Ling 566
Jan 9, 2019

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

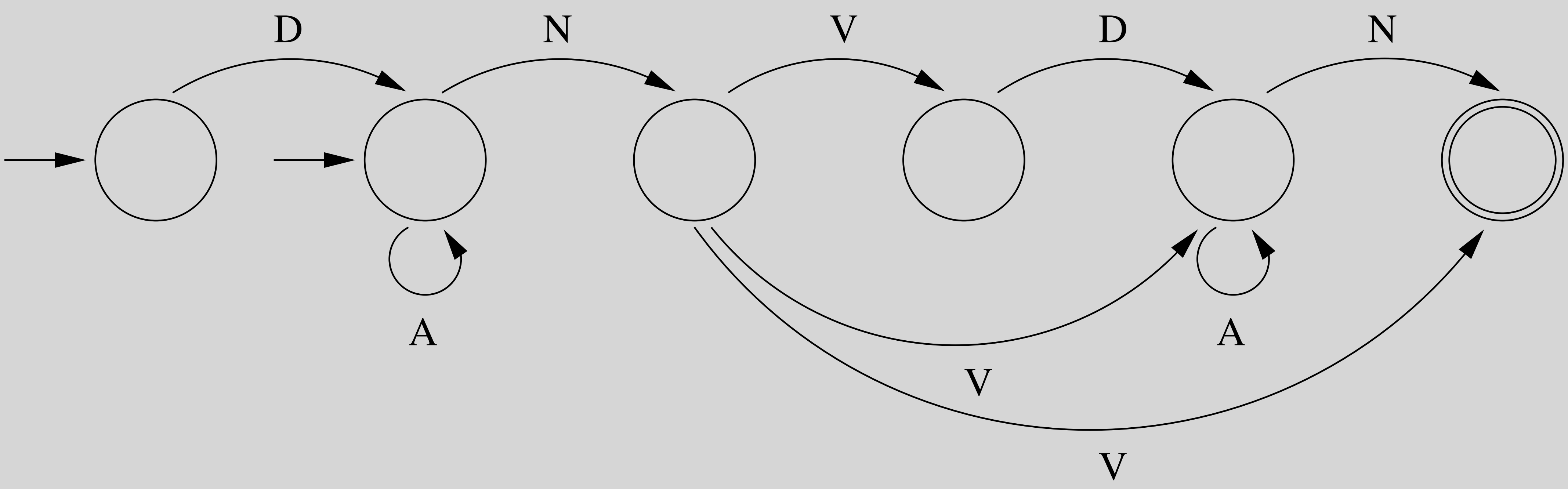
D A N V

- the noisy dogs chased the innocent cats

D A N V D A N

- $a^* = \{\emptyset, a, aa, aaa, aaaa, \ldots\}$
- $a^+ = \{a, aa, aaa, aaaa, \ldots\}$
- (D) A* N V ((D) A* N)

A Finite State Machine



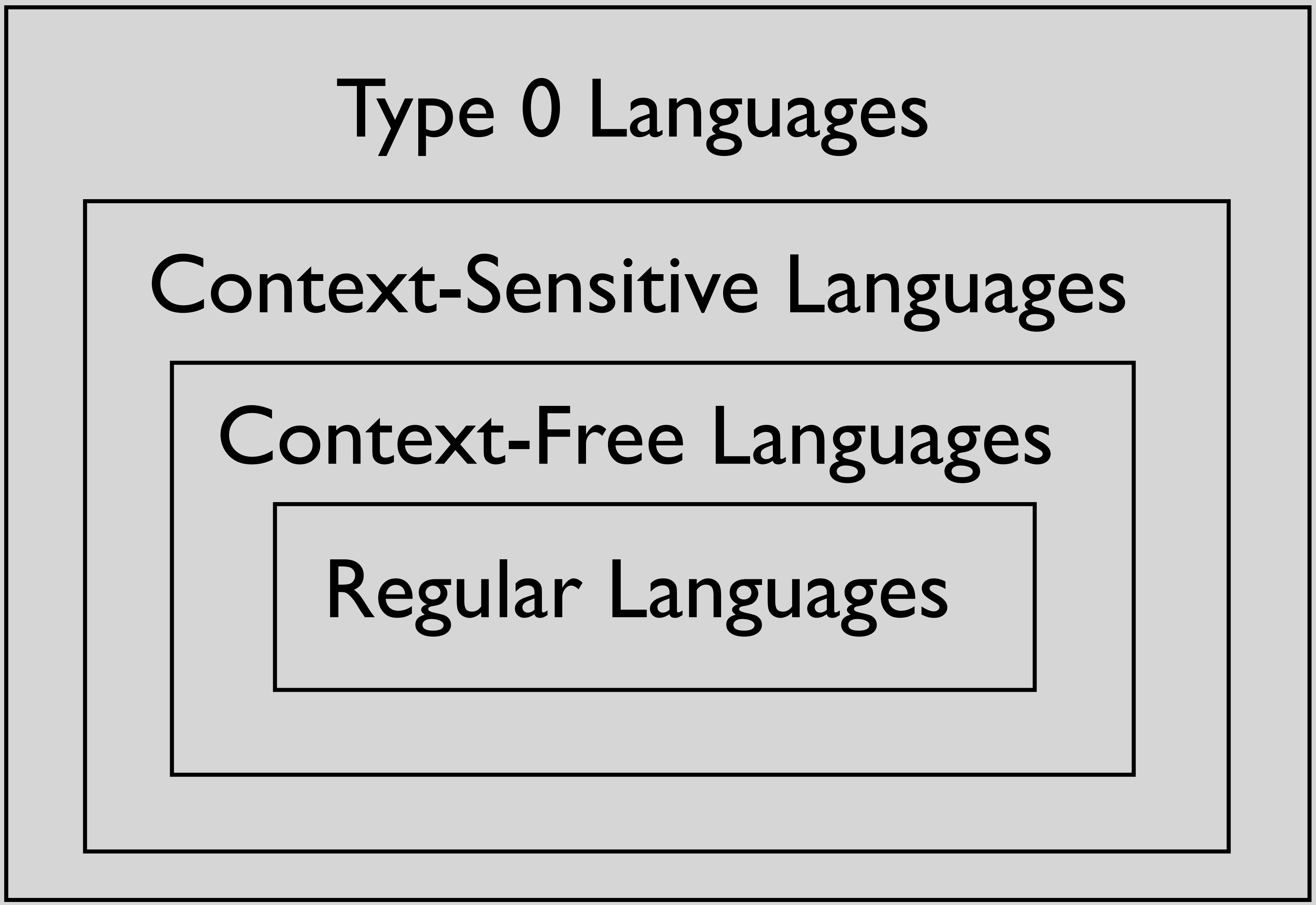
What does a theory do?

- 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



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

<table>
<tr>
<td>

<u>RULES</u>

S $\longrightarrow$ NP VP

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

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

PP $\longrightarrow$ P NP

</td>
<td>

<u>LEXICON</u>

D: the, some

A: big, brown, old

N: birds, fleas, dog, hunter, I

V: attack, ate, watched

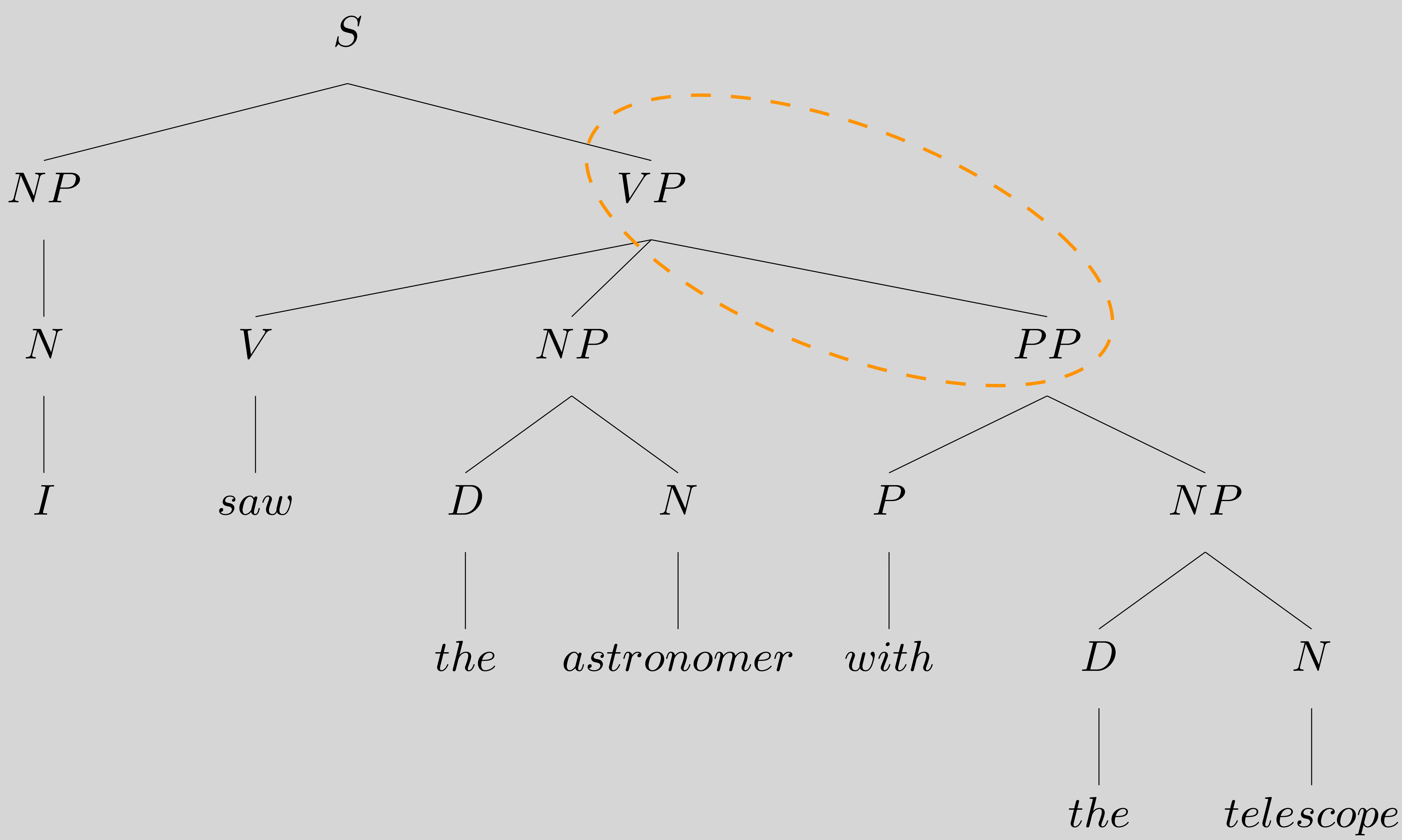
P: for, beside, with

</td>
</tr>
</table>

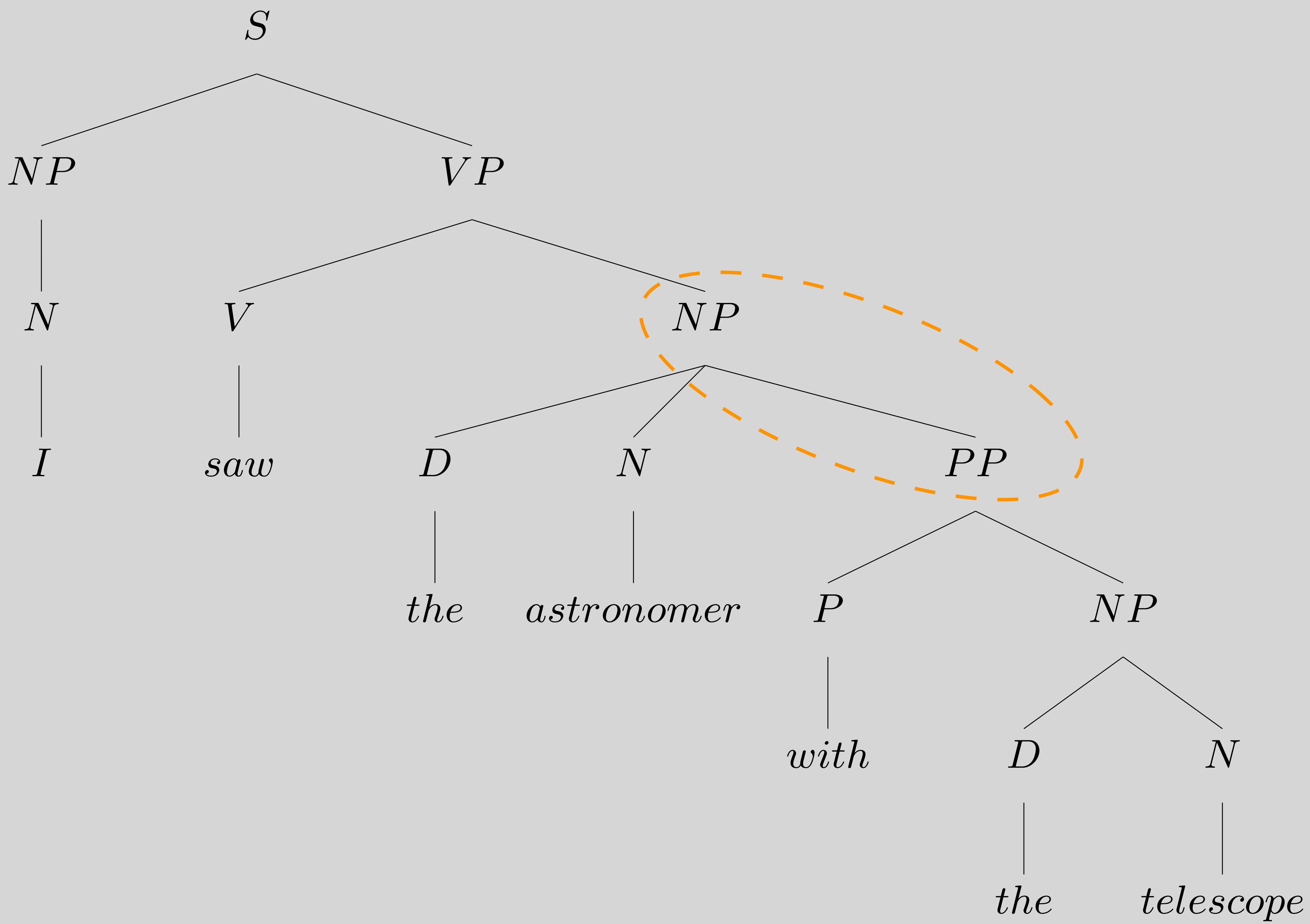
Structural Ambiguity

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.

Interim RQ

- The authors mention that topicalization and clefting are heuristics in establishing a constituent. I was wondering if there are formal or systemic ways of establishing a constituent which do not rely much on the heuristics.

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*.
  - *They arrive this morning*.

A Trivial CFG

$S \longrightarrow NP\ \ VP$

$NP \longrightarrow D\ \ N$

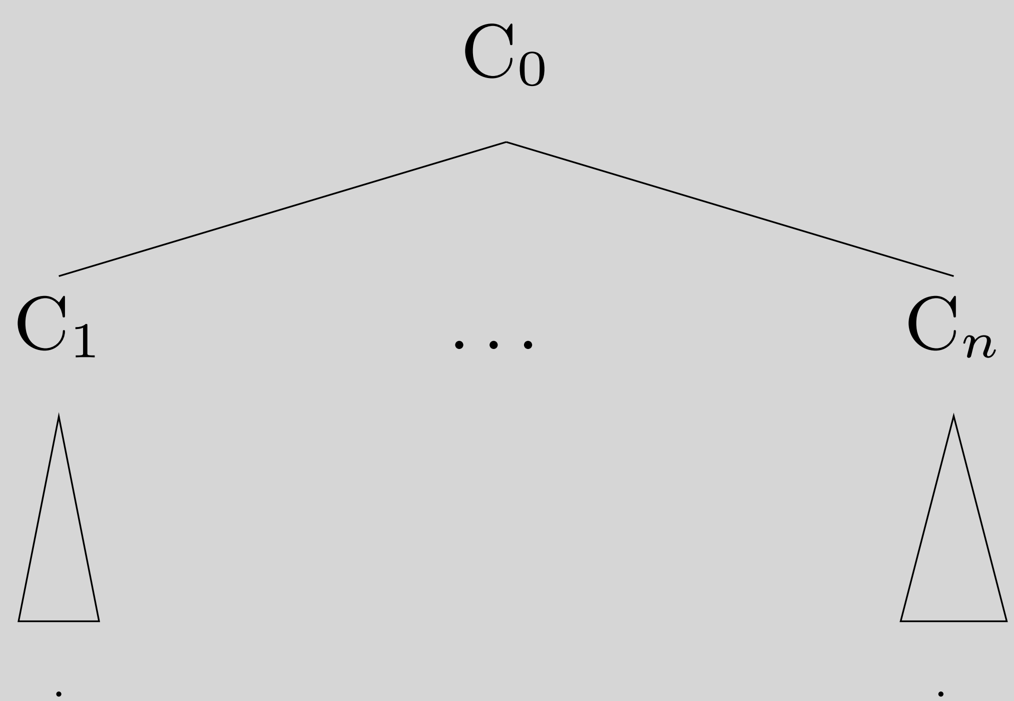
$VP \longrightarrow V\ \ NP$

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



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

are well-formed trees, and

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

Bottom-up Tree Construction

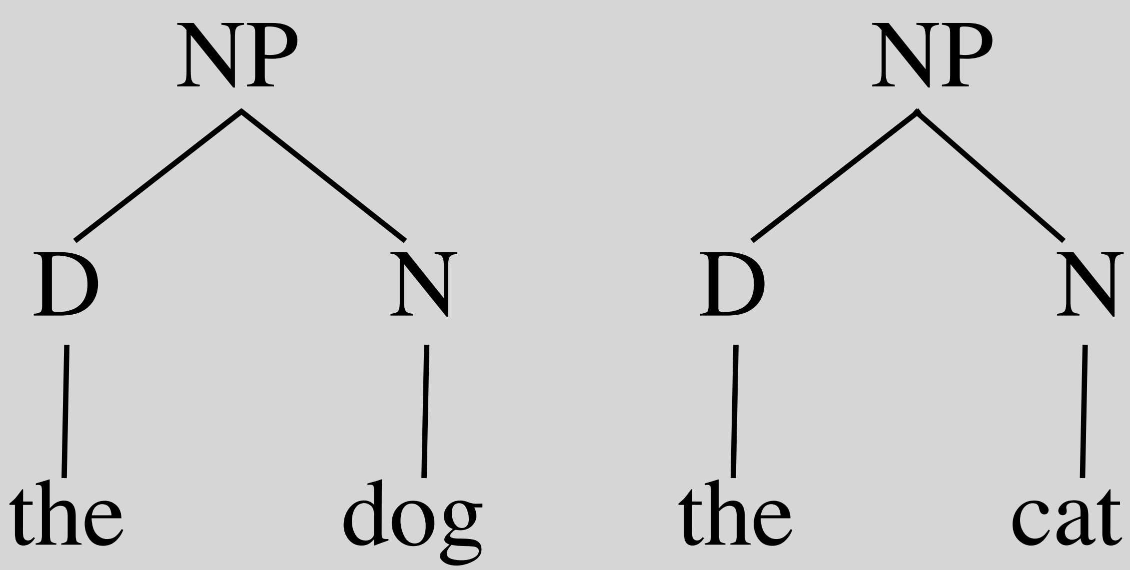
D: *the*

V: *chased*

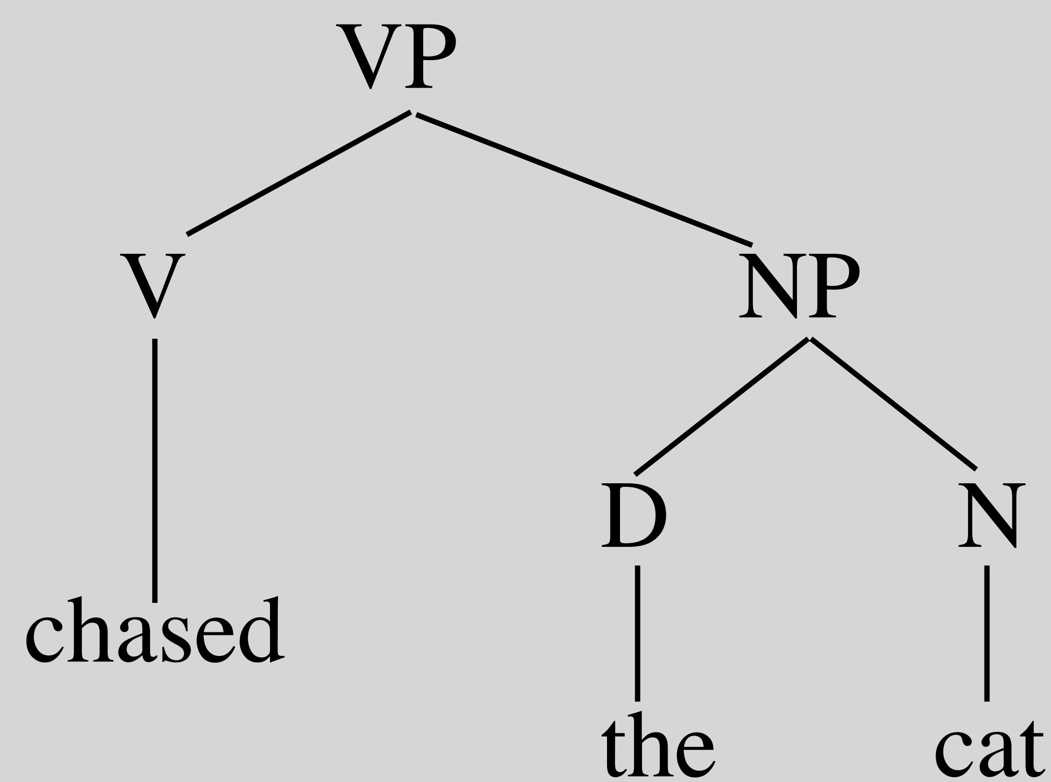
N: *dog, cat*

D	V	N	N
|	|	|	|
the	chased	dog	cat

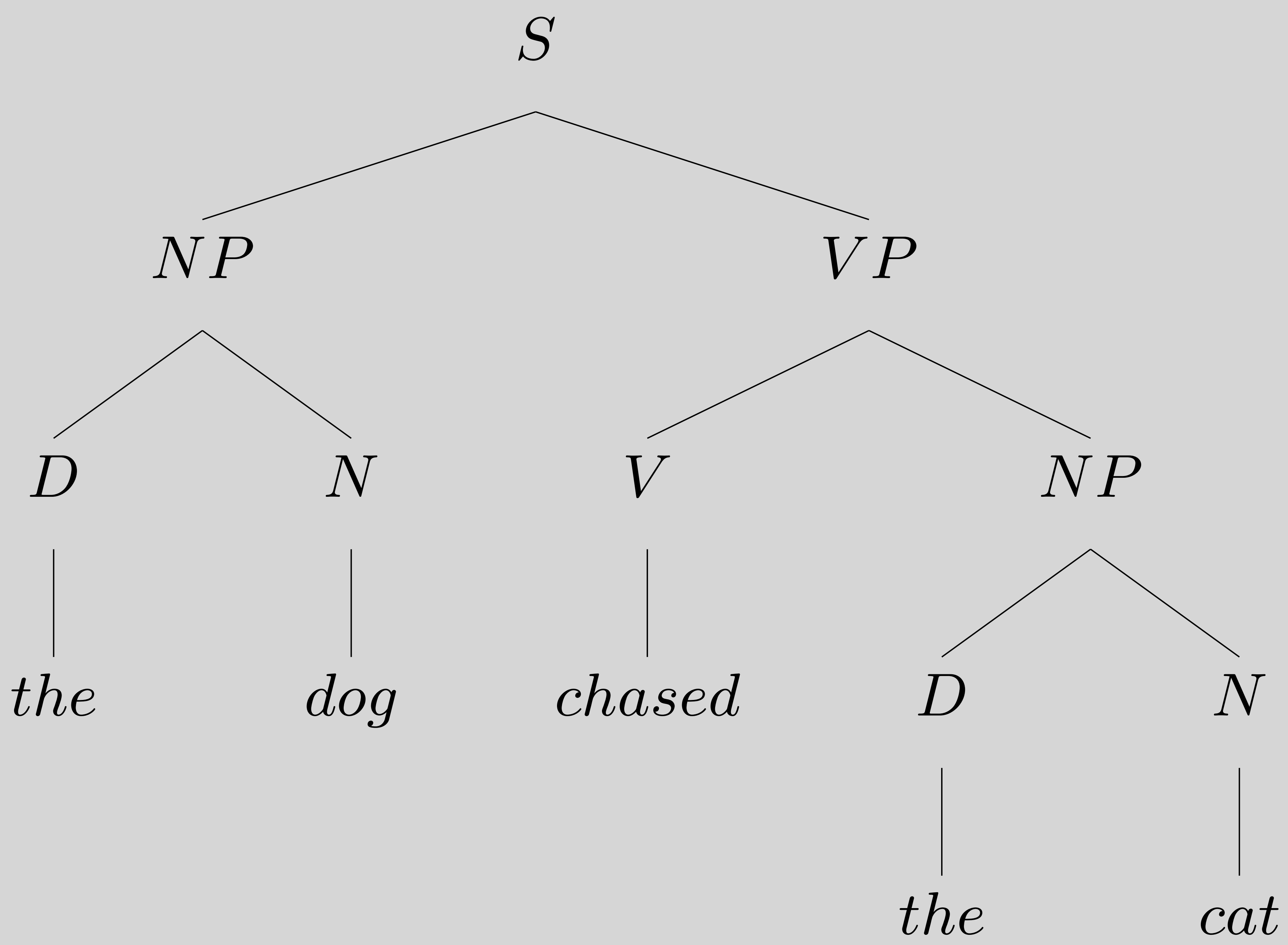
$NP \longrightarrow D \;\; N$



$VP \longrightarrow V \;\; NP$

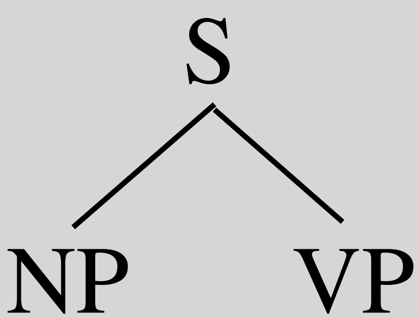


S $\longrightarrow$ NP VP

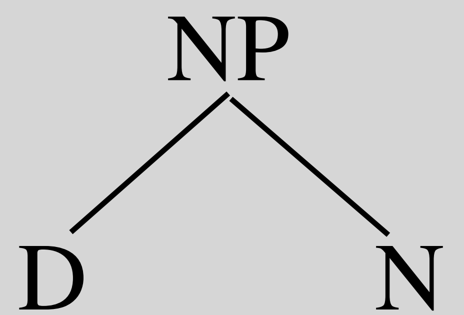


Top-down Tree Construction

$S \longrightarrow NP\ \ VP$

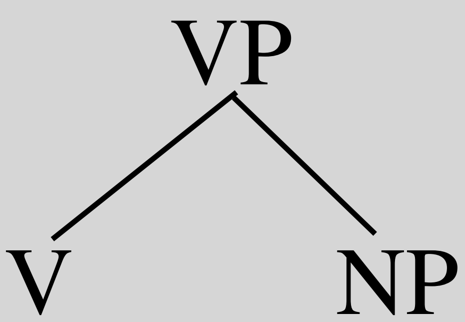


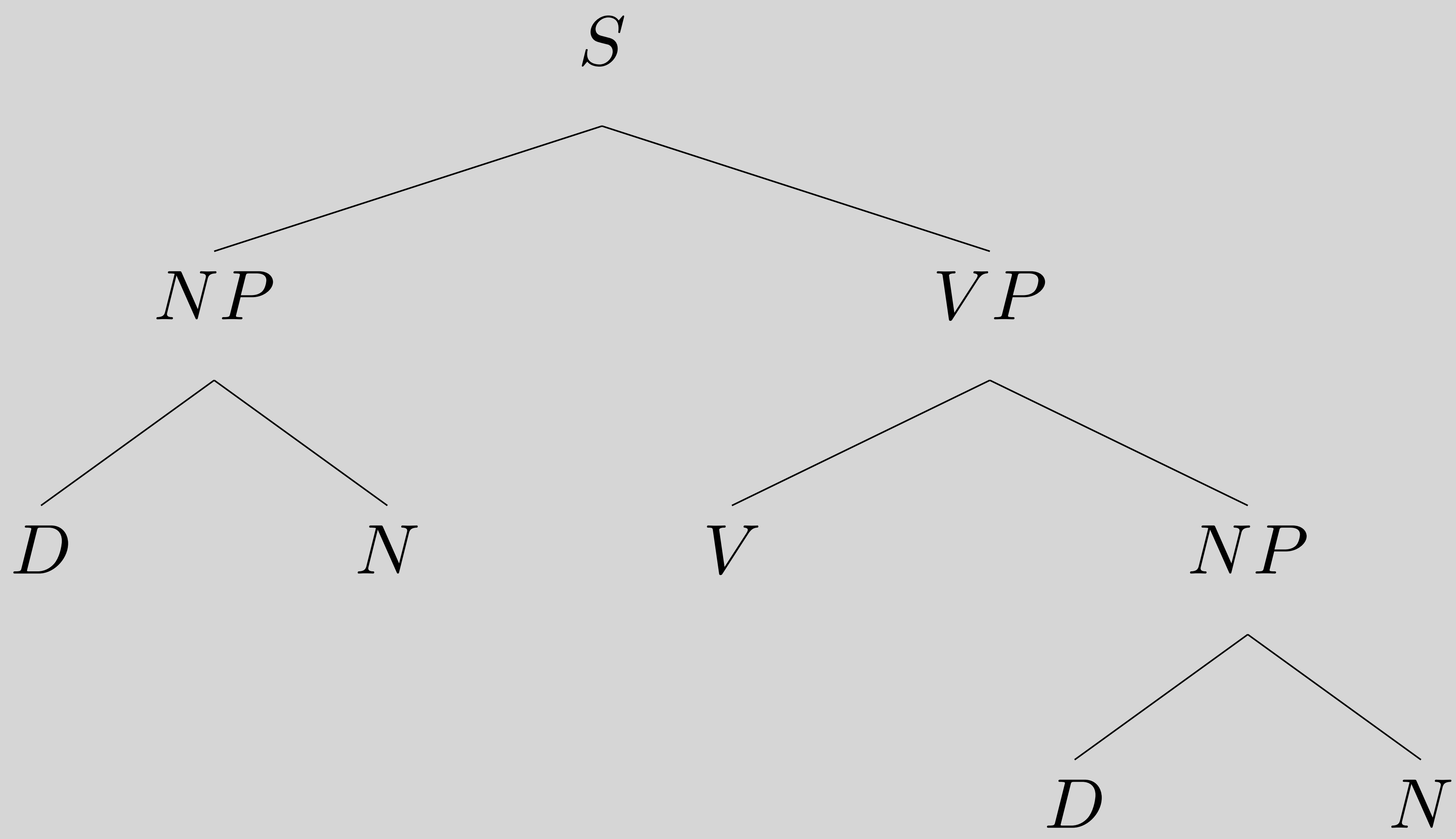
$NP \longrightarrow D\ \ N$



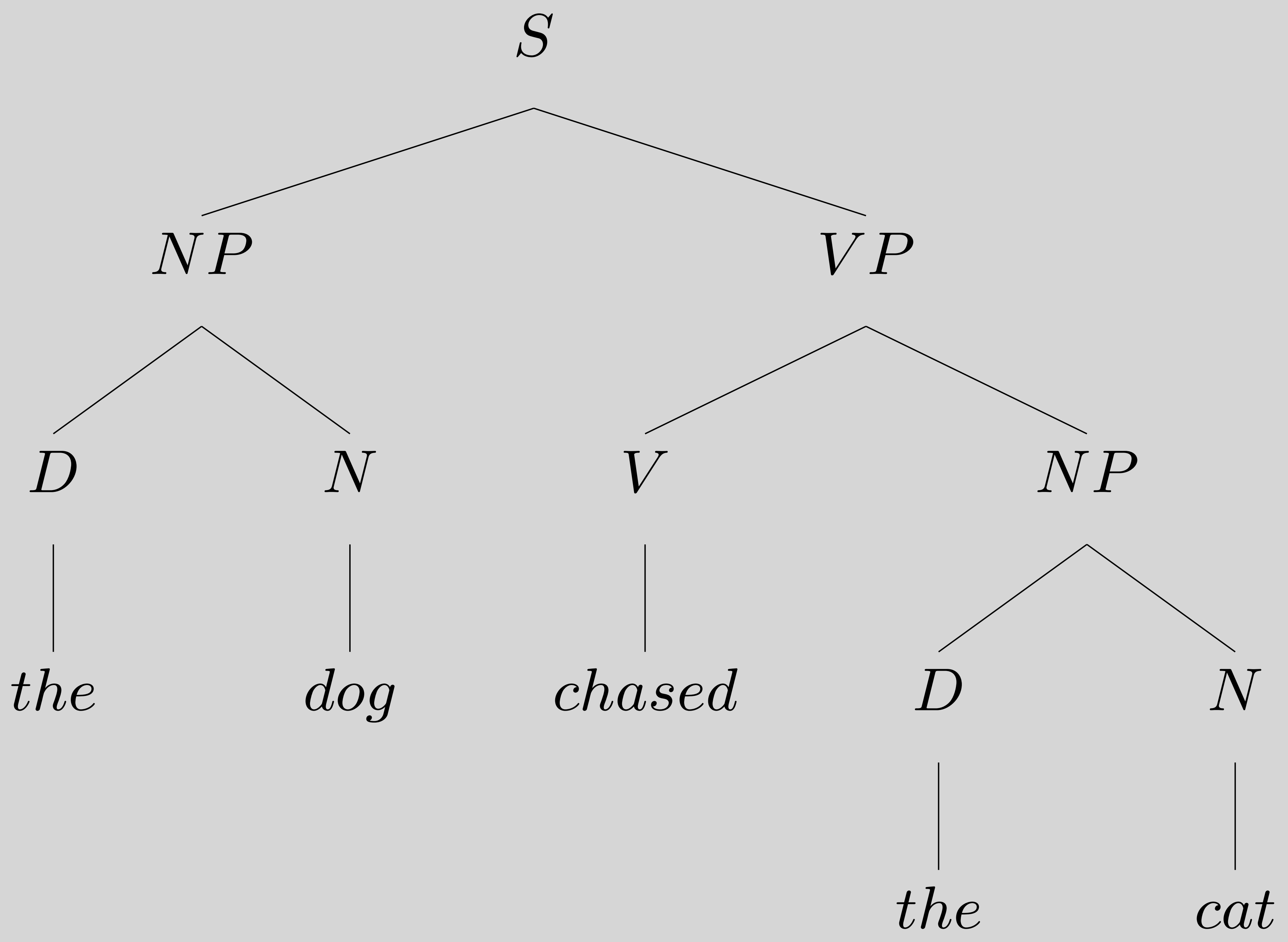
(twice)

$VP \longrightarrow V\ \ NP$




S
NP
VP
D
N
V
NP
D
N

D	V	N	N
the	chased	dog	cat


S
NP
VP
D
N
V
NP
the
dog
chased
D
N
the
cat

Weaknesses of CFG (w/atomic node labels)

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

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

- 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	$\rightarrow$	NP-SG VP-SG	VP-SG	$\rightarrow$	IV-SG
S	$\rightarrow$	NP-PL VP-PL	VP-PL	$\rightarrow$	IV-PL
NP-SG	$\rightarrow$	(D) NOM-SG	VP-SG	$\rightarrow$	TV-SG NP
NP-PL	$\rightarrow$	(D) NOM-PL	VP-PL	$\rightarrow$	TV-PL NP
NOM-SG	$\rightarrow$	NOM-SG PP	VP-SG	$\rightarrow$	DTV-SG NP NP
NOM-PL	$\rightarrow$	NOM-PL PP	VP-PL	$\rightarrow$	DTV-PL NP NP
NOM-SG	$\rightarrow$	N-SG	VP-SG	$\rightarrow$	CCV-SG S
NOM-PL	$\rightarrow$	N-PL	VP-PL	$\rightarrow$	CCV-PL S
NP	$\rightarrow$	NP-SG	VP-SG	$\rightarrow$	VP-SG PP
NP	$\rightarrow$	NP-PL	VP-PL	$\rightarrow$	VP-PL PP
	$\ldots$			$\ldots$	

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 $f(\text{SG})$:

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

- Let r be the regular language $wa^*b^*xc^*d^*y$
- $f(\text{SG}) \cap r = wa^m b^n xc^m d^n y$
- $wa^m b^n xc^m d^n y$ 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.

Overview

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

Reading Questions

- What is transformational grammar and what does it have to do with CL/NLP?
- What does the non-transformational approach get right that the transformational approach gets wrong?
- How can we tell if we are using the transformational grammar? (43) shows exactly the same representations of non-terminal symbols.

Reading Questions

- I'm curious to see how transformational grammars hold up in efficiency when compared to non-transformational CFGs and head-driven grammars. Do transformational grammars inherently have less rules? If so, because it is an abstraction from what you see or hear in language, they must be difficult to formalize. So I wonder if using the the addition of the head with CFGs can cover all agreement and variation of the features of English. In addition to capturing "headedness", does identifying a head help reduce the redundancies in CFGs?

Reading Questions

- How much evidence is generally observed before we form a rule for a language? It seems like one counter-example is enough evidence to modify a rule, however too small of a sample might lead to erroneous observations. What is generally accepted as an acceptable sample size?

Reading Questions

- What do we mean when we say we are interested in descriptively building a grammar for English? Even if we specify American English, there are many different speaker groups to choose from. In short, reading this chapter made me wonder, where are we drawing our judgments of what is acceptable or not? Where is our data coming from, and how does that get translated into a concise (and descriptive) grammar of the English language?

Reading Questions

- Like several others, I am also curious about judgement calls. I've found that my judgements are usually quite a bit more forgiving. I have no idea if this is the result of Chicano Spanish being my first language, but I would accept sentences that others might immediately rule out. For instance:
  - The children will all see the movie.
  - *The children all will see the movie.
- When looking at the tree structures for these, I understand why the second example is considered poor form, but I wouldn't mark it with an asterisk. Is it possible for a formal grammar to tell us that such "mistakes" could be acceptable to certain speakers of English?

Reading Questions

- At what level of generality are we interested in creating our grammars? It makes intuitive sense that you would need different grammars for different languages (as illustrated by Problem 4). But within a language, how do we determine whether differing dialects require differing grammars? Or do we propose dialectic augmentations to some common grammar? What level of evidence or acceptability is required to split off a new rule or grammar?

Reading Questions

- The types of grammar that this chapter tries to formulate is essentially a function that takes a sentence as the input and outputs a boolean value indicating its grammaticality/ acceptability. If that is the case, this formulation (i.e. this function) fails to account for language variation, unless it takes "the speaker" as an additional argument and gradability in judgments.

- A language-model-like formulation, then, seems to be more capable of accounting for these phenomena, yielding a *probability* of a sentence. This probability could encode both the percentage of native speakers in some population that finds a sentence acceptable, as well as the extent to which they like the sentence. Such probabilistic models have an important status in the current computer science world, but is there any relevant discussion in the linguistic literature?

Reading Questions

- While reading Ch. 2, I was thinking about how theories of syntax are/can be used with with spoken language. On page 43, the authors make a point of separating usage and abstract syntactic representations, pushing much of spoken language outside our domain of inquiry. But I still wonder how we can use syntactic theories to model spoken language when it contains “errors” and disfluencies. Would we propose an underlying complete or correct syntactic tree for such utterances? If not, then how else might it be modeled?

Reading Questions

- I'm also curious about how linguists decide what is grammatical/ should be included in the grammar of a language. I'm specifically wondering about cases where a sentence may be grammatical under a grammar of a language, but still not be understandable by a fluent speaker (such as the examples we saw in class on Monday). Do linguists generally take "acceptability" (understandable to a fluent speaker) into account when developing a grammar of a language? If not, why do we want a grammar to cover sentences that a native speaker can't parse?

Reading Questions

- I was a little uncertain on the section about nominals (NOM). I can understand why it is helpful to have a new non-lexical category for the examples given, but what would the exact English definition of the nominal category be?

- It seems that structural ambiguity is a difficult problem, considering that none of lists, regex, and CFG could solve it. However, some cases, like "noisy children and animals", are even ambiguous to human readers. Is it ever possible to develop a "correct" prescriptive grammar even when human readers are unable to identify the real meaning?

- What are some things we should know about the diversity of languages and how well they can be described by different types of grammars? I would take Pirahã and Swiss German to be on different extremes. When would CFGs be useful, though inaccurate approximations?

Reading Questions

- I do not quite understand the meaning of ‘much of the outward differences among languages can be viewed as differences in vocabulary.’(P21) Are there any concrete and convincing examples? Is the conclusion too general? Because if we take English and Chinese as an example, according to my understanding, the differences between English and Chinese are more than differences in vocabulary. And I’m also curious about how can transformational grammar account for the languages like Chinese which have no inflection and are more free in word order.