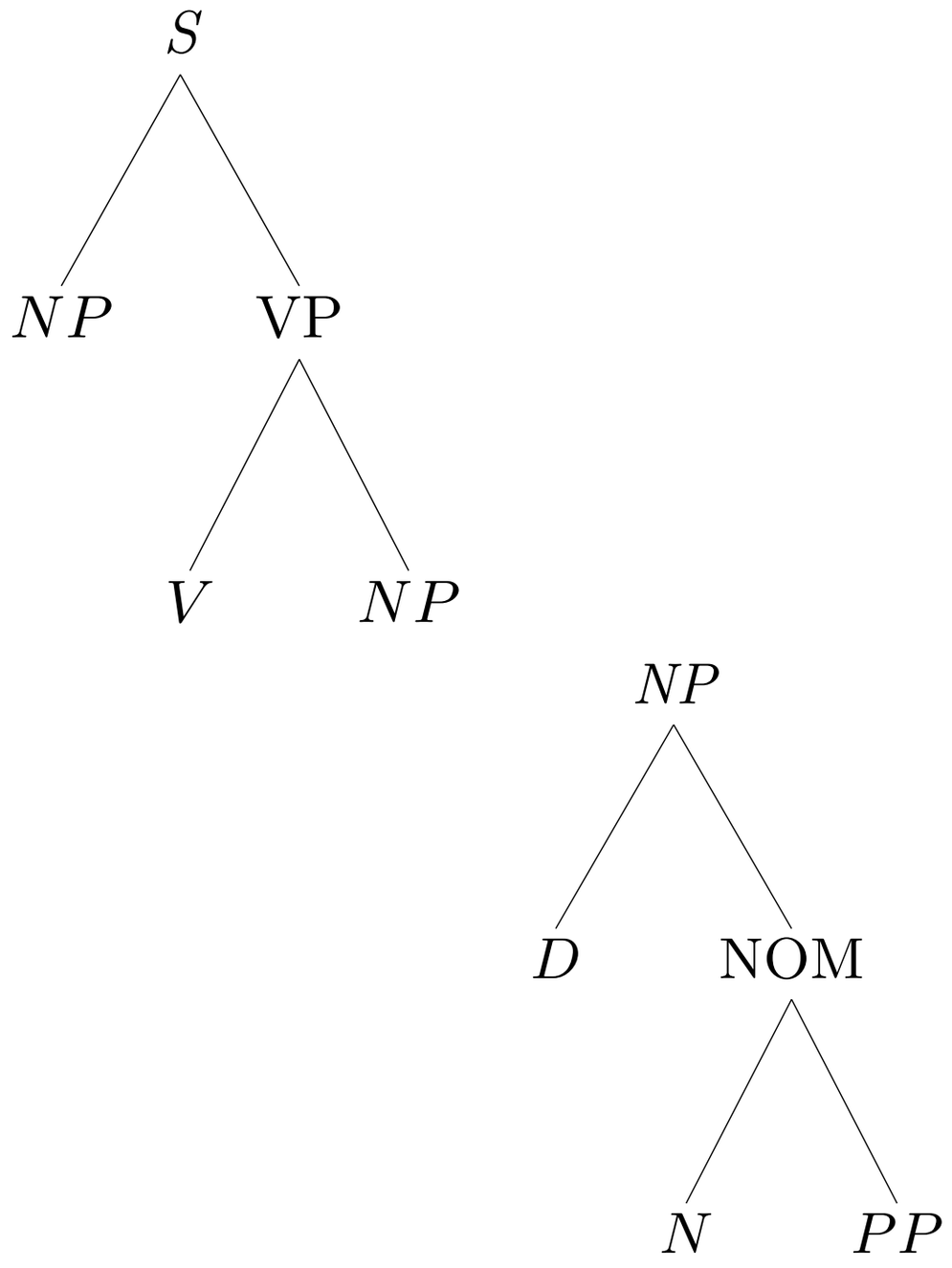
Head-Specifier Rule (Version I)

$$\begin{bmatrix} phrase \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \langle\ \rangle \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \rightarrow \boxed{2}\ \mathbf{H}\begin{bmatrix} \text{VAL} \quad \begin{bmatrix} \text{COMPS} & \langle\ \rangle \\ \text{SPR} & \langle\ \boxed{2}\ \rangle \end{bmatrix} \end{bmatrix}$$

- Combines the rules expanding S and NP.
- In principle also generalizes to other categories.
- Question: Why is SPR list-valued?

Question:

Why are these right-branching? That is, what formal property of our grammar forces the COMPS to be lower in the tree than the SPR?



Another Question…

What determines the VAL value of phrasal nodes?

ANSWER: The Valence Principle

Unless the rule says otherwise, the mother's values for the VAL features (SPR and COMPS) are identical to those of the head daughter.

More on the Valence Principle

- Intuitively, the VAL features list the contextual requirements that haven't yet been found.
- This way of thinking about it (like talk of "cancellation") is bottom-up and procedural.
- But formally, the Valence Principle (like the rest of our grammar) is just a well-formedness constraint on trees, without inherent directionality.

So far, we have:

- Replaced atomic-valued VAL features with list-valued ones.
- Generalized Head-Complement and Head-Specifier rules, to say that heads combine with whatever their lexical entries say they should combine with.
- Introduced the Valence Principle to "cancel" things off the COMPS and SPR lists.

The Parallelism between S and NP

- Motivation:
  - pairs like *Chris lectured about syntax* and *Chris's lecture about syntax*.
  - both S and NP exhibit agreement

    *The bird sings/*sing* vs. *The birds sing/*sings*

    *this/*these bird* vs. *these/*this birds*
- So we treat NP as the saturated category of type *noun* and S as the saturated category of type *verb*.

Question: Is there any other reason to treat V as the head of S?

- In standard English, sentences must have verbs. (How about non-standard English or other languages?)
- Verbs taking S complements can influence the form of the verb in the complement:

  *I insist/*recall (that) you be here on time*.
- Making V the head of S helps us state such restrictions formally

A possible formalization of the restriction on *insist*

$$\left[\begin{array}{ll}\text{HEAD} & \textit{verb} \\ \text{VAL} & \left[\begin{array}{ll}\text{SPR} & \left\langle \text{NP} \right\rangle \\ \text{COMPS} & \left\langle \left[\begin{array}{ll}\text{HEAD} & \left[\begin{array}{ll}\textit{verb} \\ \text{MOOD} & \textit{subjunctive}\end{array}\right] \\ \text{VAL} & \left[\begin{array}{ll}\text{COMPS} & \left\langle\ \right\rangle \\ \text{SPR} & \left\langle\ \right\rangle\end{array}\right]\end{array}\right] \right\rangle \backslash \left\langle \text{NP} \right\rangle\end{array}\right]\end{array}\right]$$

Note that this requires that the verb be the head of the complement. We don't have access to the features of the other constituents of the complement.

An Overlooked Topic: Complements *vs*. Modifiers

- Intuitive idea: Complements introduce essential participants in the situation denoted; modifiers refine the description.
- Generally accepted distinction, but disputes over individual cases.
- Linguists rely on heuristics to decide how to analyze questionable cases (usually PPs).

Heuristics for Complements vs. Modifiers

- Obligatory PPs are usually complements.
- Temporal & locative PPs are usually modifiers.
- An entailment test: If **X Ved (NP) PP** does not entail **X did something PP**, then the PP is a complement.

Examples

– *Pat relied on Chris* **does not entail** *Pat did something on Chris*
– *Pat put nuts in a cup* **does not entail** *Pat did something in a cup*
– *Pat slept until noon* **does entail** *Pat did something until noon*
– *Pat ate lunch at Bytes* **does entail** *Pat did something at Bytes*

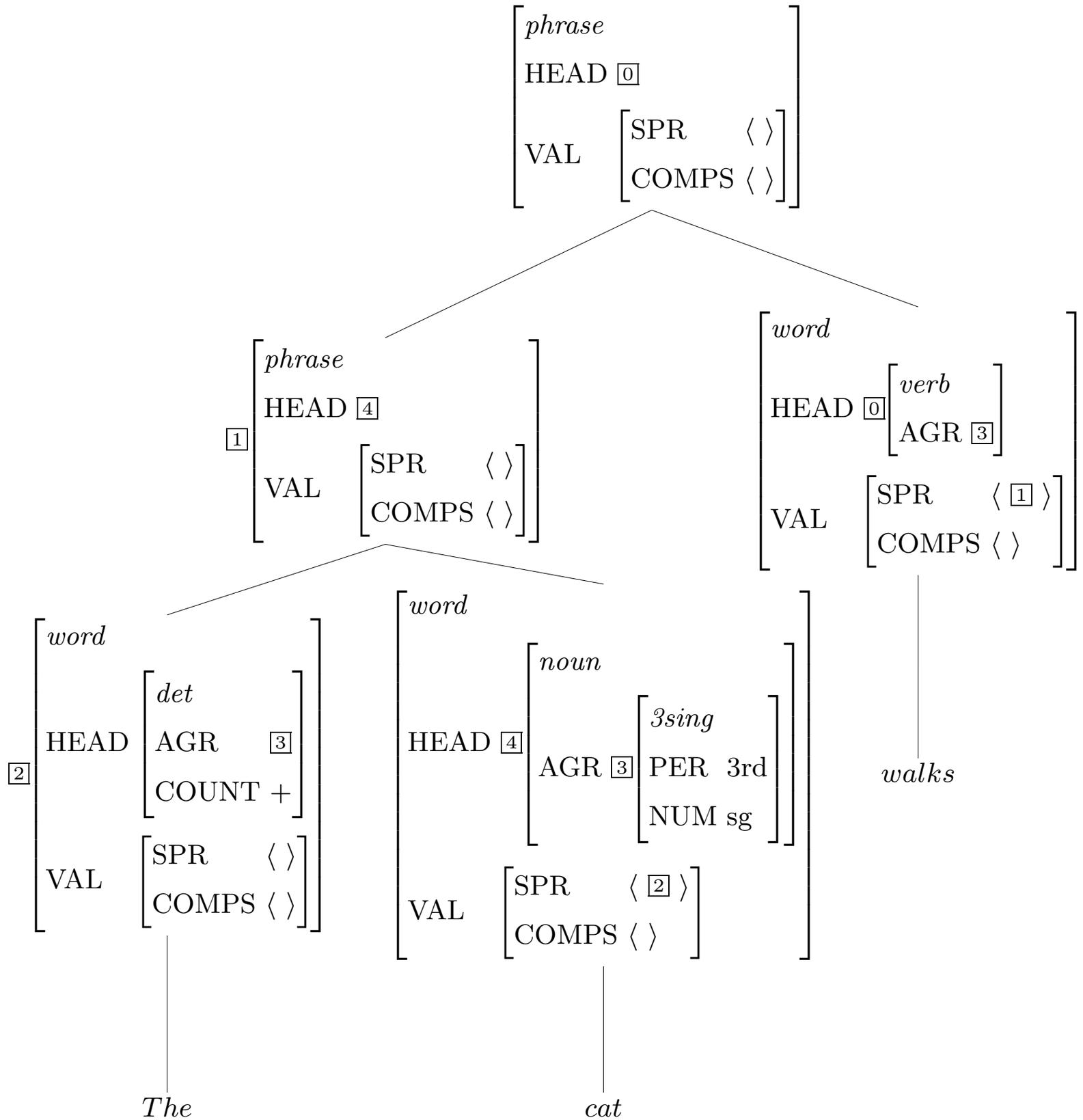
Agreement

- Two kinds so far (namely?)
- Both initially handled via stipulation in the Head-Specifier Rule
- But if we want to use this rule for categories that don't have the AGR feature (such as PPs and APs, in English), we can't build it into the rule.

The Specifier-Head Agreement Constraint (SHAC)

Verbs and nouns must be specified as:

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \left\langle \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$



The Count/Mass Distinction

- Partially semantically motivated
  - mass terms tend to refer to undifferentiated substances (*air, butter, courtesy, information*)
  - count nouns tend to refer to individuatable entities (*bird, cookie, insult, fact*)
- But there are exceptions:
  - *succotash* (mass) denotes a mix of corn & lima beans, so it's not undifferentiated.
  - *furniture, footwear, cutlery*, etc. refer to individuatable artifacts with mass terms
  - *cabbage* can be either count or mass, but many speakers get *lettuce* only as mass.
  - borderline case: *data*

Our Formalization of the Count/Mass Distinction

- Determiners are:
  - [COUNT –] (*much* and, in some dialects, *less*),
  - [COUNT +] (*a, six, many*, etc.), or
  - lexically underspecified (*the, all, some, no*, etc.)
- Nouns select appropriate determiners
  - "count nouns" say SPR <[COUNT +]>
  - "mass nouns" say SPR <[COUNT –]>
- Nouns themselves aren't marked for the feature COUNT
- So the SHAC plays no role in count/mass marking.

Overview

- Review: pizza, feature structures, well-formed trees, HFP
- A problem with the Chapter 3 grammar
- Generalize COMPS and SPR
- The Valence Principle
- Agreement
- The SHAC
- Reading Questions

Reading Questions

- The valence principle notes that the VAL features SPR and COMPS should match between mother and head daughter unless otherwise specified. Are the rules in (30) meant to illustrate this otherwise specification? What is meant by canceling appropriate elements in head daughter valence specification?

- "The effect of the Valence Principle is that: (1) the appropriate elements mentioned in particular rules are canceled from the relevant valence specifications of the head daughter in head-complement or head-specifier phrases". I don't understand what the "appropriate elements mentioned in particular rules" is here.

HCR & HSR

(30) a. Head-Specifier Rule (Near-Final Version)

$$\begin{bmatrix} \textit{phrase} \\ \text{VAL} \quad \begin{bmatrix} \text{SPR} \quad \langle\ \rangle \end{bmatrix} \end{bmatrix} \rightarrow \boxed{1} \quad \mathbf{H}\begin{bmatrix} \text{VAL} \quad \begin{bmatrix} \text{SPR} \quad \langle\ \boxed{1}\ \rangle \\ \text{COMPS} \quad \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

b. Head-Complement Rule (Final Version)

$$\begin{bmatrix} \textit{phrase} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} \quad \langle\ \rangle \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix} \textit{word} \\ \text{VAL} \quad \begin{bmatrix} \text{COMPS} \quad \langle\ \boxed{1}, \ldots, \boxed{n}\ \rangle \end{bmatrix} \end{bmatrix} \boxed{1} \ldots \boxed{n}$$

c. Head-Modifier Rule (Version II)

$$\begin{bmatrix} \textit{phrase} \end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix} \text{VAL} \quad \begin{bmatrix} \text{COMPS} \quad \langle\ \rangle \end{bmatrix} \end{bmatrix} \text{PP}$$

Reading Questions

- I thought the rule in (6) which specified a list of n COMPS was confusing. Wouldn't there be multiple possible rules, since you can specify different sets of feature structures for verbs that take different numbers of complements?

Reading Questions

- "Another thing to notice is that rules are written so that head-complement phrases are embedded within head-specifier phrases, and not vice versa. The key constraint here is the specification on the Head-Complement Rule that the head daughter must be of type word. Since the mother of the Head-Specifier Rule is of type phrase, a head-specifier phrase can never serve as the head daughter of a complement phrase." If we broke it down with our "simple terms", could we give an example and say VP-> V NP can be embedded in S->NP VP, but S -> NP VP could not be embedded in VP-> V NP in the place of the V solely because the Head-Specifier Rule mandates a word type and sentence is a phrase type?

Reading Questions

- "..the head-daughter of a head-specifier phrase need not be a word. For example, in the sentence Kim likes books, the head daughter of the head- specifier phrase will be the phrase likes books." When an entire phrase is the head-daughter, does that mean the head must agree with all the features contained within the head-daughter phrase (each word contained within the phrase)?

Reading Questions

- From what I understand, lists in COMPS and SPR are ordered, meaning they have to appear in the tree in the same order as they are in the specification. is there another way to indicate that order doesn't matter ? like a certain word that needs 2 or 3 types of complements in any given order.

Reading Questions

- In section 4.4, the line drawn between the two types of expressions—words and phrases—is blurred. The text says that "in some cases, two abbreviations may apply to the same node. For instance, the node above Alex in (26) may be abbreviated as either NP or N. Similarly, the node above opera may be abbreviated as either NOM or N." But how could Alex and opera be phrases? Aren't they sort of atomic in the sense that they are leaves? Also, the grammar rules seem to be pretty reliant on word/phrase distinctions to some extent or another. Is the irrelevance of the distinction simply contextual?

Reading Questions

- There were at least two paragraphs in Chapter 4 that described an expression as being "saturated". One is at the bottom of page 104 that describes an expression that is saturated as "has no unfulfilled valence features". In reading "unfulfilled" as "empty", but I'm not sure that is the correct reading. This same term is used on the bottom of p.116 in the phrase "non-saturated constituents" which defines "non-saturated" as "those with non-empty SPR or COMPS values". These two definitions did not seem to agree with one another ("non-saturated" should mean unfulfilled valence features, correct?).

Reading Questions

- (26) on page 104 gives a tree for the sentence "Alex likes the opera." Here, the word "opera" selects a specifier <[3]>, which is "the." If the sentence were "Alex likes opera" instead, presumably the SPR value of "opera" would be empty: <>. Would this mean that the grammar needs to have two different lexical entries for "opera," which differ in values for the SPR feature?

Reading Questions

- In the discussion of dine, devour, and eat, we see that even though all the words are semantically related, they display a variety of different valence structures. We seem to be able to separate the syntax from the semantics fairly easily here. Is this problematic in other languages? That is, are there cases where the semantics define the syntax? Are we able to keep these two areas distinct consistently?

Reading Questions

- Is PP [SPR <>] or not? In my opinion, PP should be [SPR -], because PPs are often used with verbs or adjective, like 'put it to the table' or 'is fond of this action'. Still, because sometimes PP can be placed in the beginning of a sentence like 'On the table, there is a toy car', I am not very sure how to fill the SPR property.

Reading Questions

- Why isn't COUNT part of the noun's AGR?
- Is COUNT -/+, or something similar, used for patterns like "fewer than X PLNOUN" and "less than X UNITS SGNOUN". Also, if there are cases where either many or much can be used, how would that be represented? Do you just drop the count, or would you explicitly specify it can be - or +.

Reading Questions

- How might this grammar handle incomplete lexical information? The feature COUNT, in particular, seems like it would only function with a lexicon that differentiates between mass and count nouns. What happens when the grammar can't easily identify the lexical head of a sentence or phrase?

Reading Questions

- Does a lexical approach to case (where the grammar rules don't mention it) work for languages with real case systems?

Reading Questions

- If N is the head of both subject-verb agreement and determiner-noun agreement then why is VP the head of S?
- We spent a lot of chapter 4 removing redundancy in our phrase structure rules, but I still don't feel completely satisfied with AGR. The value of AGR is of type agr-cat which, when fully specified, has both PER and NUM features. This makes sense for NP-VP agreement, but PER is redundant for D-N agreement, right? And what does it really mean for a determiner to be 3sing?

Reading Questions

- Our coordination rule allows for multiple NPs to be joined in a phrase. It seems that NPs in coordination phrases have interesting agreement properties.

*A teacher and some students drink.*drinks coffee*

- It appears that SHAC doesn't apply to multiple NPs in a coordination phrase. Why is agreement with a verb like drink handled in such a way for NPs in a coordination phrase?

Reading Questions

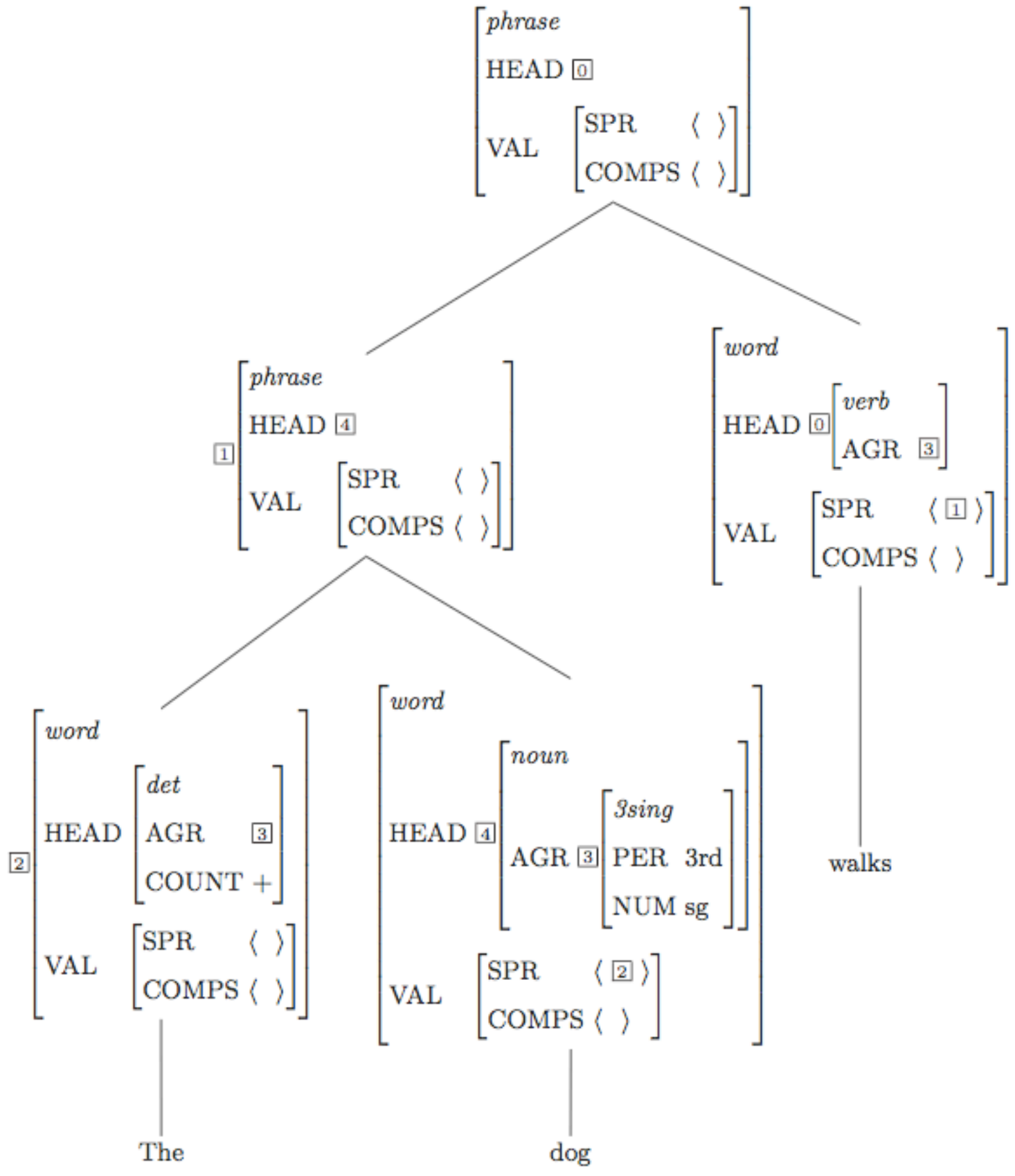
- Directionality: In (30a), the empty COMPS list is noted on the head daughter rather than the mother (where it's implied by the Valence Principle). Likewise in (52), the detailed is "pushed down" the tree. All of this seems to suggest a bottom-up approach that was dismissed earlier as metaphorical. Why is this preferred? Is it more than just a convention?

- How can properties, such as AGR, can travel through the tree hierarchy in the upper direction first, or the lateral direction, then cascade down? When the page 115 mentions, "AGR value of the verb(3) is identical to that of the NP it selects as a specifier (1)", I am assuming "NP" here is used just because the example illustrates NP, but it could be anything else? And the level that AGR can cascade down is not limited to just 1 level, is that correct assumption?

Reading Questions

- I'm having some trouble getting my head around agreement tags (when to tag, which tags correspond with which). More concretely, in the tree in (52) on p.115: The det and noun agree with tag 3 and is tagged with the same agreement, but why doesn't the noun phrase have an agreement mark? How does the agreement pass fromt the det and noun to the verb without going through the NP node?

(52)



Reading Questions

- The HCR says COMPS < > on the mother, but what about a sentence like "John saw the dog in the store", where the intended parsing is that John was in the store when he saw the dog, and the dog was not necessarily in the store. In this case, "in the store" serves as an adjunct rather than a complement, meaning that the prepositional phrase would have to be attached at a phrase-level above, instead of as a sister to the lexical head saw. Wouldn't that make this PP a sister to the VP that licenses saw the dog? And if so, that VP's COMP list would not be empty, but contain a reference to its sister PP? Is this legal under the Head-Complement rule?

- On page 98 it says; "there are certain kinds of PP that seem to be able to co-occur with almost any kind of verb, such as temporal or locative PPs". What are the fringe cases where a verb cannot occur with these kinds of PP?

Reading Questions

- The way we are formulating the new head complement rule got me wondering about how we would treat particle phrasal verbs like 'bring up' or 'take over', etc. On the surface, it seems like the only way to analyze a phrase like 'take over the country' is as a V with a COMPS <PP> value. Intuitively, though, this is wrong because it seems like the particle is more a part of the verb and helps to create its meaning, in the same way that 'blow' and 'blow up' mean very different things. At the same time, it is not truly part of the verb because it is a separate word and does not have to appear right next to the verb, as in 'take the country over'/'take it over'. How does the grammar treat constructions like these?

Reading Questions

- I'm curious how COMPS might work for a phrasal verb like pick up, structurally I would want the analysis to be different than the one given in (8) for relied e.g., we can say On Leslie, we relied or On the table, we put the flowers but not *up the kids, we picked. For some inseparable phrasal verbs (come across, back out of) analyzing it as a single lexical unit might work, but for the example above we'll need to somehow capture the fact that we can say I picked up the kids, I picked the kids up, I picked them up, but not *I picked up them. Do we have a way to do that yet?

Reading Questions

- In this chapter, I saw the use of the triangle symbol previously used for PP (seems to be shorthand for phrases rather than to continue to divide the phrase into syntactical segments) extended to NP phrases. This made me question whether when asked for the complete syntactical tree if I should also separate PP -> P + NP. When is this shorthand notation appropriate?

Reading Questions

- I have noticed so far that each time we study the drawbacks of a certain representation of the language we try to add more aspects to the grammar to cover the special cases that are not covered by the current representation. That is fine from a theoretical point of view, I was wondering what would be the effect of this addition of this much of complexity in real life applications of NLP? Would it still be possible to map all these theories and edge cases and make a computer program generalize them to further understand natural language?