Ling 566
Sept 30, 2025
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?

An example hypothetical language (p.22)

- Some sentences go on and on
- *Sentences some go on and on
- *Some sentences go on and on and on
- Sentences some go on and on and on
- Some sentences go on and on and on and on
- *Sentences some go on and on and on and on

Insufficient Theory #2: FSMs

- the noisy dogs left

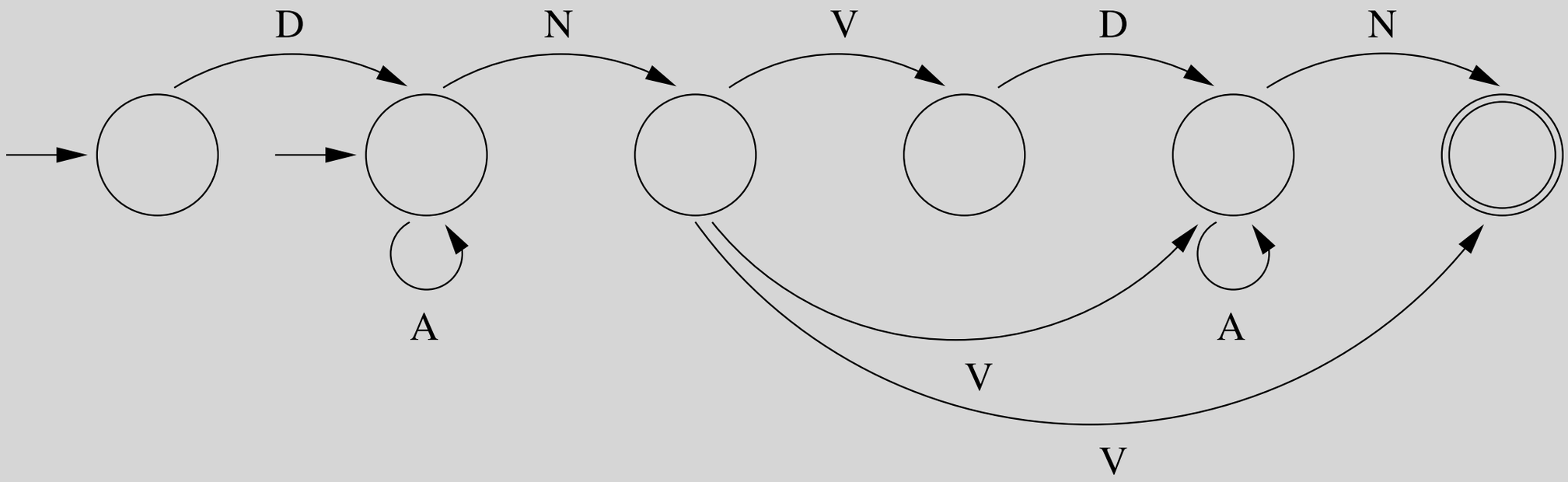
D A N V

- 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



Insufficient Theory #2: FSMs

- What's wrong with this?
- What can't it model?

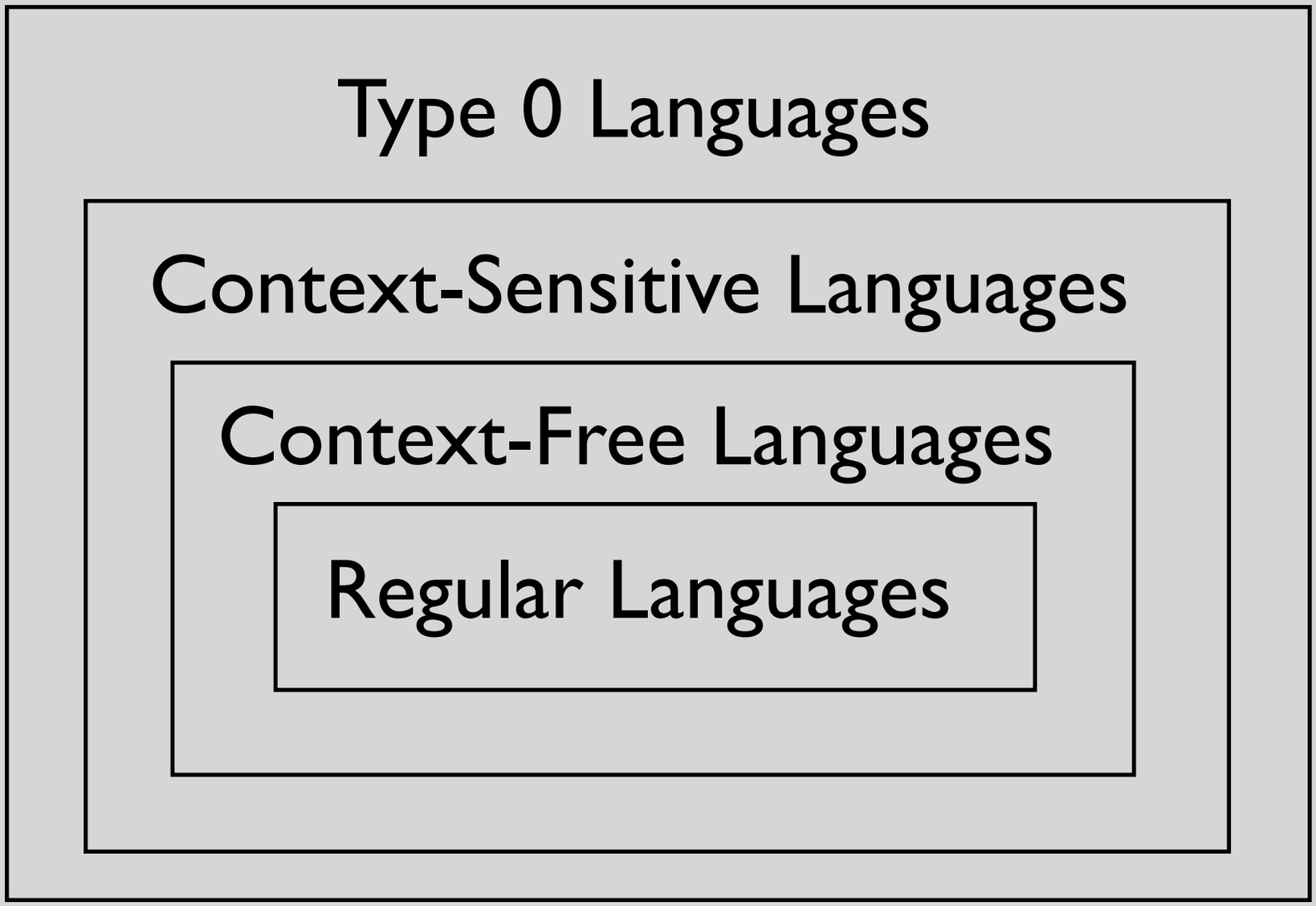
What does a theory do?

- 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

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

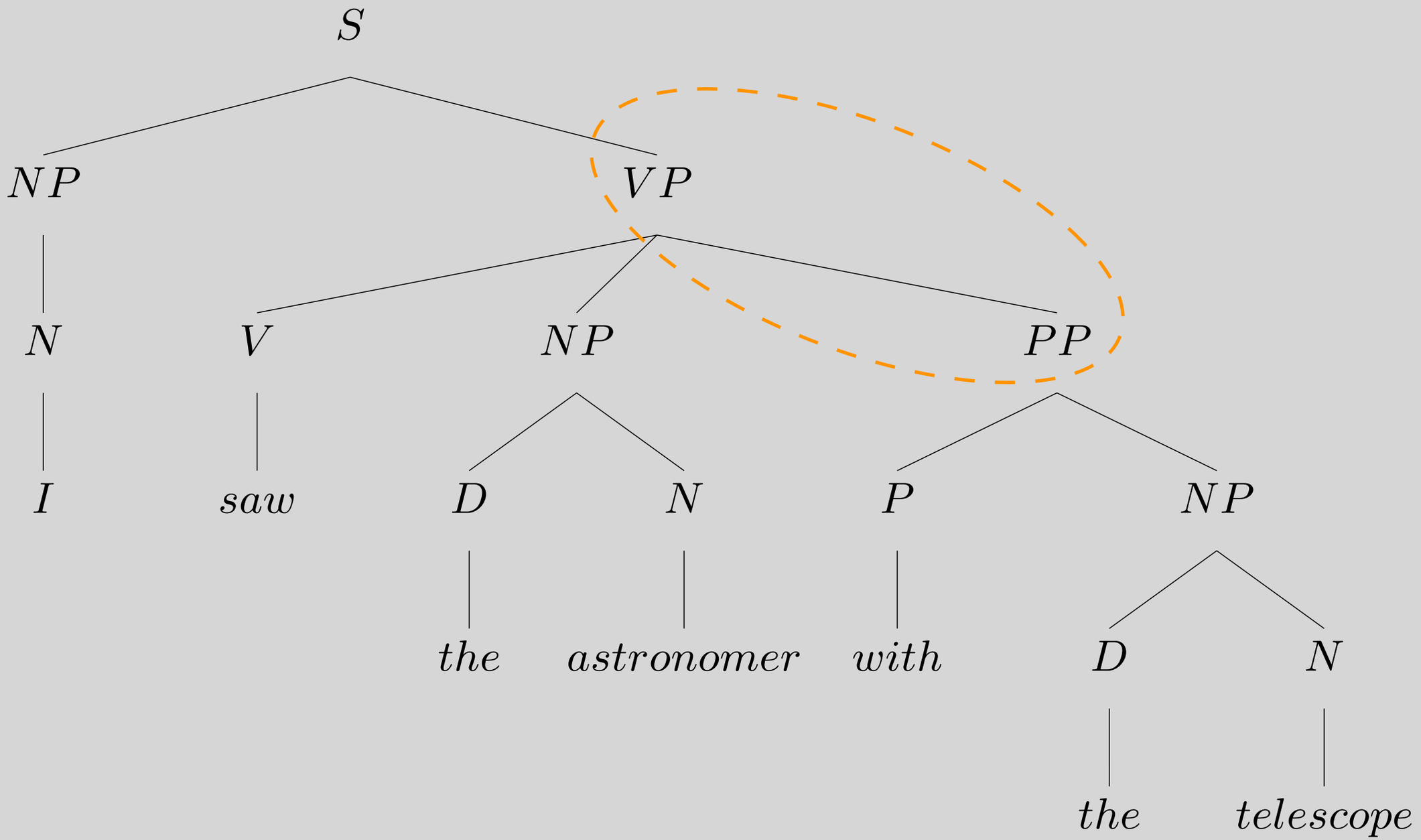
V: attack, ate, watched

P: for, beside, with

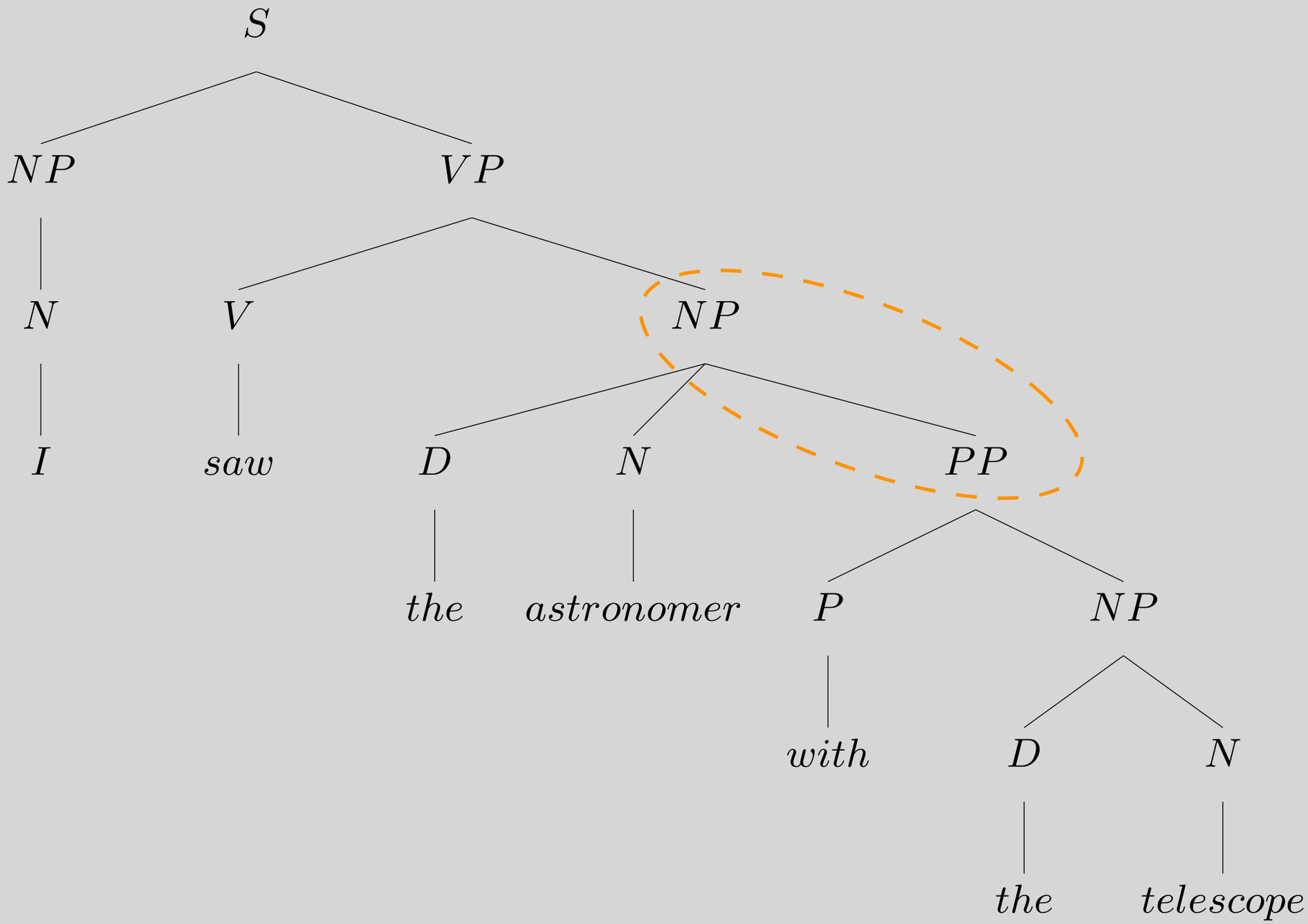
Structural Ambiguity

I saw the astronomer with the telescope.

Structure 1: PP under VP



Structure 2: 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→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

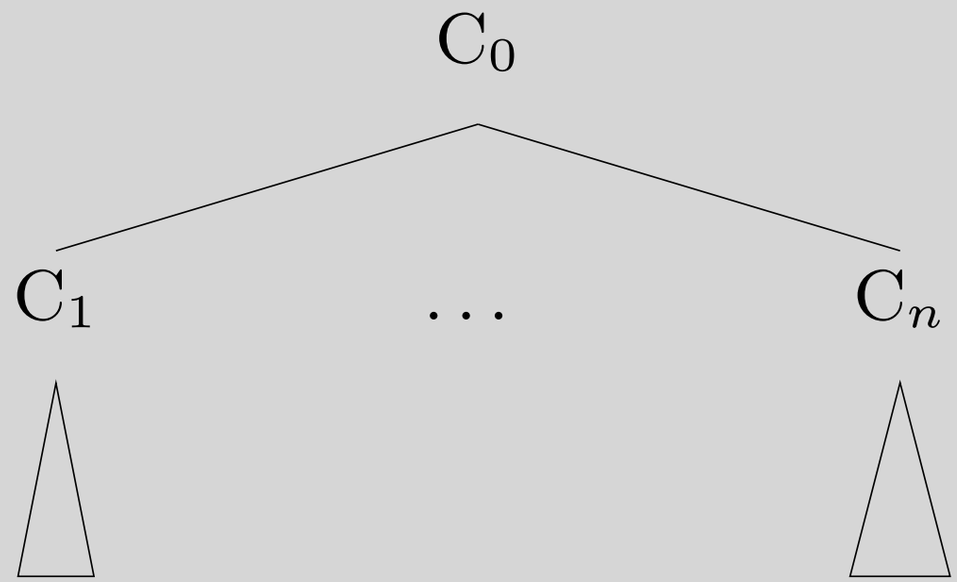
VP $\longrightarrow$ V NP

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



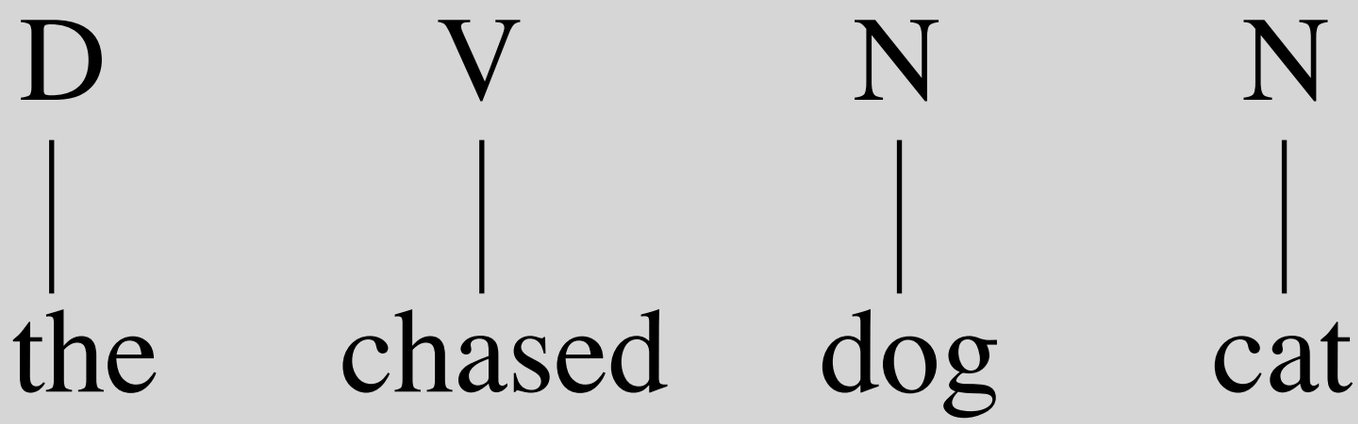
is a well-formed nonlexical tree if (and only if)

are well-formed trees, and

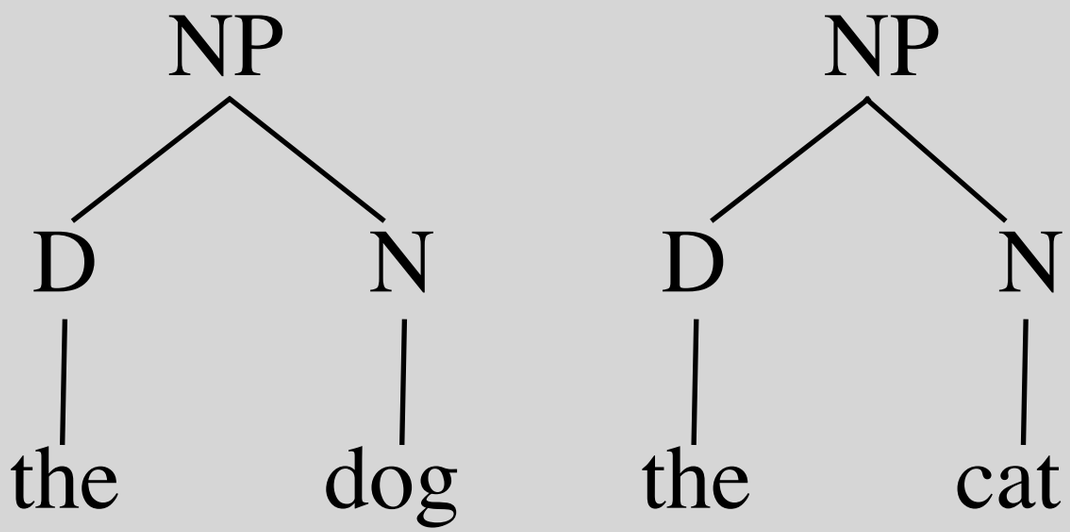
$C_0 \rightarrow C_1 \ldots Cn$ is a grammar rule.

Bottom-up Tree Construction

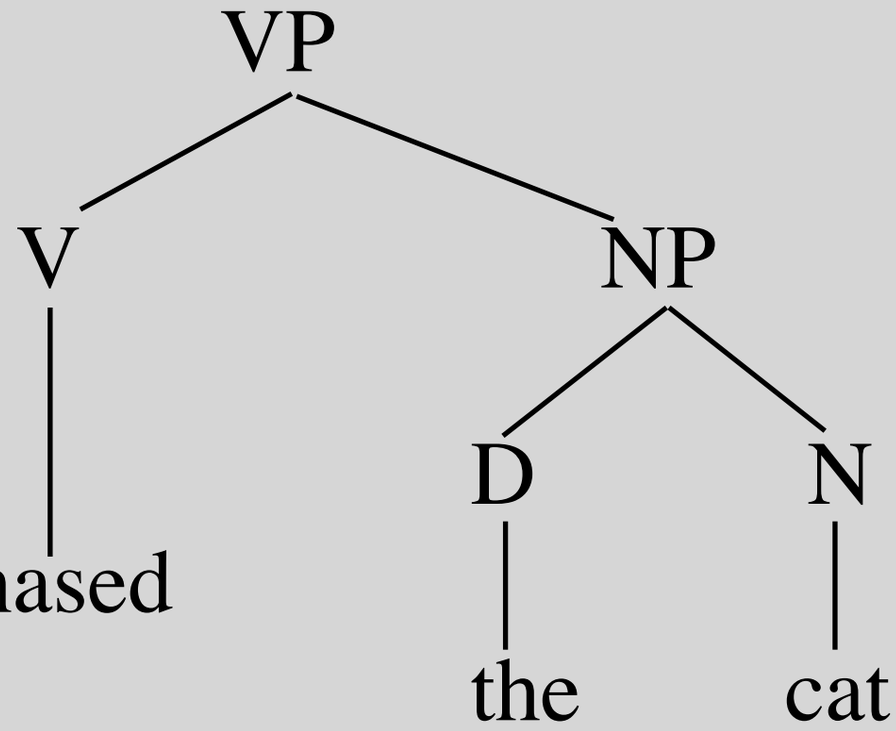
D: *the*
V: *chased*
N: *dog, cat*



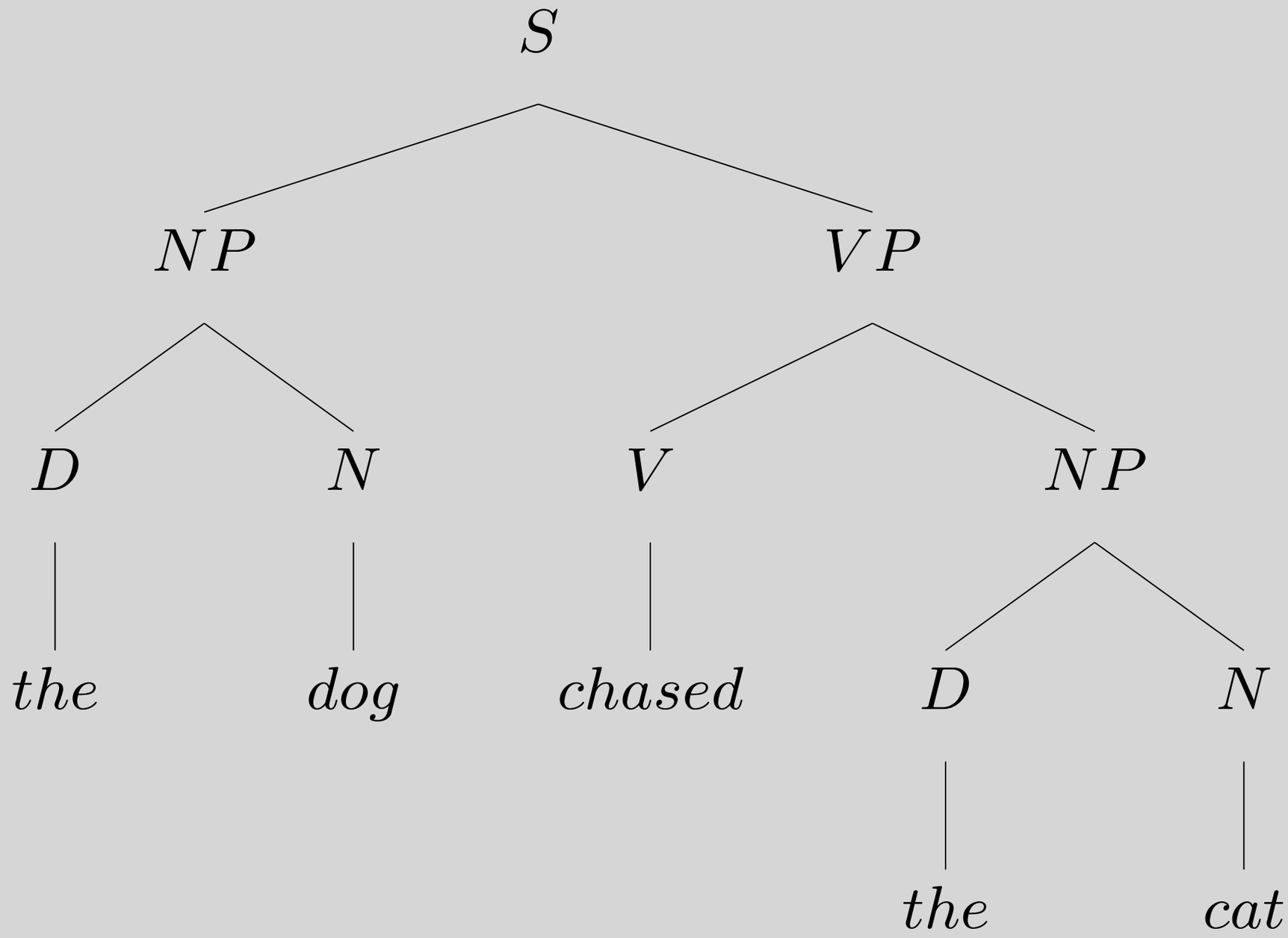
NP ⟶ D N



VP ⟶ V NP

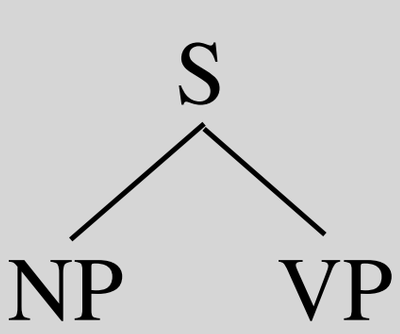


S ⟶ NP VP

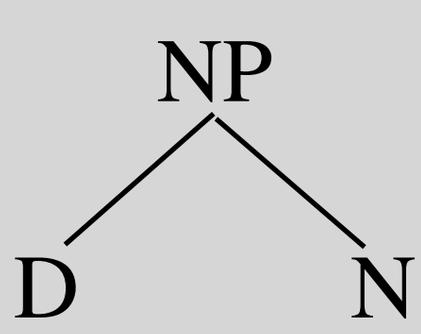


Top-down Tree Construction

$S \longrightarrow NP\ \ VP$

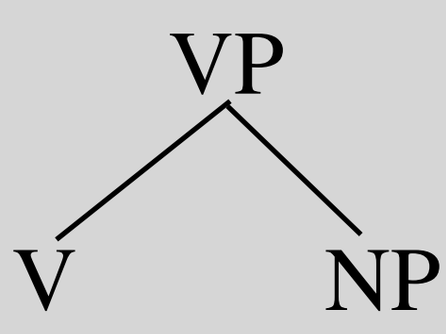


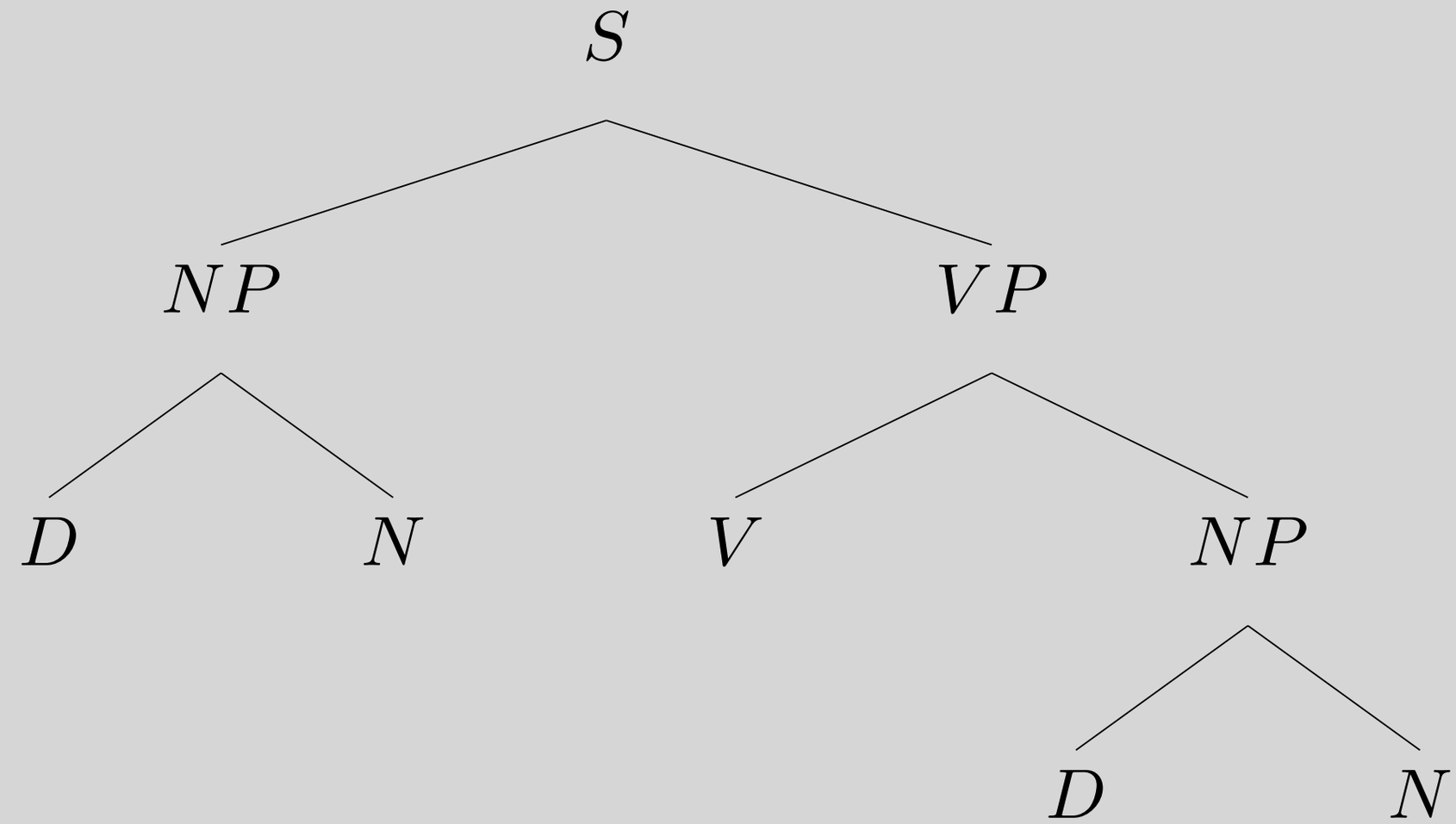
$NP \longrightarrow D\ \ N$



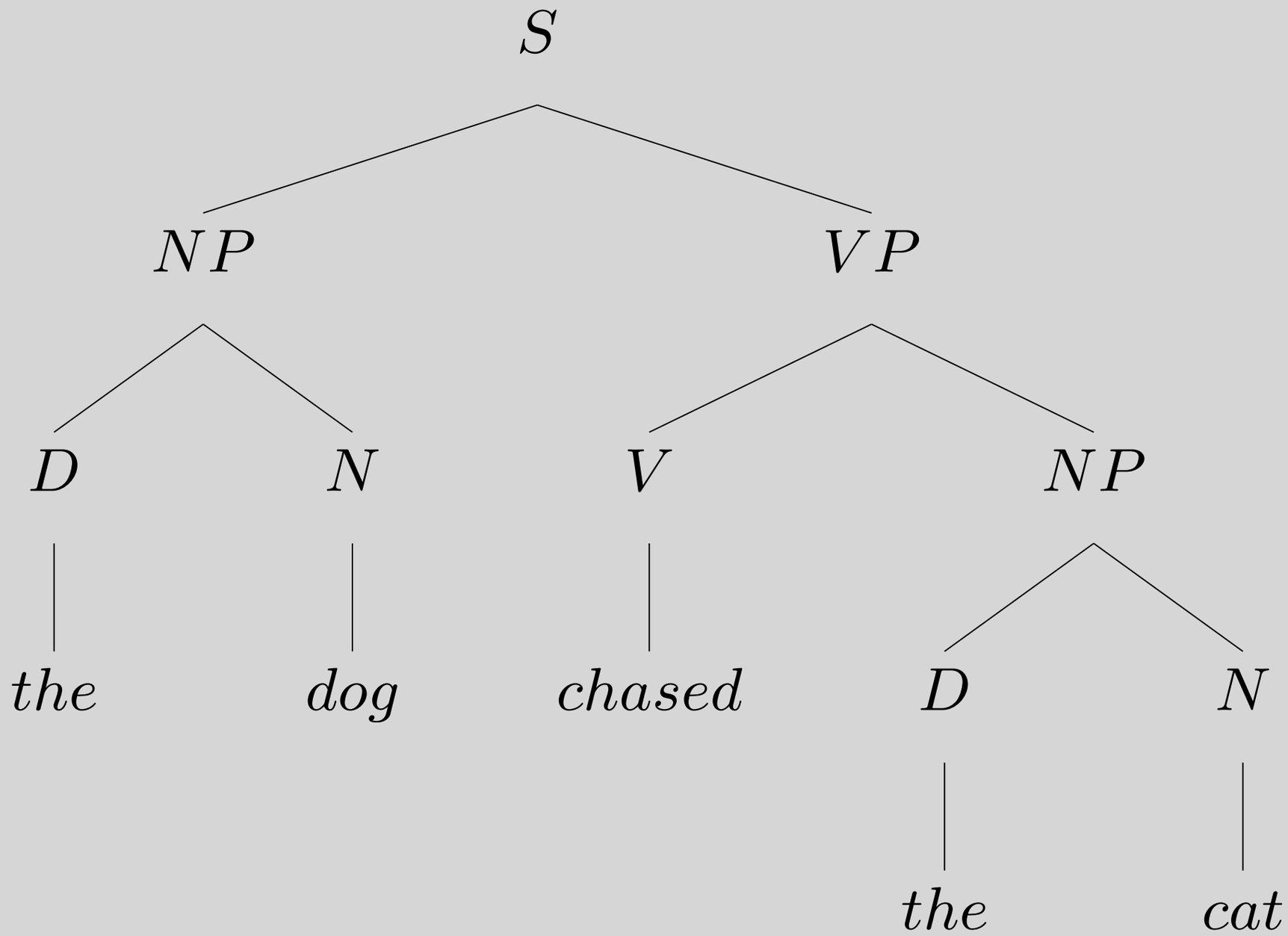
(twice)

$VP \longrightarrow V\ \ NP$





D	V	N	N
the	chased	dog	cat



For parsing, which style feels most intuitive to you?

0

Bottom up

0

Top down

0

Left to right

0

Something else

0

None of the above

0

Weaknesses of CFG (w/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	$\rightarrow$	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
	$\rightarrow$	...		$\rightarrow$	...

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(SG):

f(d'chind)	=	a	f(Jan säit das mer)	=	w
f(em Hans)	=	b	f(es huus)	=	x
f(lönde)	=	c	f(aastriiche)	=	y
f(hälfe)	=	d	f([other])	=	z

- 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. w/reg languages
- $\therefore$ f(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 SG is not weakly context-free and a fortiori not strongly context-free.

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

RQ: NOM

- Why NOM and not N-bar / N'?
- Page 32: How does NOM solve a problem that couldn't be explained using just N and NP? In example (21) *no painting by Miro or drawing by Klee*, it seems like the same structure could be represented with NP coordination rule instead of introducing NOM --> N or NOM --> NOM PP.

Which rules (w/PP->P NP, NP-> D N, and coord rule) license "no painting by A or drawing by B was displayed"?

0

NP -> NP PP

NP -> D NOM; NOM -> NOM PP; NOM -> N

NP -> (D) N (PP)

None of the above

Which rules (w/PP->P NP, NP-> D N, and coord rule) license "no painting by A or drawing by B was displayed"?

0

NP -> NP PP

0

NP -> D NOM; NOM -> NOM PP; NOM -> N

0

NP -> (D) N (PP)

0

None of the above

0

Which rules (w/PP->P NP, NP-> D N, and coord rule) license "no painting by A or drawing by B was displayed"?

0

NP -> NP PP

0

NP -> D NOM; NOM -> NOM PP; NOM -> N

0

NP -> (D) N (PP)

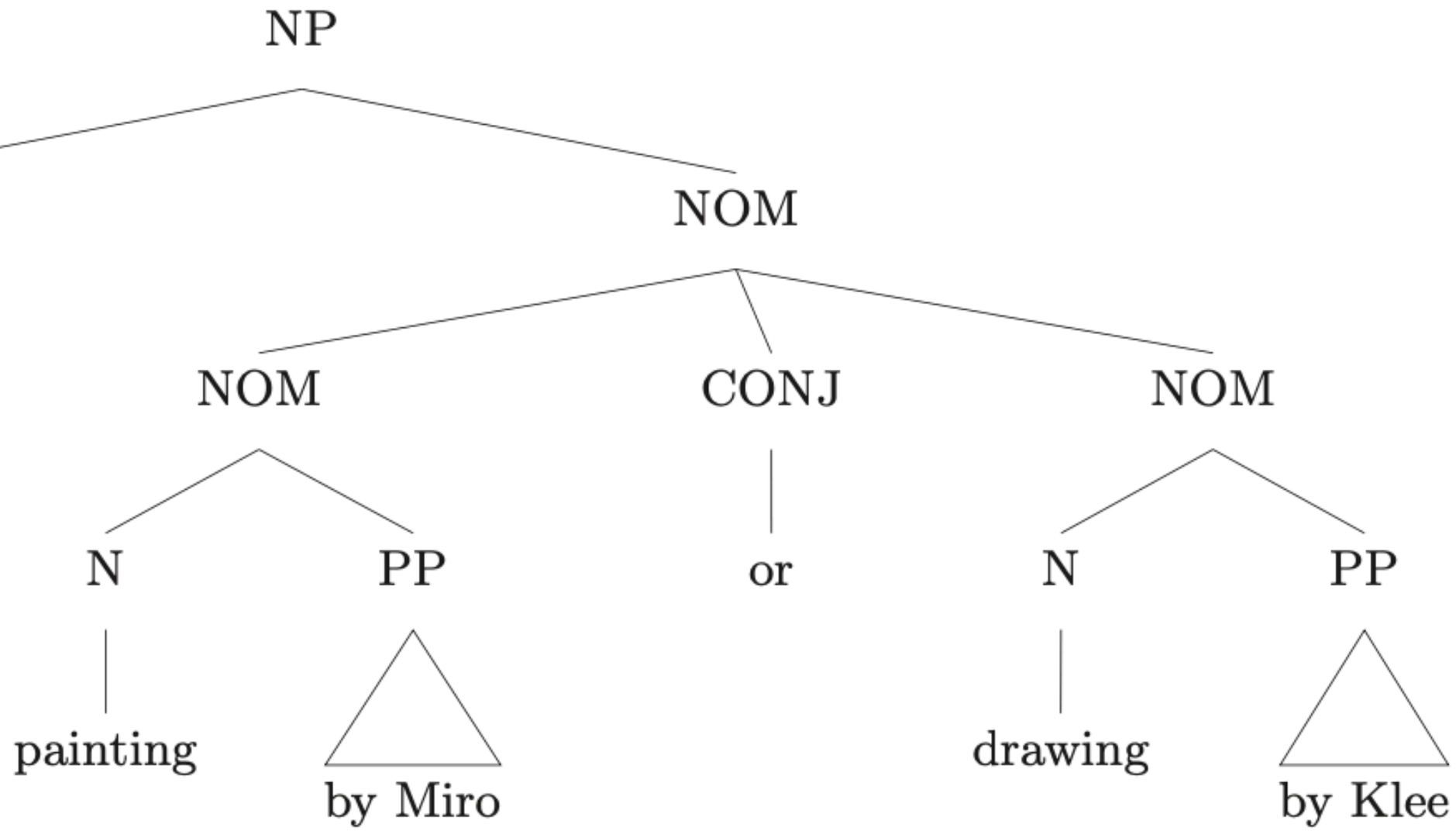
0

None of the above

0

RQs: NOM

(22)



RQ: POS

- When assigning lexical categories to words, how would we handle words "acting" as a different lexical category? For example, in the phrase "the admired man", would "admired" still be treated as a verb and need to be accounted for under "V" in our grammar, or would we just treat "admired" like an adjective because its position in the phrase has it modifying a noun? Similarly, how would we treat adjectives predicating sentences in languages such as Japanese?

RQ: Ambiguity

- If CFG explains ambiguity by allowing multiple full parses of the same string, does that mean our brains are actually building a bunch of different trees at the same time when we hear something ambiguous? Or could it be that we just build one structure at first and then revise it on the fly, maybe using probabilities or semantics to guide us?

RQ: Heads/headedness

- "The notion of headedness is a problem for CFG because it cuts across many different phrase types, suggesting that the rules are too fine-grained."
- What is meant by headedness "[cutting] across many different phrase types."?

RQ: Tree terminology

- Page 34 details terminology for describing tree structures. In the example given (see image), "the" and "cats" are parts of a single NP, but that NP itself does not have a lexical item associated with it. Does this make "NP" itself the "mother" node, or should the implied lexical constituent be considered the mother node? Are the "true" daughter nodes of the NP "D" and "NOM" or their leaf nodes, "the" and "cats", or are these not separable components (the lexical trees)?

RQ: Transformational Grammar

- Transformational grammars are briefly described in this chapter. Is a transformational grammar more difficult to use in computational linguistics (than HPSG)? If so, what makes it more difficult?

- I was a bit confused by the idea of transformations. It seems like the idea is there is an "underlying" form of the sentence which may appear as one of several alternatives such as plural, passive, etc.

- How does that transformational agreement rule actually work? How does it reduce the ambiguity?

RQ: Context

- This chapter introduced context-free grammar without much discussion of "context." I'm curious how “context” is captured in context-sensitive grammars.

RQ:

- "There are some aspects of language that are primarily manifestations of individual speakers' representations..." - could we go over some examples?
- “Some believe grammars are a theory of people's mental representation of language”: Is there really any way to prove these theories though? It's probably very hard to actually concretely find out what people's mental representation of language are. How then do we verify that any of our grammars and theories or do we actually not care too much because grammars provide other uses and if so what other uses?

RQ: Linguistic theory

- Throughout reading this chapter, what struck me the most was the discussion of "natural" vs "unnatural", and how to best capture "linguistically significant generalizations". My (broad) question is, where do these judgments come from? In 2.7.1 there was mentions of these "unnatural" phrase structure rules that no linguist would ever want to write. What is considered more natural? Something that makes more logic sense cross-linguistically? In 2.2.1 it mentions that certain theories can fail to capture certain significant generalizations. Where is the consensus reached on what is significant? How do we determine these generalizations in the first place?