Ling 566
Oct 3, 2023

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

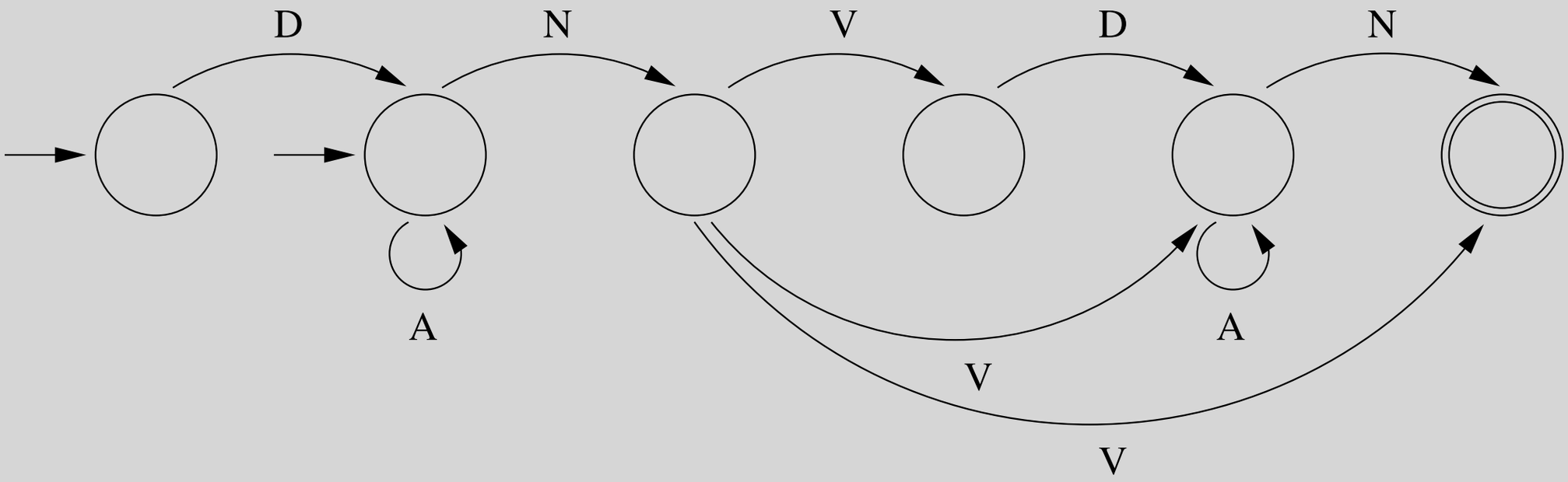
D A N V

- 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?

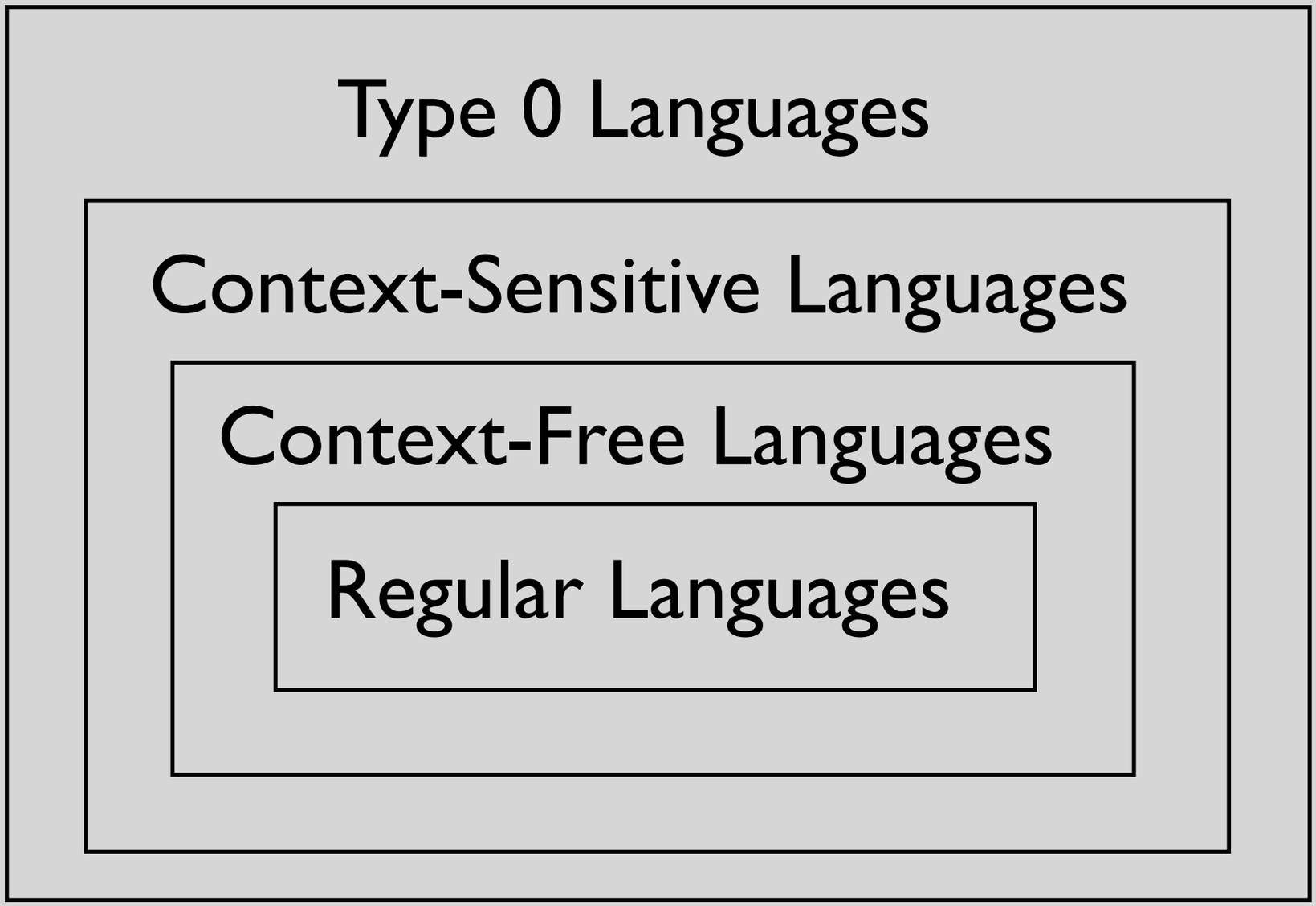
What does a theory do?

- 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

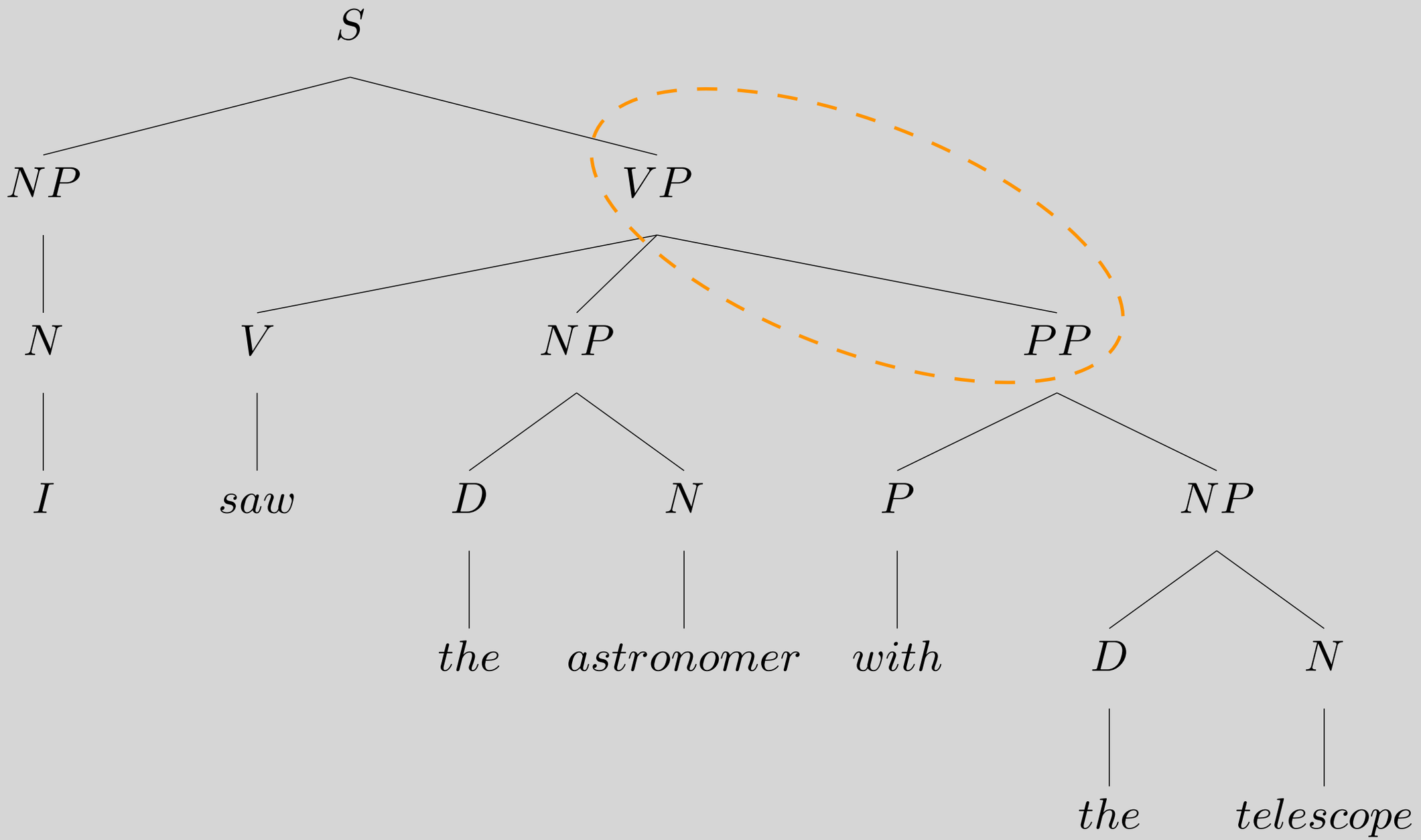
V: attack, ate, watched

P: for, beside, with

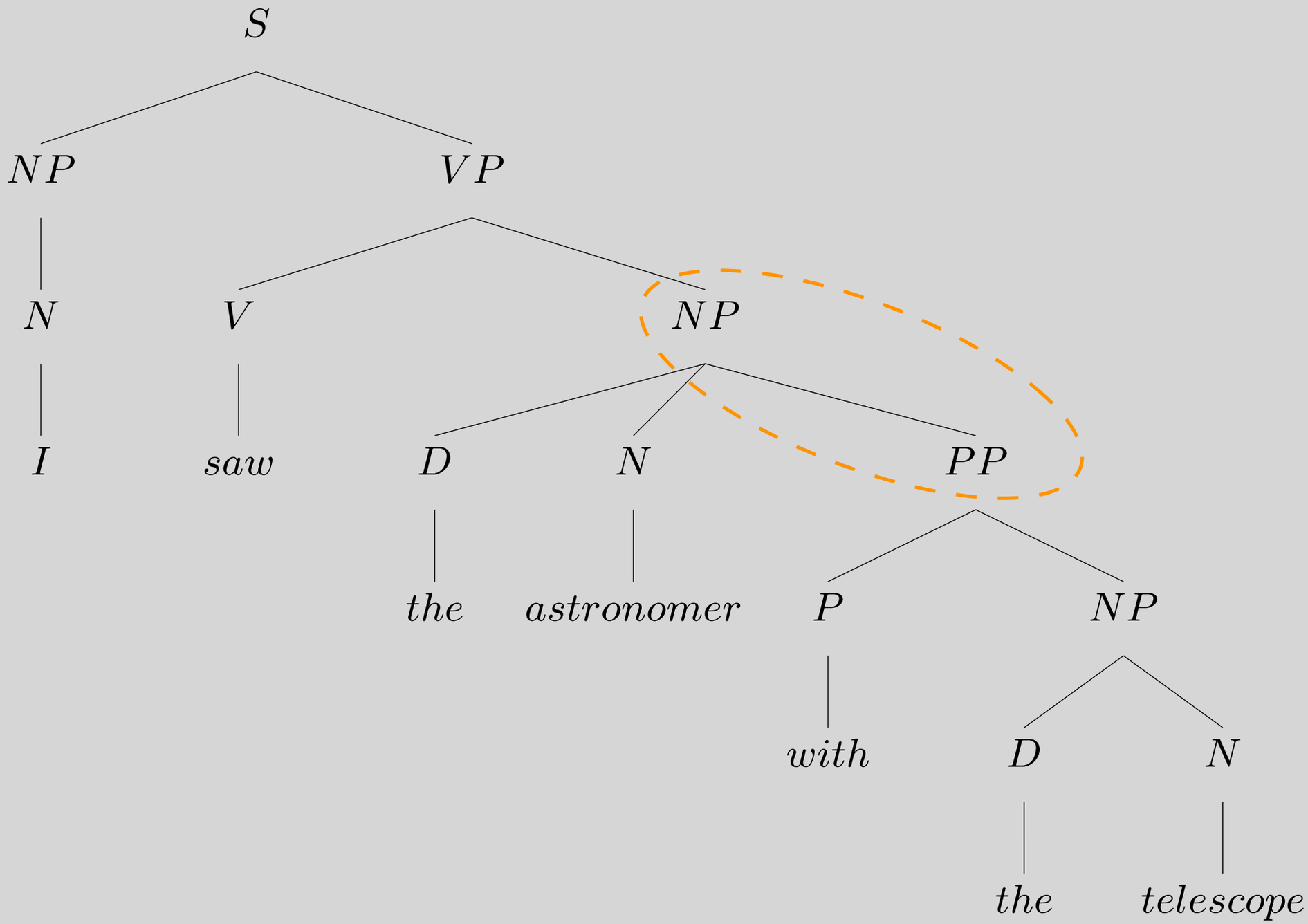
Structural Ambiguity

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

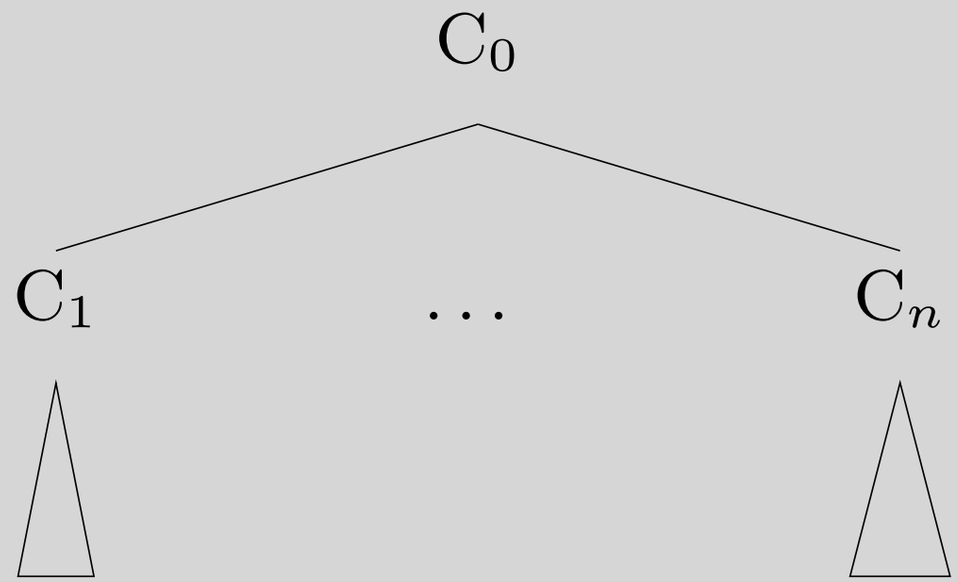
VP $\longrightarrow$ V NP

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



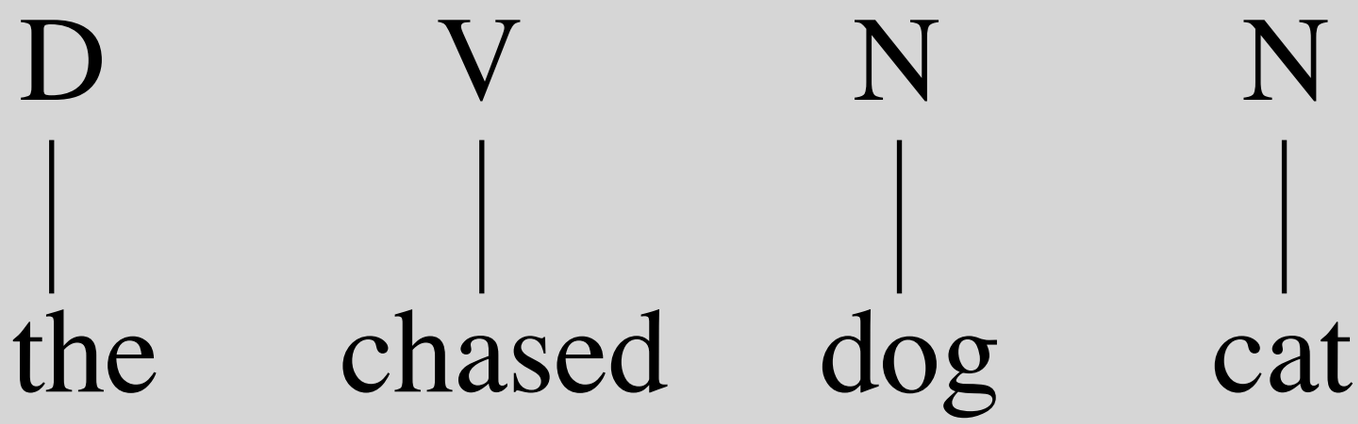
is a well-formed nonlexical tree if (and only if)

are well-formed trees, and

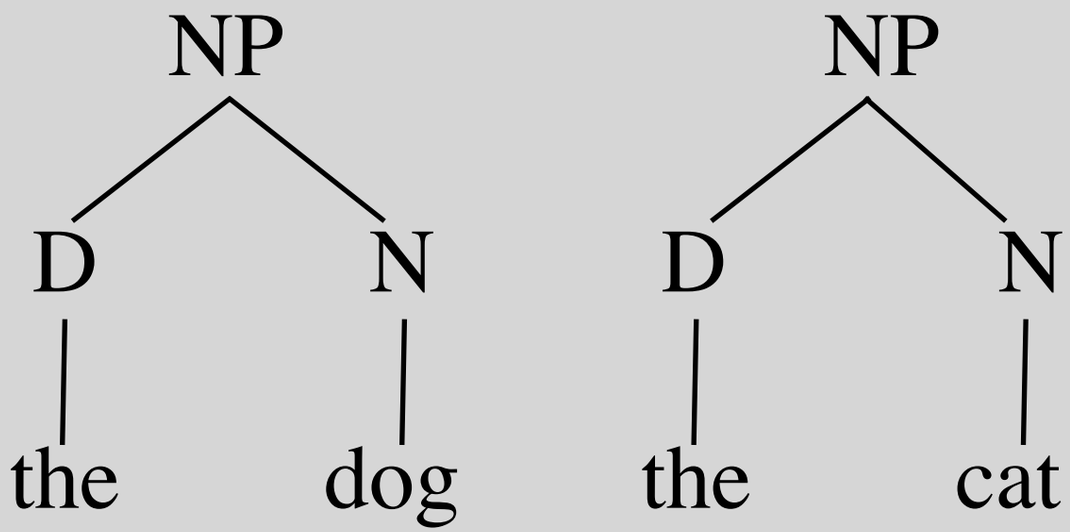
$C_0 \rightarrow C_1 \ldots Cn$ is a grammar rule.

Bottom-up Tree Construction

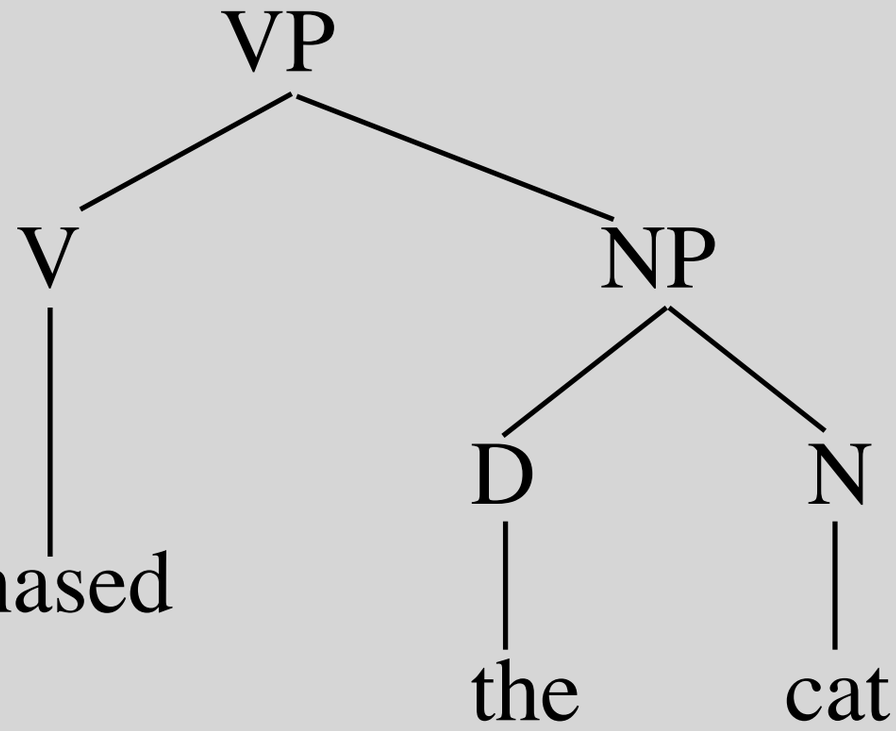
D: *the*
V: *chased*
N: *dog, cat*



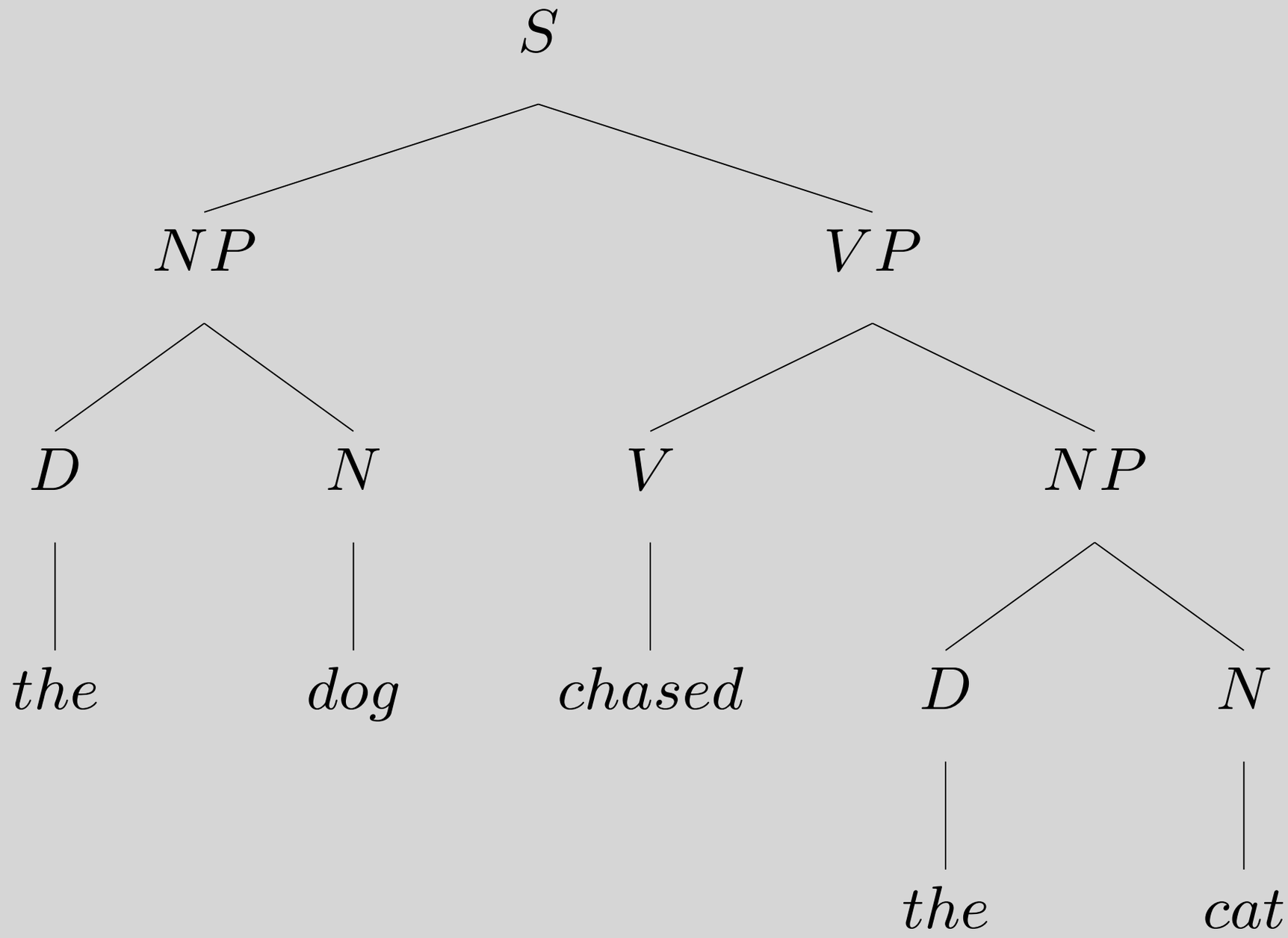
NP ⟶ D N



VP ⟶ V NP

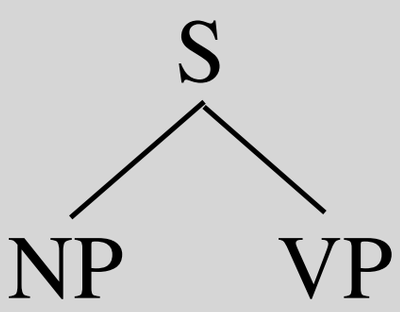


S $\longrightarrow$ NP VP

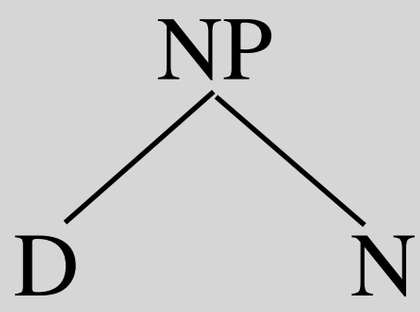


Top-down Tree Construction

$S \longrightarrow NP\ \ VP$

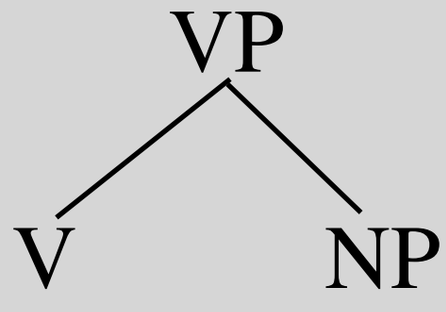


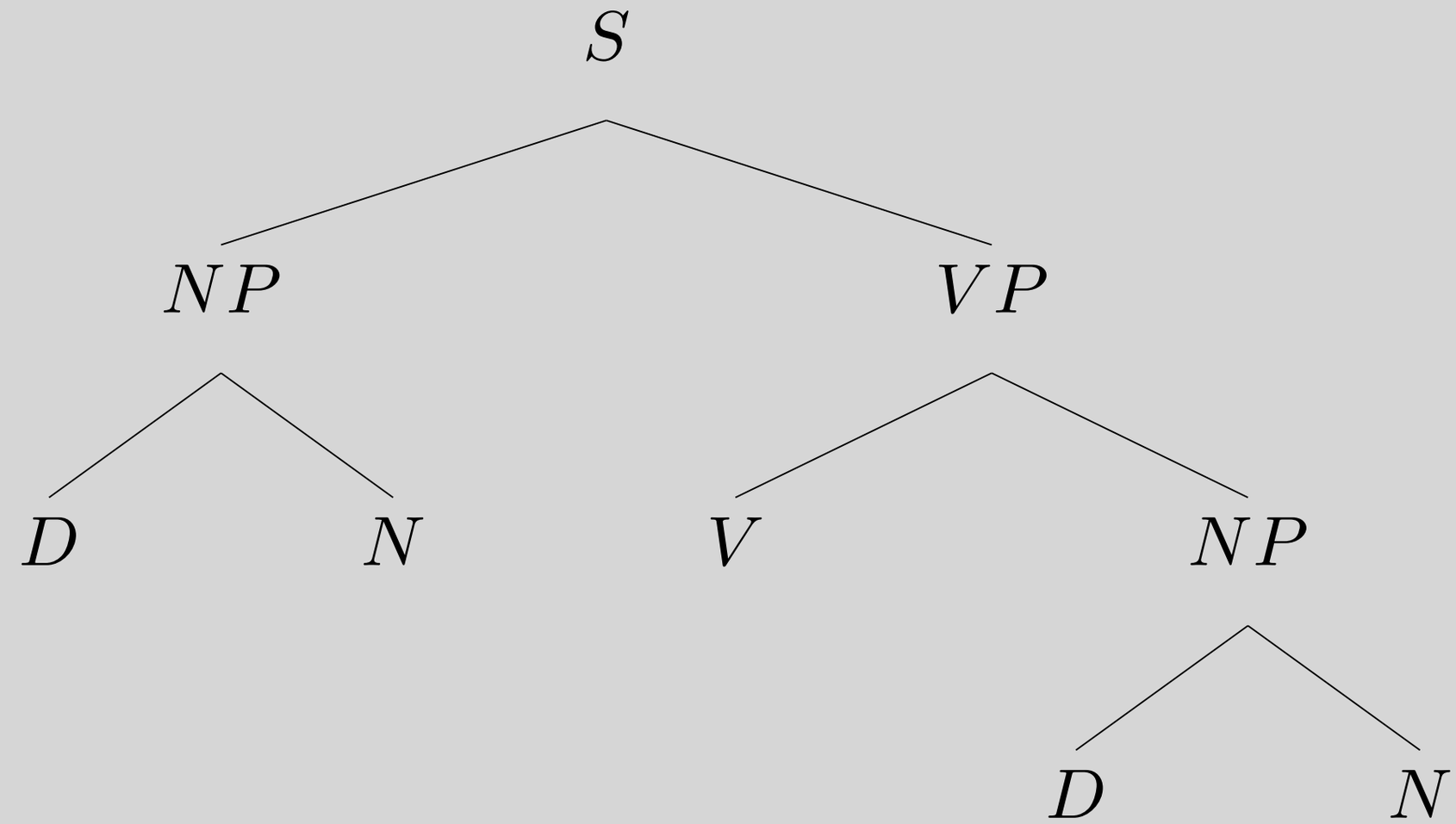
$NP \longrightarrow D\ \ N$



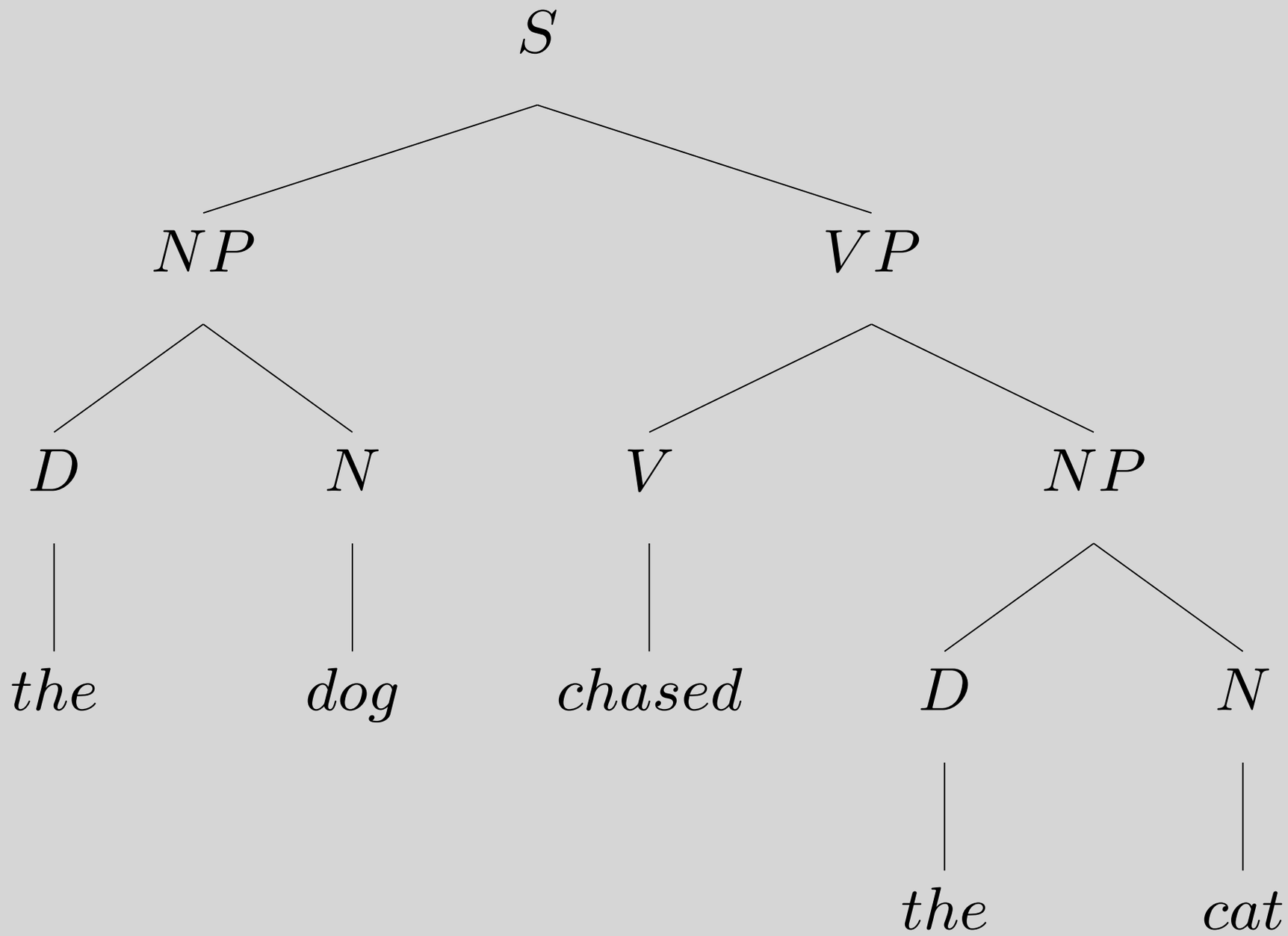
(twice)

$VP \longrightarrow V\ \ NP$





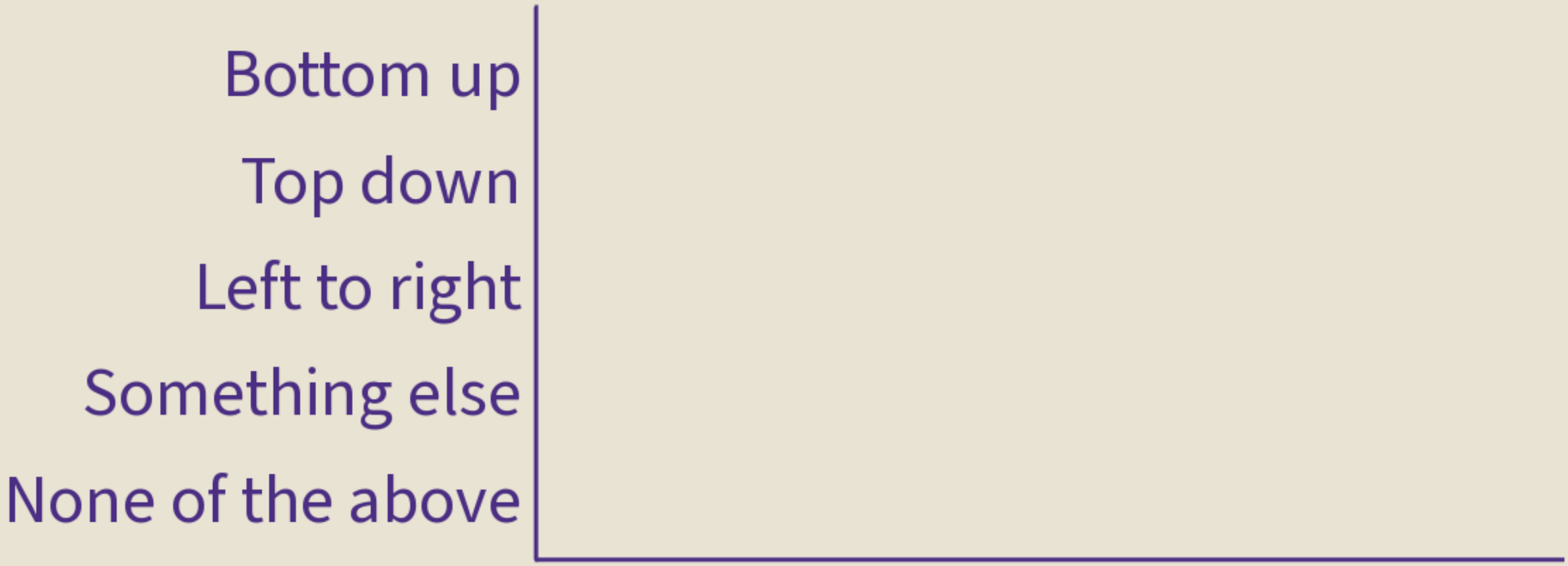
D	V	N	N
the	chased	dog	cat



When poll is active, respond at **pollev.com/emb**

Text **EMB** to **22333** once to join

For parsing, which style feels most intuitive to you?



Total Results: 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.
- 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

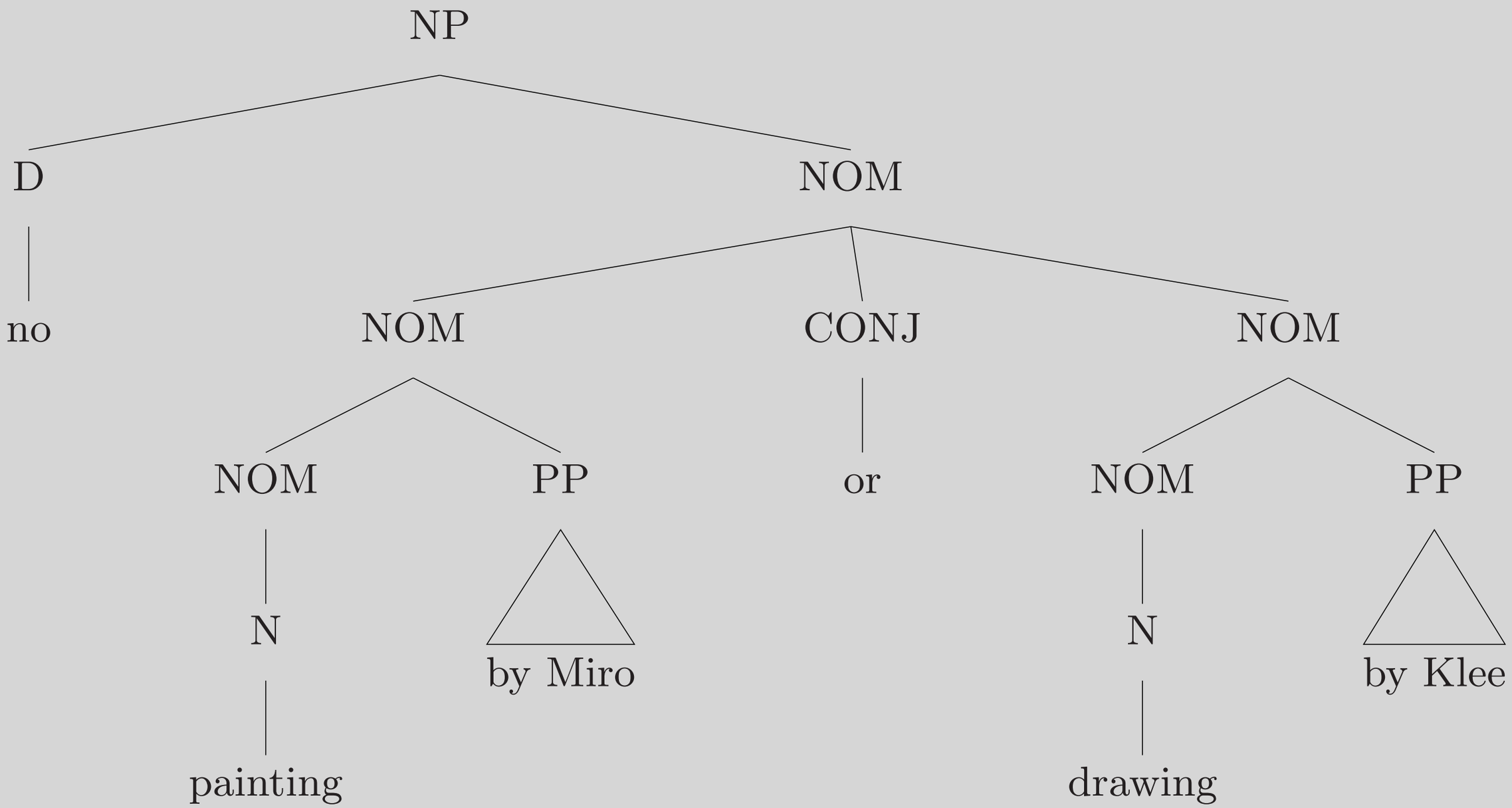
RQs: NOM

- Why can't we just collapse the rules containing NOM in (23) to NP -> (D) N (PP+)?
- (See poll next slide)

W

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

Total Results: 0



RQs: NOM

- Why can't we just collapse the rules containing NOM in (23) to NP -> (D) N (PP+)?
- Why do we have to distinguish N and NOM? What distinguishes them?

RQs: NOM

- I'm curious about the reason for introducing the concept of 'NOM' instead of having phrase structure rules like the ones below:
- DP → D NP
- NP → N
- NP → NP PP
- In this case, DPs are always headed by a D (whether null or not) and NPs by an N, which I think is a more generalizable approach compared to the rules involving NOM (as shown in (20), p.31). I don't understand the primary motive behind creating this new category that differs from regular phrases. Is it to represent nominal adjuncts and complements through different structures?

RQs: NOM

- This may be quite a simple question or even a non-question, but I noticed the analysis of category D as an optional part of category NP. This contrasts with my experience in my undergraduate syntax class, where category D was analyzed as the head of category DP, where the D slot was posited to be occupied by ø in cases where no determiner was present. I was wondering if there was a reason for choosing the NP analysis over the DP one or if that is simply a stylistic choice to be made when forming a theory.

RQs: NOM

- Why are the only predicates of the NOM rules "NOM → N" and "NOM → NOM PP" ? And likewise, why is NOM only a child of D?
- The examples show why this is a useful abstraction to establish what scope different structures lie inside, but I'm not sure why you couldn't do something like "NOM → VP" or "NOM → NOM VP" or other rule predicates.
- I think I must be misunderstanding when to use NOM; is it only if you have several items at the same level of the tree? It seems like it could be applied in a lot of scenarios.

RQs: NOM

- While reading our "current rules" in our grammar, I became curious as to why we need both NOM -> N and NOM -> NOM PP. Is there a reason why we cannot simplify that into NOM -> N|NOM (PP)?

RQs: NOM

- I am wondering how to treat cases like NOM prefixed with more than one determiner, like *both my sisters*, *your every move*, *the first thing* or *my other cat*.

RQs: NOM

- Right after example (25), the book got into a caution advisory and claims that "NOM cannot appear at the beginning of a sentence/cleft". What are the reasons for this? One explanation I can think of involves the NOM's place of occurrences from the summary grammar rules, and if we put this in a cleft sentence's tree structure using Tense Phrases (TP), putting NOM in a position it does not belong to would break the TP rule for the predicate.

- NOM can't appear in a cleft, so what about:

- *It is paintings by Klee and drawings by Miro that the museum displays*

RQs: Details of CFG

- Is it necessary for everything on the right side of the arrow, excluding what's in parentheses, to be present for that specific phrasal category?

RQs: Arbitrariness of categories

- Why can't a PP be directly under a Sentence (S → PP)? Are they always directly attached to an NP? Doesn't this sentence: At noon, the students gathered under the tree. have the PP separate from the NP, "the students"?

- 2.7.1, "if our theory of natural language syntax were nothing more than CFG, our theory would fail to predict the fact that certain kinds of CF rules are much more natural than others." What does "much more natural than others" mean here? Is there a spectrum of correctedness?

RQs: Other theories

- Beginning with (19), I'm confused why some of the CFG rules have the same phrasal categories on both sides of the arrow (such as VP > VP PP). In my experience, I remember the phrase type was made up of the head and any complements or adjuncts. You would only see XP generate the same XP in a tree if there was coordination or perhaps an appositive modifying a noun in which case an NP could generate another NP. I'm curious how a tree diagram would represent this. (Similarly, the tree in (22) shows the NOM generating N and PP but the rule in (23) states that NOM > NOM PP. What is going on here?)

RQs: Other theories

- My undergrad syntax instructors told us that each mother node in a syntactic tree can only have 2 daughters. Throughout Ch.2 there are many syntactic trees with nodes that have 3 or more daughters. Does this difference result from different theories of syntax? What is the benefit of one approach over another?

RQs: Transformational Grammar

- - How are transformations delineated within a transformational grammar? Do they use similar mappings as shown in a context free grammar?

- - To avoid the redundancy shown in figure (36), would it be possible to somehow map the agreement rules out separately? Instead of mapping out separate rules for each instance of a word type (e.g. NOM-SG, NOM-PL), would there be a way to indicate within the grammar how to delineate singular and plural agreement that could then be applied to the whole word type? (Is that, in effect, what a transformational grammar is?)

- - The text also says that the grammar we will be using is non-transformational. In that case, will there be a need to delineate these singular and plural agreements within our work?

RQs: Transformational Grammar

- What is the base 'unit' in a syntactic tree? It doesn't always seem to correspond to what we traditionally consider as words. When we have transformational features like tense and number, is it the feature itself that is part of the lexicon?

RQs: Transformational Grammar

- Pg. 41 states that "In a transformational grammar, then, each sentence is associated not with a single tree structure, but with a sequence of such structures." Does that mean that not all of these structures have to be well-formed then? Or this just mean that it's a series of transformations that leads to well-formed structures such as agreement transformation? If so, how do we properly label them?

RQs: Transformational Grammar

- On page 35 the book states "In the early 1960s, several scholars published arguments purporting to show that natural languages exhibit properties beyond the descriptive capacity of CFGs. The pioneering work in the first two decades of generative grammar was based on the assumption that these arguments were sound." To me, this implies that the pioneering work on generative grammar took place after the theory of CFGs, which confused me because I thought CFG was a type of generative grammar?

RQs: Compound, morphology

- Are words always considered atomic units of grammatical categories?
- Is punctuation not considered in either of non-lexical or lexical categories because it's more prescriptive rule?
- How are grammar trees designed for agglutinative languages? Do the words need to be broken up to be matched to leaf nodes, or are there more categories for each word type?

RQs: Compound, morphology

- How would we modify the grammar to add noun-noun compounds?
- In cases of hyphenated word phrases, if the resulting phrase is composed of words from different types, how can we modify the grammar rules to be compatible with these cases. For example, *a text-based solution*, *a once-quiet place*.

RQs: Free word order

- How would CFG look for languages with free word order, such as Ukrainian or Russian. Would it simply require many transformations to account for things such as gender, emphasis derived from word order, noun cases, etc? On page 42, it then talks about example (45) and how agreement cannot be determined until after a transformation has been made to decide which NP is the subject. As sentences become more complex in the languages previously mentioned, do certain transformations become prioritized over others?

RQs: Headedness

- Does 'headedness' occur in all human languages?
- The book describes a head-driven grammar as having "lexical heads" of "syntactic phrases." However, our rules allow for heads to be a phrase as well, as is clear from VP -> VP PP where I assume VP is the head of VP. Is this true? Or will we later discover that the head must be a lexical item and not a phrase?

RQs: Headedness

- The textbook mentions that the head is the lexical category that a phrasal category derives its name from, and that phrasal categories were named after the lexical category that is an obligatory part of that phrase (e.g. N for NP, since D is optional). This implies that the definition for a head is "the obligatory lexical category of a phrasal category". I was wondering how "obligatory" should be interpreted under this context, since, for example, the phrasal category PP → P NP would require both P and N (derived from NP) in order to be complete, how do we determine which one is more "obligatory" to select its head? Or is it that only immediate lexical categories are considered for this analysis?

RQs: Well-formedness

- How should we handle choosing whose judgments are well formed? If we were to say, take a vote, or choose a "standard" form of a given language, at what point does that become prescriptive? Could we consider a judgment to be descriptive for one speaker while prescriptive for another? Or do we need to establish a variety of a given language within a sociological group before we view the judgments that way? Are there other criteria I'm not considering here?

RQs: Process-neutrality

- I understand direction neutrality in a tree, but I'm a bit confused on the idea of process neutrality. How can a person think of "constraining the set of all phrase structure trees" without thinking about how sentences are generated in the language? Is that not the first step in thinking if things in a language are well-formed?

- I'm still confused as to what 'well-formed' means. If a tree was not 'well-formed' would it not have direction and process neutrality?

RQs: Chomsky Hierarchy

- “Because phrasal categories can appear on the right-hand sides of rules, it is possible to have phrases embedded within other phrases. This permits CFGs to express regularities that seem like accidents when only regular expressions are permitted.” Until this part, I had taken "regular expression" as a given term. What, specifically, does expressing regularities mean?

RQs: Chomsky Hierarchy

- What is the difference between a lexical category, nonlexical category/phrasal category, and regular expression? Section 2.3 reads, "To distinguish the two types, we will sometimes use the terms' lexical category' (for parts of speech) and 'nonlexical category' or 'phrasal category' to mean types of phrase." Does this mean that phrases are the set of rules, i.e., S -> NP VP, and lexical categories are the words, i.e., dog, run? Are NP and VP also called phrase categories, or are they regular expressions?

RQs: Chomsky Hierarchy

- In terms of the memory capabilities of an FSA and a PDA, what properties do phrases require to be processed? Or: what can the PDA (CFG) do that an FSA (regular expression) can't do in order to process phrases?

RQs: Chomsky Hierarchy

- CFG is referred as a generative grammar. However, when generating a sentence, we are not only care about the part-of-speech, the grammar, but also the meaning. As it is said by Chomsky, "Colorless green ideas sleep furiously." this sentence is grammatically correct but semantically wrong, so it can't be judged as acceptable by our intuitive. Then should we consider the rules used to generate this invalid?

RQs: Syntax v. Grammar

- Since HPSG winds up stuffing things like semantic senses of words and relationships between words in sentences in its grammars (iirc), where is the line between syntax and grammar in HPSG parlance? It seems like you encode all the syntax rules for a language or dialect into a grammar, so is syntax just the more general study of how we go about representing things (as grammars)?

RQs: Bigger Picture

- What is considered the best grammar of English? Is there anything universally accepted? Does it depend on the goal of the grammar? Do different linguistic schools of thought produce different kinds of grammars?

RQs: Bigger Picture

- We have seen that CFG is not a complete or perfect representation of English, but are there practical applications for CFGs? Maybe the grammar generates sentences that are not well-formed, but it if it accepts sentences that are well-formed, maybe that's good enough for some use cases.

RQs: Bigger Picture

- I am wondering whether these shortcomings can be overcome by sufficient computational power. For example, even though CFGs fail to capture "headedness" (and maybe some other characteristics of natural languages), will we be able to compute these characteristics on the fly during analysis? Even though massive redundancy is required in order to analyze the agreement and subcategorization patterns of natural languages, will we be able to hand over this redundancy handling to algorithms and only work on the level we care about during analysis? Or are these shortcomings of CFGs fundamental flaws that cannot be overcome without introducing extensions to CFGs?

RQs: Bigger Picture

- Within the computational context, how do language models solve the agreement using its grammar?