



Ling 566

Oct 3, 2017

Context-Free Grammar

Overview

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

Insufficient Theory #1

- A grammar is simply a list of sentences.
- What's wrong with this?

Insufficient Theory #2: FSMs

- the noisy dogs left

D A N V

- the noisy dogs chased the innocent cats

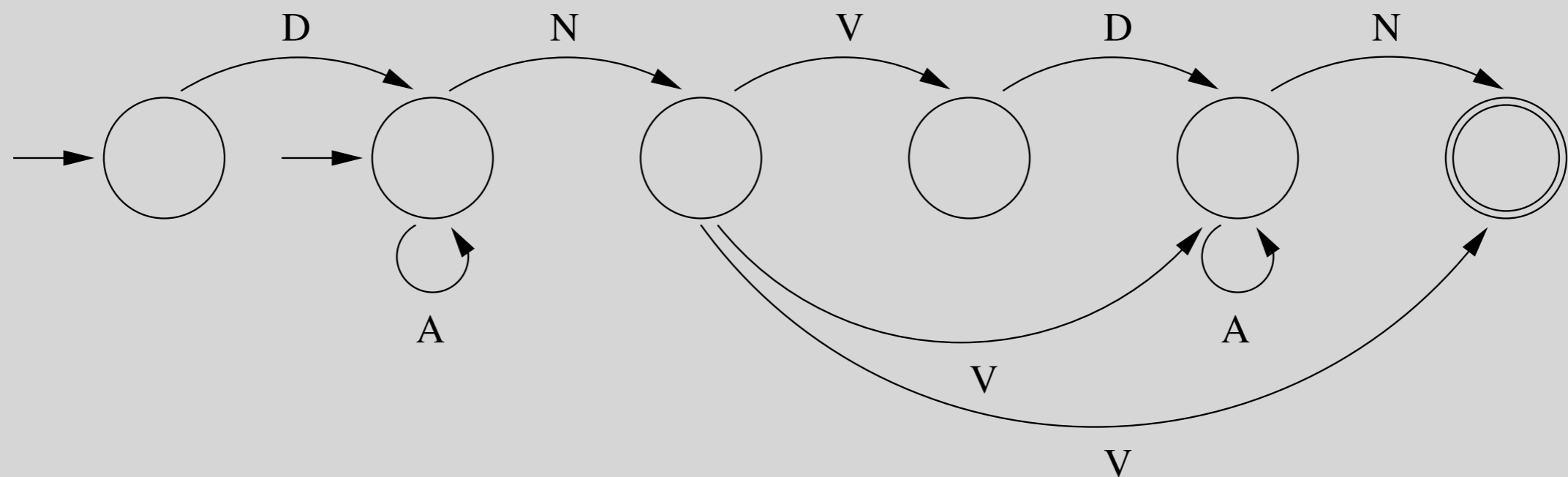
D A N V D A N

- $a^* = \{\emptyset, a, aa, aaa, aaaa, \dots\}$

- $a^+ = \{a, aa, aaa, aaaa, \dots\}$

- $(D) A^* N V ((D) A^* N)$

A Finite State Machine



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

Chomsky Hierarchy

Type 0 Languages

Context-Sensitive Languages

Context-Free Languages

Regular Languages

Context-Free Grammar

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

A Toy Grammar

RULES

$S \longrightarrow NP VP$

$NP \longrightarrow (D) A^* N PP^*$

$VP \longrightarrow V (NP) (PP)$

$PP \longrightarrow P NP$

LEXICON

D: the, some

A: big, brown, old

N: birds, fleas, dog, hunter, I

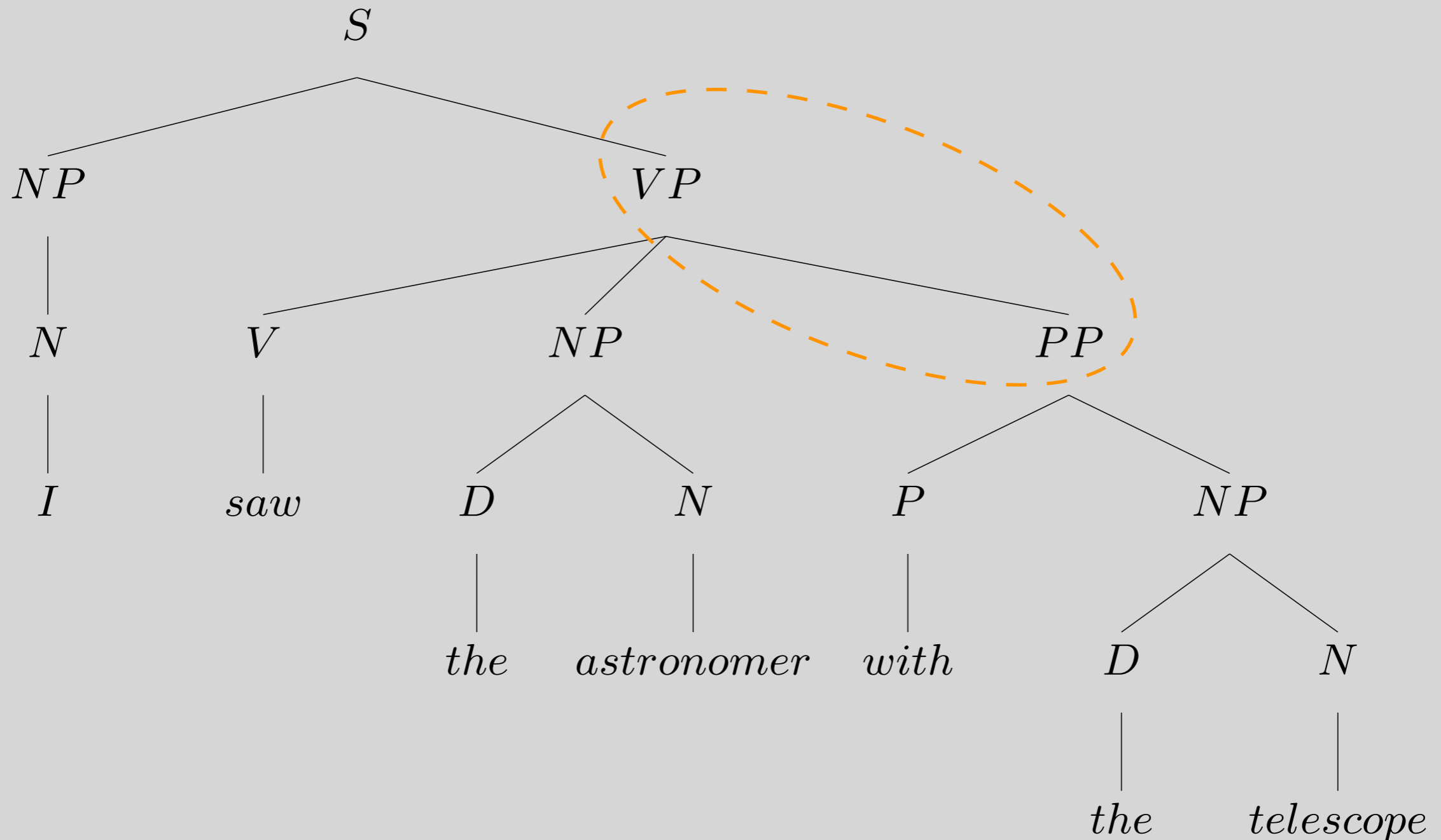
V: attack, ate, watched

P: for, beside, with

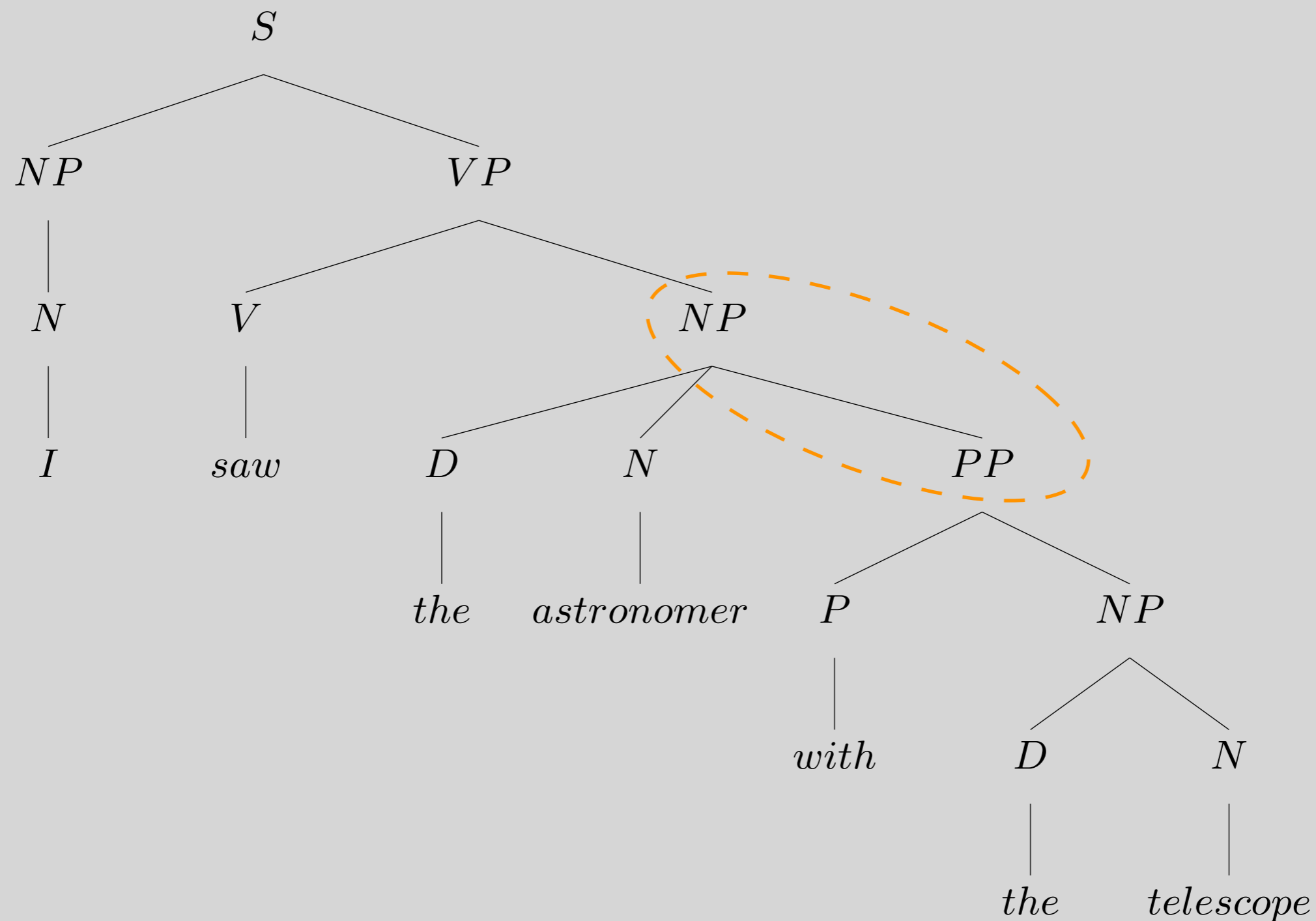
Structural Ambiguity

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

- 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

- 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 $A \rightarrow X$, 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.*

A Trivial CFG

$S \rightarrow NP VP$

$NP \rightarrow D N$

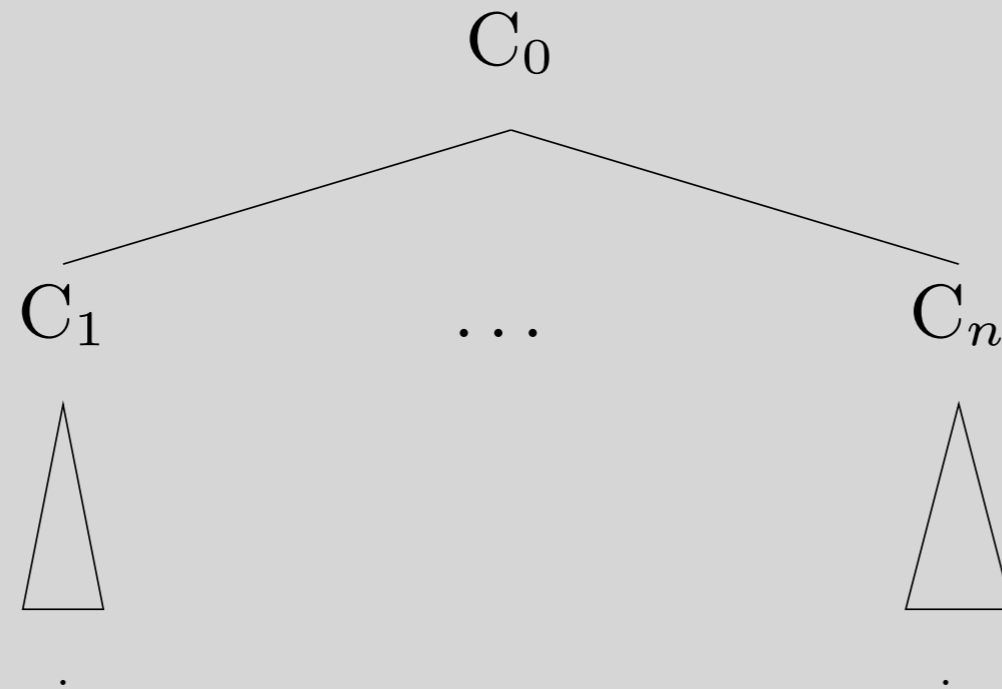
$VP \rightarrow V NP$

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



is a well-formed nonlexical tree if (and only if)

C_1 , ... , C_n are well-formed trees, and



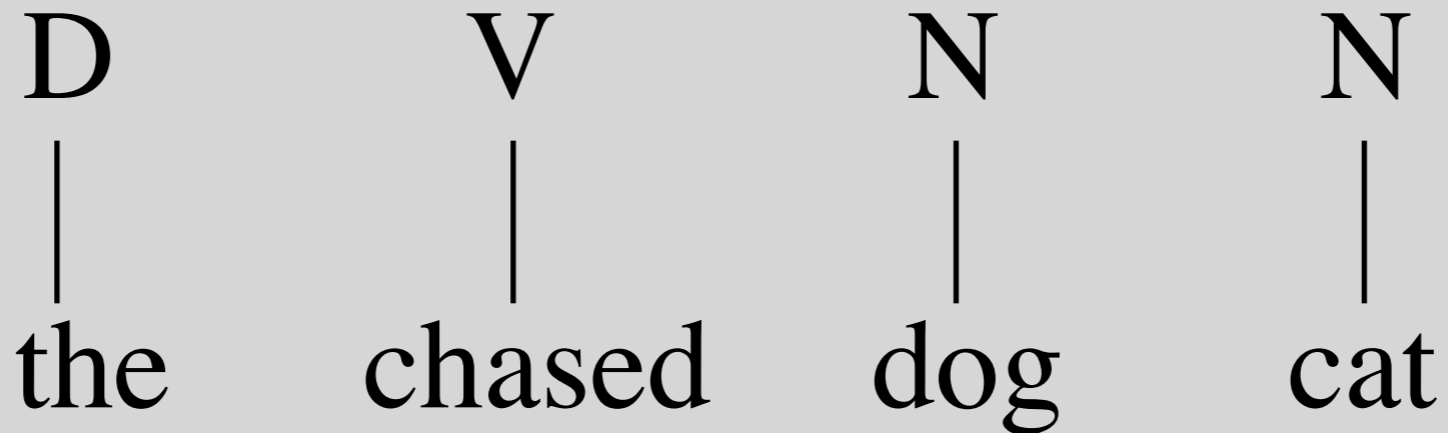
$C_0 \rightarrow C_1 \dots C_n$ is a grammar rule.

Bottom-up Tree Construction

D: *the*

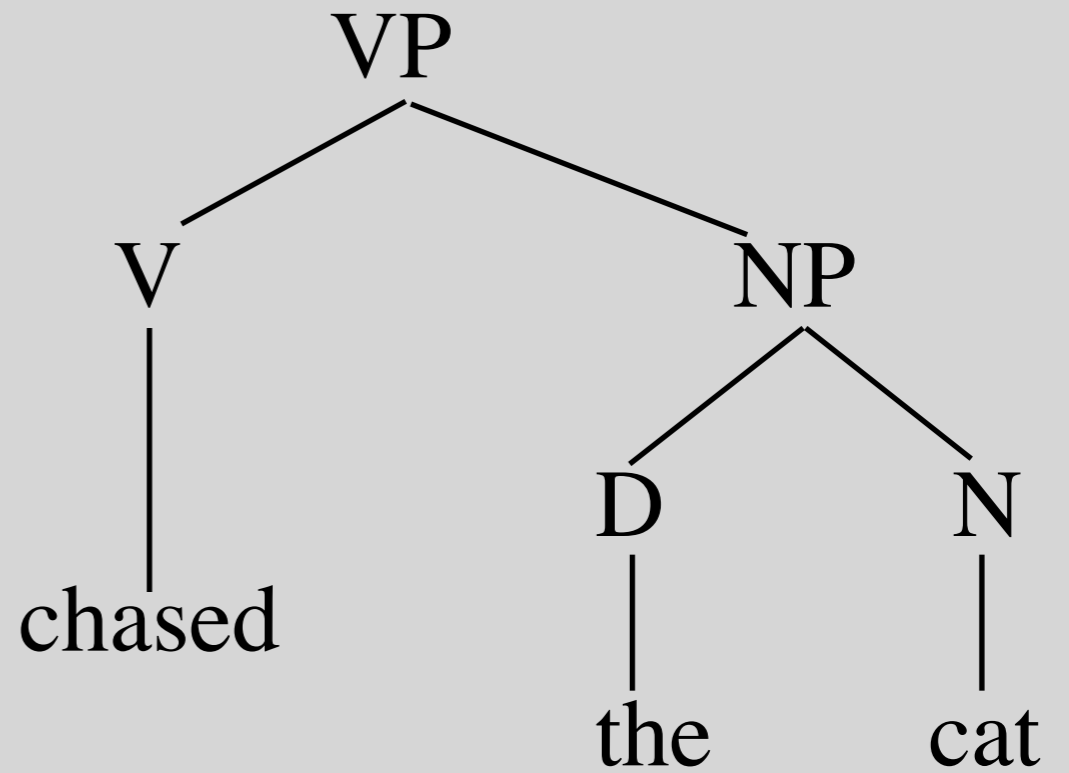
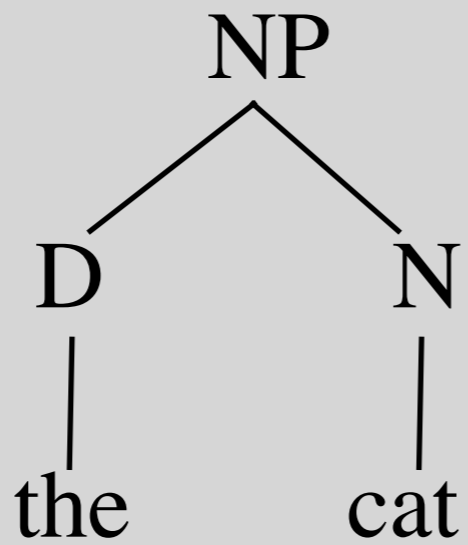
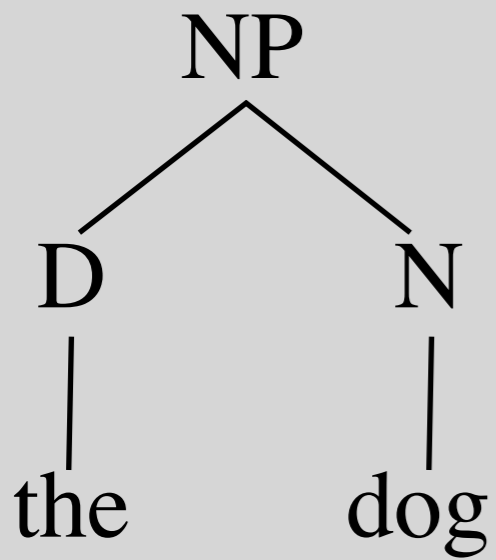
V: *chased*

N: *dog, cat*

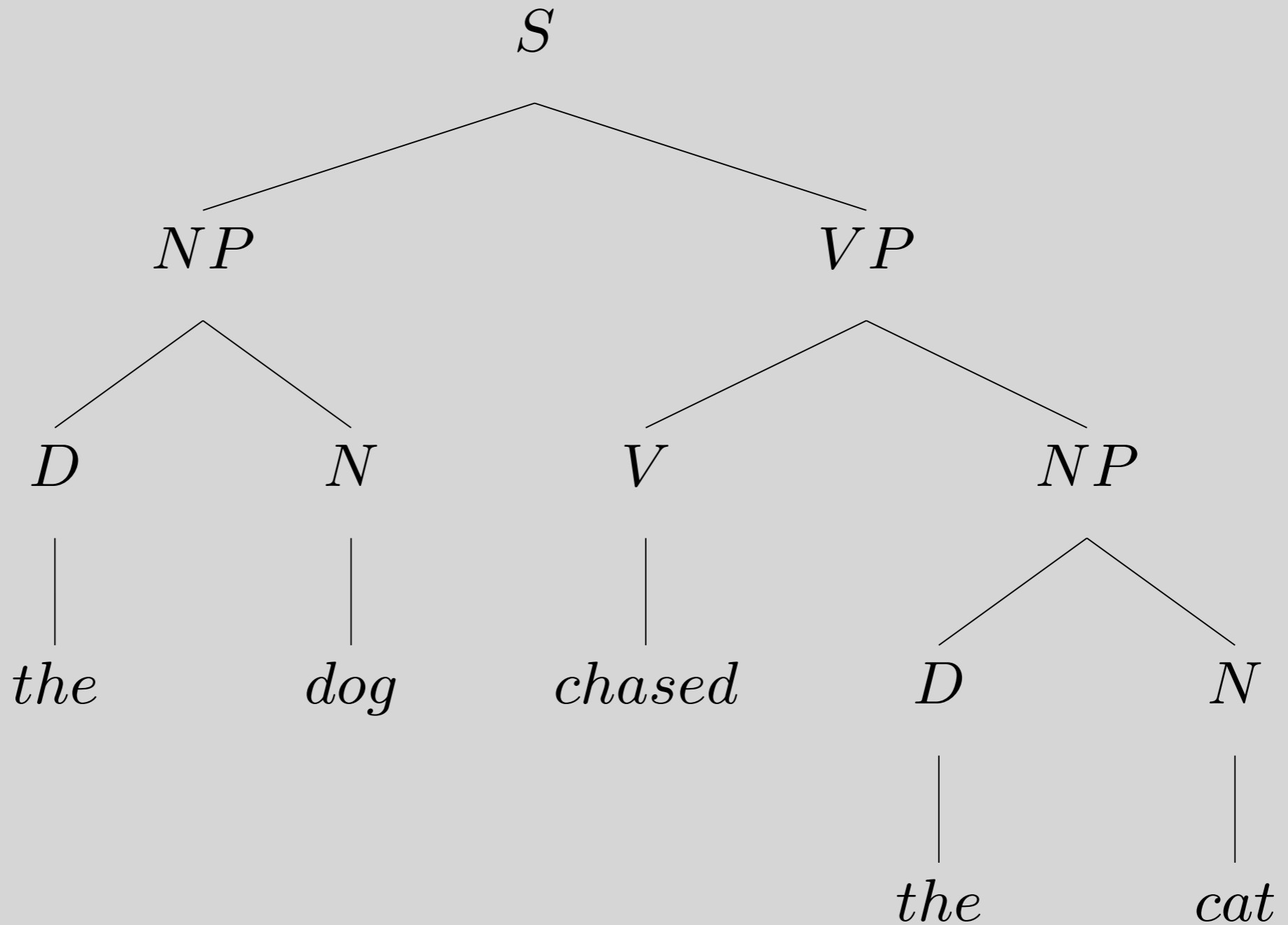


NP \rightarrow D N

VP \rightarrow V NP

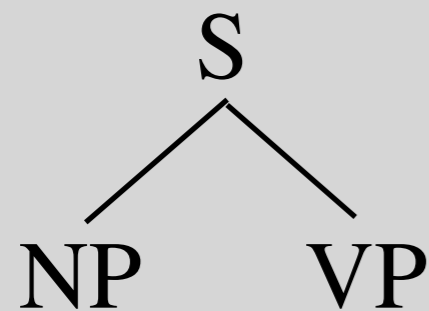


$S \rightarrow NP VP$

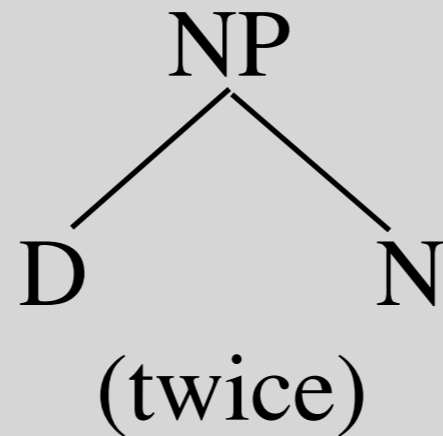


Top-down Tree Construction

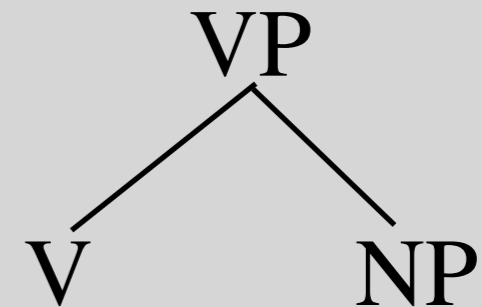
$S \longrightarrow NP VP$

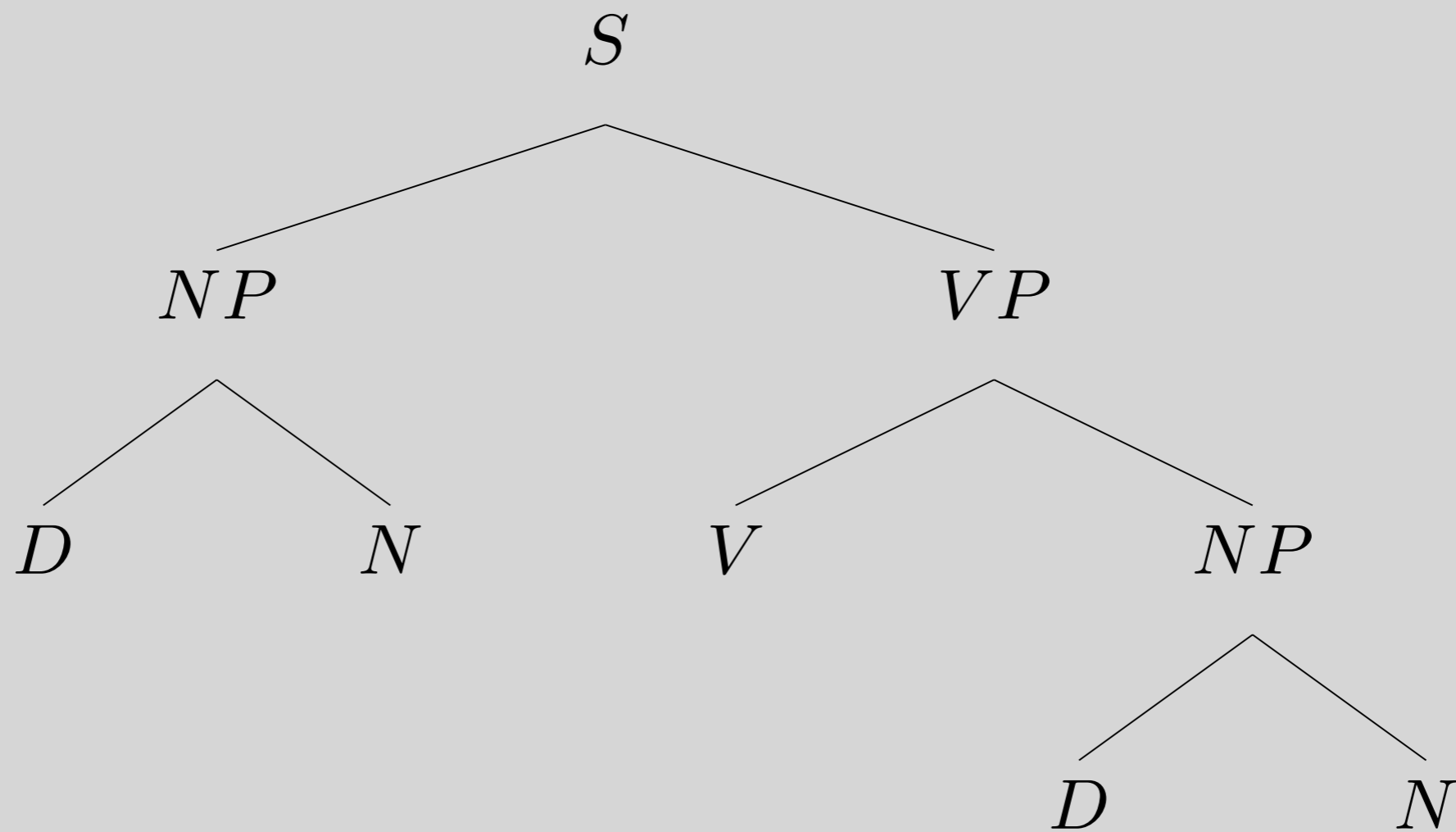


$NP \longrightarrow D N$



$VP \longrightarrow V NP$



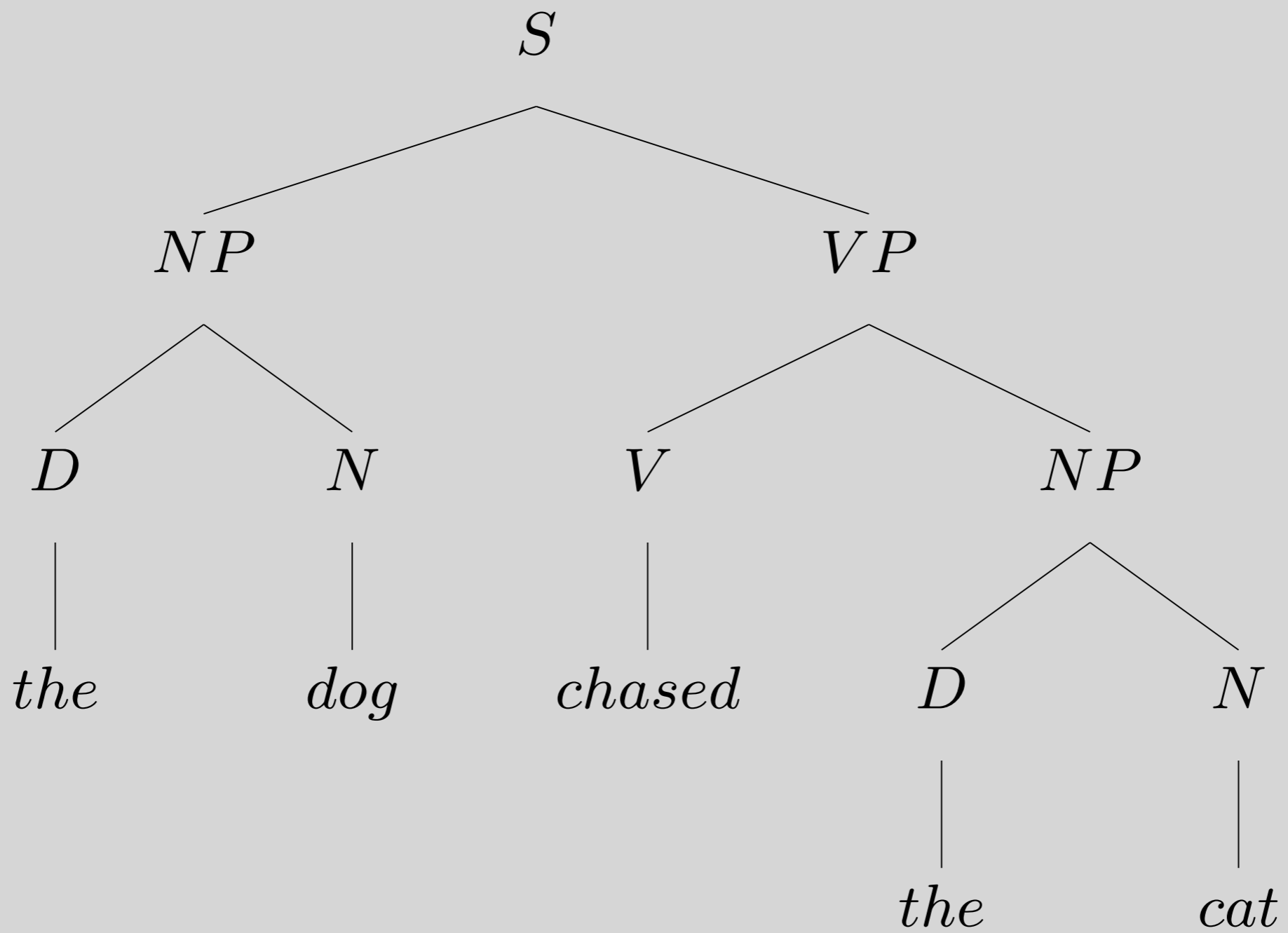


D
|
the

V
|
chased

N
|
dog

N
|
cat



Weaknesses of CFG (w/atomic node labels)

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

$$VP \rightarrow P NP$$
$$NP \rightarrow VP S$$

- 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 $f(\text{SG})$:

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

- Let r be the regular language $wa^*b^*xc^*d^*y$
- $f(\text{SG}) \cap r = wa^mb^nc^md^ny$
- $wa^mb^nc^md^ny$ 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

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.

Overview

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

Reading Questions

- The chapter stated that 'ambiguous sentences are often one which have multiple possible valid divisions of constituents'. This is a very neat and tidy way to think of ambiguity, and it made me wonder if ambiguity-ratings had ever been used in NLP, perhaps in a sort of "if ambiguity rating exceeds threshold, look farther than usual for additional context" or something along those lines.
- When we apply the CFG to build a tree for a sentence, are we supposed to build all the possible tree structure based on the CFG rules rather than using our own intuition to build the only "correct" tree?

Reading Questions

- ‘A lexical structure is well-formed if a particular word is listed under its corresponding grammatical category’. Via the concept of well-formedness, one can deduce the well-formedness of non-lexical trees by the Theorem given in this page.
 - 1. Is there a rigorous proof of this Theorem? Or perhaps a clarified version of definition of well-formedness?
 - 2. '*V--like*' is given to exemplify well-formedness of a lexical structure. Suppose there exists another lexical structure '*P--like*', using one of the other connotations of the word *like* in a different grammatical category, within the same S as the former structure is. Will well-formedness be defied?

Reading Questions

- What "context" is "context-free grammar" free of?
- Why is headedness a problem for CFG?
- On page 44, one of the suggested further readings is a work by Chomsky “arguing against the use of context-free grammars”. I am a bit confused about how Chomsky’s approach differs from a CFG and I was wondering if we could break down the differences between a CFG and the Chomskyian proposal.

Reading Questions

- Although the rules in (36), (39), (40) and (41) are redundant for human beings, they are not a problem for computer. So, does CFG play a more important role in computational linguistics?
- I was wondering why it was that the CFG treats sentence parsing and sentence generation equally. How are the "top-down" and "bottom-up" processes equally efficient?
- parsing vs generating: What drives generation?

Reading Questions

- What's the point of NOM?
- I am also curious about the introduction of NOM on page 31. What is the process for coming up with such nonlexical categories and their corresponding rules?
- Why aren't we using the more general X' notation?
- Why is $VP \rightarrow VP PP$ better than a version with PP^* ?
- Why have an explicit CONJ node in the tree on p. 21 ?

Reading Questions

- Are some languages more head-driven than others?
- English is order-sensitive in regards to the subject and object of the sentences, so is it inherent in the CFG? Or regarding other languages that are not order-sensitive, how is the CFG applied to them?
- After reading about CFGs and generative grammars in p.37 I am wondering about if there has been attempts to make cross-language grammars and how do they look like.

Reading Questions

- Is there an effort to explicitly define Chomsky's Universal Grammar in anyway or is UG simply an abstract/loosely defined principle? Is it not a useful problem since the Universal Grammar itself isn't a natural language and might not be constrained to the same rules of the languages we seek to describe?

Reading Questions

- If I have a lexicon and two context free grammars in hand, what is the criteria based on which I should decide which of the two is better ?
- Am I correct if I assumed that better CFG: (1) should be less ambiguous, (2) should provide better modeling of grammatically correct sentences of the language in question, and (3) should not accept grammatically incorrect sentences ?
- Any other criteria to be considered ?

Reading Questions

- The statement on P.40 that "there are verbs that only appear in other environments; for example, some verbs require following PPs or Ss" makes me wonder whether it is possible to generalize thorough rules to represent natural language. It appears to me that there are too many possible combinations of constituents to generalize. There are many cases in which certain verbs with similar meanings can not be interchanged. Second language learners as myself often come up with grammatically correct expressions that sounds weird to native speakers. And native speakers often find it difficult to explain the reason. I think this is because language is often used by chunk, and I think it is far more complicated than what CFG or Transformational Grammar can generalize.

Reading Questions

- p.47: The textbook states that subject-verb agreement is handled by assuming that number is an "intrinsic property of nouns, but not of verbs." What is the motivation for nouns having this intrinsic property instead of verbs? Is it just arbitrary?

Reading Questions

- In Section 2.4.1, the text mentions the below rule:

$$X \rightarrow X + \text{CONJ} X$$

- The interpretation for this is that elements of a category can be conjoined in the same way. What can we conclude about the conjunction of elements that do not belong to the same category? Is it correct to conclude that conjunction of elements of different categories is ungrammatical?
- "Coordinate conjunction is used as a test for constituency." I would like to discuss this in class so that I can comprehend it better. Specifically, how does this test work? Is this a sufficient and complete condition for constituency?

Reading Questions

- Are there theories of grammar that revolve around chains of sentences that do not sound correct together? As a general notion, we understand that there are well-formed paragraphs in a similar way that we have well-formed sentences. Are there methods for defining what makes a paragraph sound "grammatically correct"?
- Are commas and punctuation ever considered as being part of a grammar, are they not also determining whether or not a sequence of strings is grammatically correct?