Ling 566
Oct 6, 2015

Context-Free Grammar

Overview

- Formal definition of CFG
- Constituency, ambiguity, constituency tests
- Central claims of CFG
- Weaknesses of CFG
- Reading questions

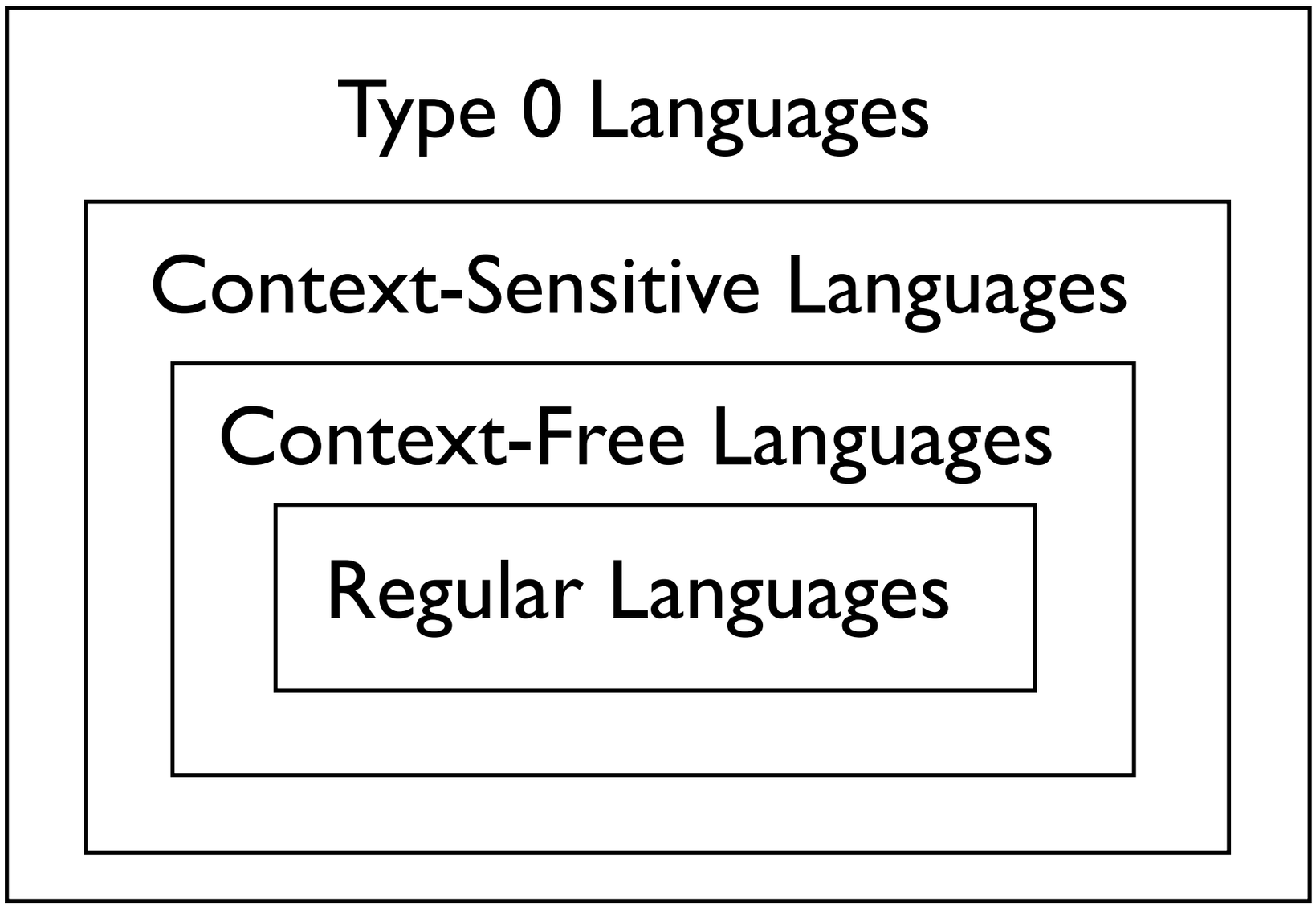
What does a theory do?

- Monolingual
  - Model grammaticality/acceptability
  - Model relationships between sentences (internal structure)
- Multilingual
  - Model relationships between languages
  - Capture generalizations about possible languages

Summary

- Grammars as lists of sentences:
  - Runs afoul of creativity of language
- Grammars as finite-state machines:
  - No representation of structural ambiguity
  - Misses generalizations about structure
  - (Not formally powerful enough)
- Next attempt: Context-free grammar (CFG)

Chomsky Hierarchy



Context-Free Grammar

- A quadruple: $< C, \Sigma, P, S >$
  - C: set of categories
  - Σ: set of terminals (vocabulary)
  - P: set of rewrite rules $\alpha \rightarrow \beta_1, \beta_2, \ldots, \beta_n$
  - S in C: start symbol
  - For each rule $\alpha \rightarrow \beta_1, \beta_2, \ldots, \beta_n \in P$
    $\alpha \in C;\ \beta_i \in C \cup \Sigma;\ 1 \leq i \leq n$

A Toy Grammar

RULES

S ⟶ NP VP

NP ⟶ (D) A* N PP*

VP ⟶ V (NP) (PP)

PP ⟶ P NP

LEXICON

D: the, some

A: big, brown, old

N: birds, fleas, dog, hunter, I

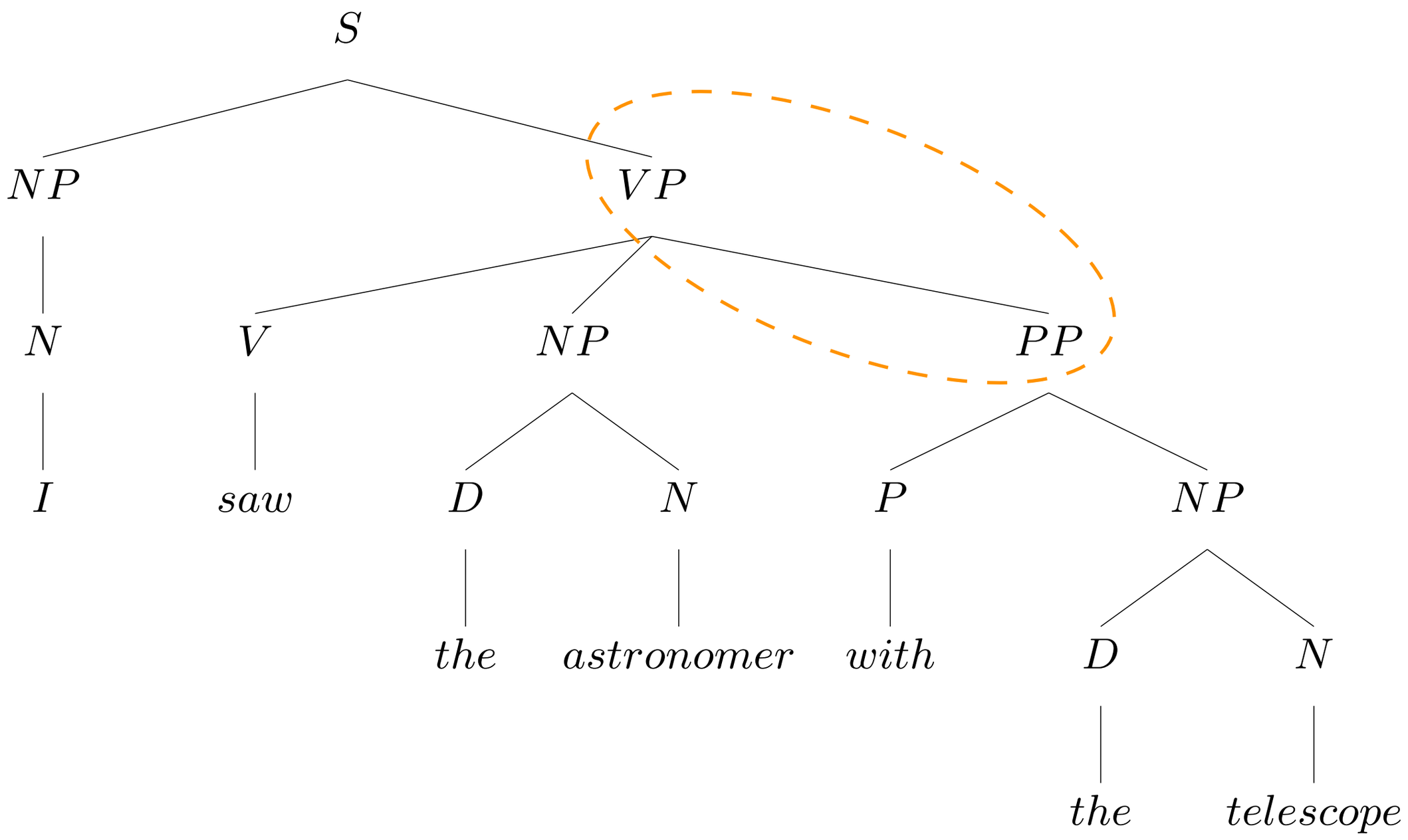
V: attack, ate, watched

P: for, beside, with

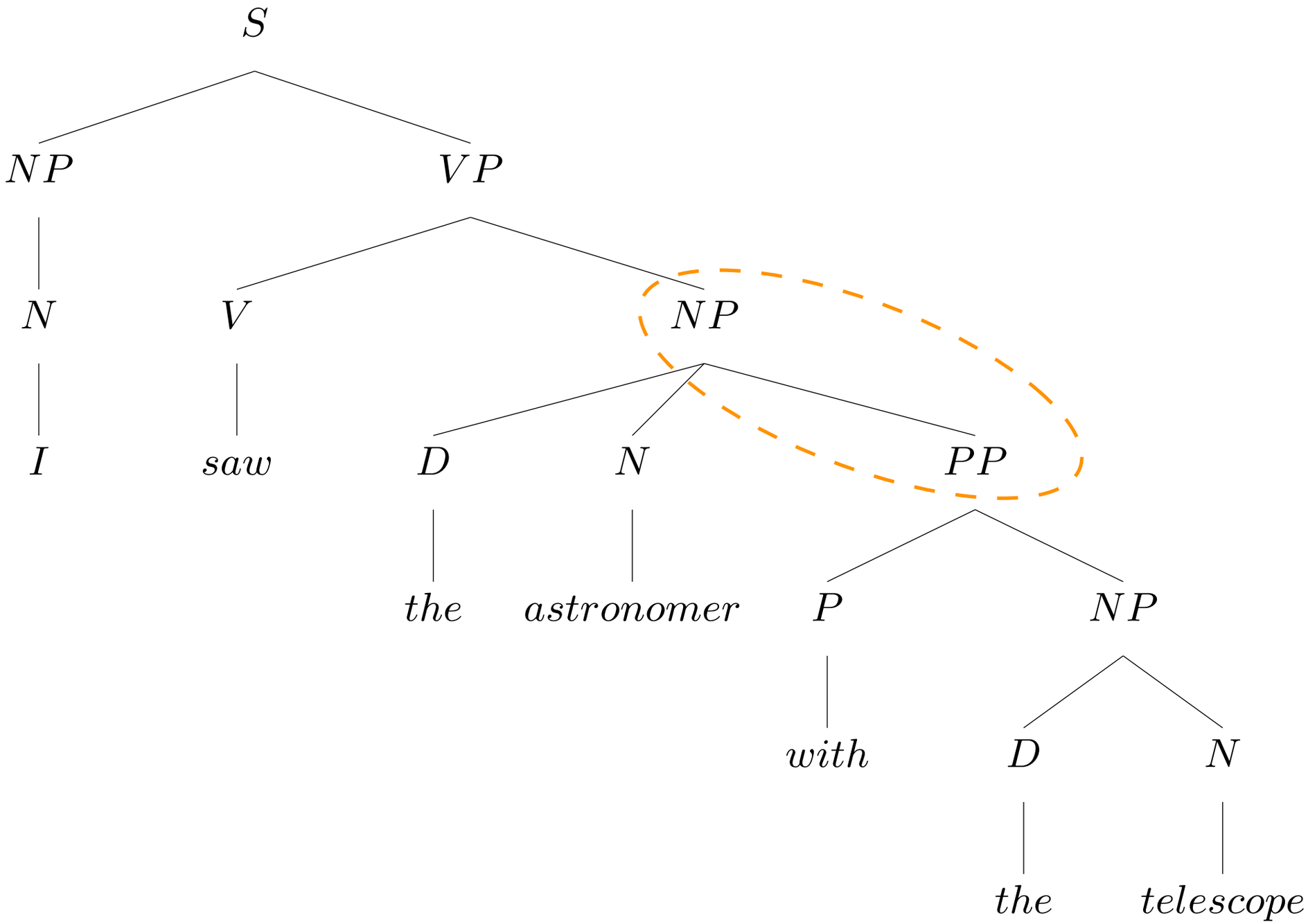
Structural Ambiguity

I saw the astronomer with the telescope.

Structure 1: PP under VP



Structure 1: PP under NP



Constituents

- How do constituents help us? (What's the point?)
- What aspect of the grammar determines which words will be modeled as a constituent?
- How do we tell which words to group together into a constituent?
- What does the model claim or predict by grouping words together into a constituent?

Constituency Tests

Constituency Tests

- Recurrent Patterns

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over *the lazy brown dog with one ear*.

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail *jumped over* *the lazy brown dog with one ear*.

- Coordination

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over *the lazy brown dog with one ear*.

- Coordination

  The quick brown fox with the bushy tail and *the lazy brown dog with one ear* are friends.

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over *the lazy brown dog with one ear*.

- Coordination

  The quick brown fox with the bushy tail and *the lazy brown dog with one ear* are friends.

- Sentence-initial position

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over *the lazy brown dog with one ear*.

- Coordination

  The quick brown fox with the bushy tail and *the lazy brown dog with one ear* are friends.

- Sentence-initial position

  The election of 2000, everyone will remember for a long time.

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over *the lazy brown dog with one ear*.

- Coordination

  The quick brown fox with the bushy tail and *the lazy brown dog with one ear* are friends.

- Sentence-initial position

  The election of 2000, everyone will remember for a long time.

- Cleft sentences

Constituency Tests

- Recurrent Patterns

  The quick brown fox with the bushy tail jumped over the lazy brown dog with one ear.

- Coordination

  The quick brown fox with the bushy tail and the lazy brown dog with one ear are friends.

- Sentence-initial position

  The election of 2000, everyone will remember for a long time.

- Cleft sentences

  It was a book about syntax they were reading.

General Types of Constituency Tests

General Types of Constituency Tests

- Distributional

General Types of Constituency Tests

- Distributional
- Intonational

General Types of Constituency Tests

- Distributional
- Intonational
- Semantic

General Types of Constituency Tests

- Distributional
- Intonational
- Semantic
- Psycholinguistic

General Types of Constituency Tests

- Distributional
- Intonational
- Semantic
- Psycholinguistic

... but they don't always agree.

Central claims implicit in CFG formalism:

1. Parts of sentences (larger than single words) are linguistically significant units, i.e. phrases play a role in determining meaning, pronunciation, and/or the acceptability of sentences.

2. Phrases are contiguous portions of a sentence (no discontinuous constituents).

3. Two phrases are either disjoint or one fully contains the other (no partially overlapping constituents).

4. What a phrase can consist of depends only on what kind of a phrase it is (that is, the label on its top node), not on what appears around it.

- Claims 1-3 characterize what is called ‘phrase structure grammar’

- Claim 4 (that the internal structure of a phrase depends only on what type of phrase it is, not on where it appears) is what makes it ‘context-free’.

- There is another kind of phrase structure grammar called ‘context-sensitive grammar’ (CSG) that gives up 4. That is, it allows the applicability of a grammar rule to depend on what is in the neighboring environment. So rules can have the form A→X, in the context of Y_Z.

Possible Counterexamples

- To Claim 2 (no discontinuous constituents):

  A technician ***arrived*** *who could solve the problem*.

- To Claim 3 (no overlapping constituents):

  I read what was written about me.

- To Claim 4 (context independence):

  - *He arrives this morning*.
  - **He arrive this morning*.
  - **They arrives this morning*.
  - *They arrive this morning*.

Weaknesses of CFG (atomic node labels)

- It doesn't tell us what constitutes a linguistically natural rule

$$\text{VP} \rightarrow \text{P NP}$$
$$\text{NP} \rightarrow \text{VP S}$$

- Rules get very cumbersome once we try to deal with things like agreement and transitivity.

- It has been argued that certain languages (notably Swiss German and Bambara) contain constructions that are provably beyond the descriptive capacity of CFG.

Agreement & Transitivity

S	→	NP-SG VP-SG	VP-SG	→	IV-SG
S	→	NP-PL VP-PL	VP-PL	→	IV-PL
NP-SG	→	(D) NOM-SG	VP-SG	→	TV-SG NP
NP-PL	→	(D) NOM-PL	VP-PL	→	TV-PL NP
NOM-SG	→	NOM-SG PP	VP-SG	→	DTV-SG NP NP
NOM-PL	→	NOM-PL PP	VP-PL	→	DTV-PL NP NP
NOM-SG	→	N-SG	VP-SG	→	CCV-SG S
NOM-PL	→	N-PL	VP-PL	→	CCV-PL S
NP	→	NP-SG	VP-SG	→	VP-SG PP
NP	→	NP-PL	VP-PL	→	VP-PL PP
	...			...	

Shieber 1985

- Swiss German example:

...	mer d'chind	em Hans	es huus	lönd	hälfe	aastriiche
...	we the children-ACC	Hans-DAT	the hous-ACC	let	help	paint
...	we let the children help Hans paint the house					

- Cross-serial dependency:
  - *let* governs case on *children*
  - *help* governs case on *Hans*
  - *paint* governs case on *house*

Shieber 1985

- Define a new language $f(\text{SG})$:

$$\begin{array}{rcl@{\qquad}rcl}
f(\text{d'chind}) & = & \text{a} & f(\text{Jan säit das mer}) & = & \text{w} \\
f(\text{em Hans}) & = & \text{b} & f(\text{es huus}) & = & \text{x} \\
f(\text{lönde}) & = & \text{c} & f(\text{aastriiche}) & = & \text{y} \\
f(\text{hälfe}) & = & \text{d} & f([\text{other}]) & = & \text{z}
\end{array}$$

- Let r be the regular language $wa^*b^*xc^*d^*y$
- $f(\text{SG}) \cap r = wa^m b^n x c^m d^n y$
- $wa^m b^n x c^m d^n y$ is not context free.
- But context free languages are closed under intersection.
- $\therefore$ $f(\text{SG})$ (and by extension Swiss German) must not be context free.

Strongly/weakly CF

- A language is *weakly* context-free if the set of strings in the language can be generated by a CFG.
- A language is *strongly* context-free if the CFG furthermore assigns the correct structures to the strings.
- Shieber's argument is that SW is not *weakly* context-free and *a fortiori* not *strongly* context-free.
- Bresnan et al (1983) had already argued that Dutch is *strongly* not context-free, but the argument was dependent on linguistic analyses.

On the other hand....

- It's a simple formalism that can generate infinite languages and assign linguistically plausible structures to them.
- Linguistic constructions that are beyond the descriptive power of CFG are rare.
- It's computationally tractable and techniques for processing CFGs are well understood.

So.....

- CFG has been the starting point for most types of generative grammar.
- The theory we develop in this course is an extension of CFG.

Reading Questions

- How could you tell if you need a rule like S -> S PP in addition to VP -> VP PP? How can you tell if particular PPs belong under S or VP?

- In CFG the sentence (S->NP VP) is considered the 'initial symbol' of words which is 'licensed' by the grammar. I was wondering how CFG understand utterances which are being used in common speech that do not form the initial S.

- In the example sentence "The brown dog watched the birds beside the hunter", my intuition tells me that the PP "beside the hunter" belongs in the NP instead of the VP. I can see how this would be difficult to do structurally though. Do formal grammars ever consider the PP in sentences of this form to be attached to the NP? And if so how do they model it?

Reading Questions

- **The defendant denied*: Maybe this is syntactically well-formed, but the problem arises because a rule of semantics is being violated.

- Transitivity, etc., are talked about in Chapter 2 as syntactic classes that license certain behavior (e.g. The teacher handed John the book vs. *The teacher handed John, etc.). However, we could argue that certain semantic classes of verbs appear in certain grammatical contexts as well (e.g. movement verbs with locative PP's; transactional verbs with indirect objects). (I understand that the locative prepositions are themselves semantic and that indirect objects are not syntactic constituents in the sense of phrase-level structures—perhaps that's where my answer lies.) This syntactic/semantic interplay makes me think of Beth Levin's work in VerbNet, etc. Why is there often this correlation between meaning and syntactic class?

Reading Questions

- In the reading, section 2.7.1, it states that headedness is a problem for CFGs, saying that it "cuts across many phrase types, suggesting that the rules are too fine-grained". What exactly is meant by this? CFG's certainly have a convention for labeling phrase structures that imply the head-type, ie again VPs contain a V head, NPs contain an N head, etc, and therefore shouldn't it be relatively easy to detect and disregard unnatural hypothetical phrase structures rules like the example given on page 36? Is the issue that natural languages are far more complex, and would require a far more extensive pairing system beyond the simple NP->N, VP->V, etc pairings?

Reading Questions

- In section 2.3, it's mentioned that the definition given for CFGs allows regular expressions on the right side of rules, which according to footnote #10, differs a little from the usual formal definition of CFGs, which only allow finite strings on the right side of rules. Is there any reason to represent CFGs in this way, other than just making the notation more convenient?

- "CFG thus provides us with a straightforward mechanism for expressing such ambiguities, whereas grammars that use only regular expressions don't" I don't understand why this is the case when the CFG has regular expressions in the rules.

Reading Questions

- Does the lack of a base case in VP -> VP PP type recursion lead to any problems?
- Through coordination and other tools of recursion a sentence can be infinitely extended, but the human brain has limitations. Information cannot be processed in an infinite utterance. A "well-formed" sentence would follow the stipulated rules, but these rules ignore biological and information processing requirements.
- Or on the other side, does infiniteness mean, that the language has to be flexible enough to allow extensions and changes so to adapt to new grammatical conventions or semantic expressions that may become a new norm?

Reading Questions

- The infinite nature of natural language has always been a blocking issue when building language models for ASR systems. Is it possible to have some kind of graphemes/ characters CFG, where the terminals are the characters? Is there any research about this kind of grammars where the building blocks are characters not words?

Reading Questions

- Language universals: Does this generally mean "common to all existing human languages" or "common to all hypothetical human languages"? If the latter, does it account for different hypothetical environments for hypothetical languages to have developed (like what if language was developed by people who had technology to efficiently display videos of their thoughts?), and does it assume no additional variation in the biological basis of cognition?

- The book makes the interesting observation that a grammar of English should "abstract across different speakers, too." (p.43) That strikes me as an odd thought. We're told that "a grammar is a theory about the mental representation of linguistic knowledge." If that's the case, I suppose I don't see why abstracting for different speakers is an issue. Wouldn't a comprehensive enough grammar of a certain language characterize structure for each speaker of that language? Does this boil down to prescriptive vs. descriptive rules?

Reading Questions

- The claim that "A syntactic theory that sheds light on human linguistic abilities ought to explain why such patterns do not occur in human languages," seems fairly bold. If we are not positing something like Universal Grammar, or making the claim that our formalisms somehow mimic underlying mental processes, how would we even begin to posit an underlying "why"? I'm curious, if the claim above means a syntactic theory ought to be predicting the patterns we "should" see and the patterns we "shouldn't" see, (and again, if so on what basis) or whether it just means that the theory should present a best-fit formalism for that which is "attested" and "unattested" in natural language, and from their posit what we might be "likely" or "unlikely" to find.

Reading Questions

- Since natural languages aren't fully CFLs, can they be accurately represented with a combination of CFG generative rules and transformational rules?
- How can we learn more about transformational analyses?
- What are some other examples of transformation interactions?
- Why are we doing HPSG and not transformational grammar?

Reading Questions

- I understand that HPSG deals differently with the phenomena that transformational grammar attempts to explain, but I wondered what the reasoning is behind using this theory over something like MP for NLP/computational linguistics applications. Is it simply that HPSG lends itself better to computational applications? Is there currently any work being done in NLP/computational linguistics using more transformational approaches?

Reading Questions

- Sections 2.8 and 2.9 mention the role grammar, especially transformational grammar, has played in psychological study. I see the merit in using language structure to assist in researching mental processes, but this seems overly simplistic to me—to take what seems to be a somewhat artificial process and apply blindly to the human mind. (Based on some quick research, it looks like the transformational grammar informs language processing concept hasn't panned out.) How do psycholinguists currently use grammar to help in describing performance and competence processing? To what extent does it inform psycholinguistic hypotheses?

Reading Questions

- I wonder what risks, if any, these approaches have in terms of achieving the goal of elegance and enlightenment at the expense of an accurate understanding of neurological or sociological realism about how people actually deploy language skills. Might the "rules" deployed by our brains be as complex as those of 2.7.3, despite our conscious ability to simplify them through techniques found in transformational grammars or HPSG, or does the relative ease with which our brains seem to process language suggest the prudence of embracing parsimony as far as possible? How have the more successful grammars faired when combined with the findings of neuro/sociolinguists?

Overview

- Formal definition of CFG
- Constituency, ambiguity, constituency tests
- Central claims of CFG
- Weaknesses of CFG
- Next time: Feature structures