



# Ling 566

## Oct 5, 2021

### Context-Free Grammar

# Self-intro

- Please call me Emily
  - But Dr./Prof. Bender is okay
- Pronouns she/her

# 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?

# Reading Questions

- “A syntactic theory that sheds light on human linguistic abilities ought to explain why such patterns do not occur in human languages. But a theory that said that grammars consisted only of lists of sentences could not do that.” - how does CFG explain various patterns not appearing in human languages, then? CFG also just presents some patterns for languages, right?

🌐 When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



# **W** "Patterns that exist in human languages" here refers to:

Word orders in a single sentence

Patterns of possible sentences

Something else

# 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  
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

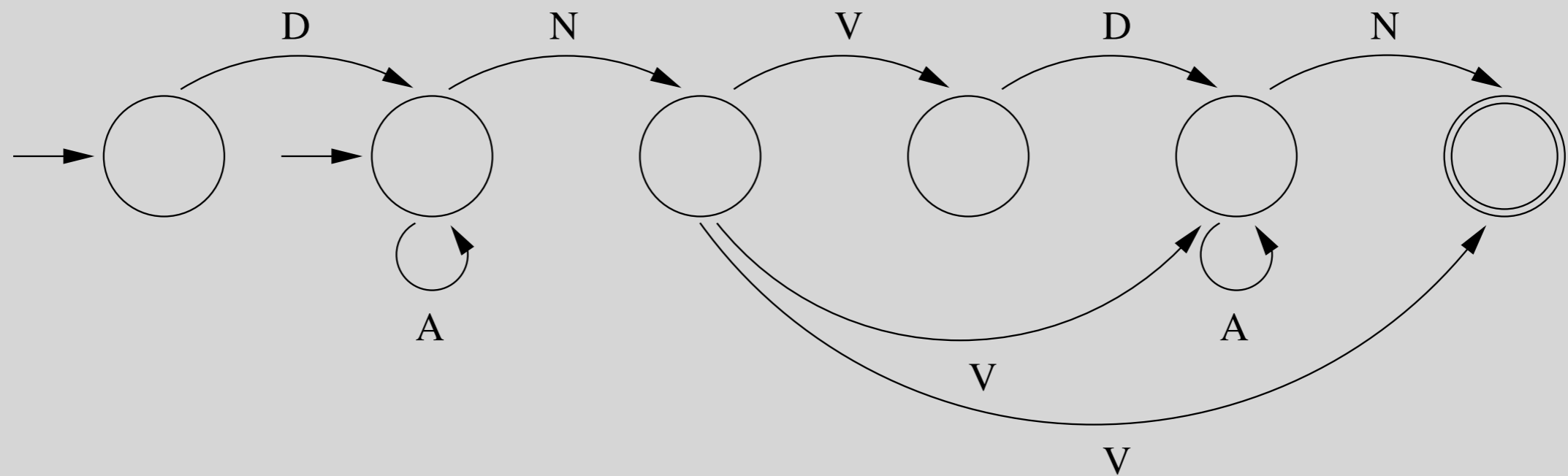
- $a^* = \{\emptyset, a, aa, aaa, aaaa, \dots\}$

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

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



# A Finite State Machine



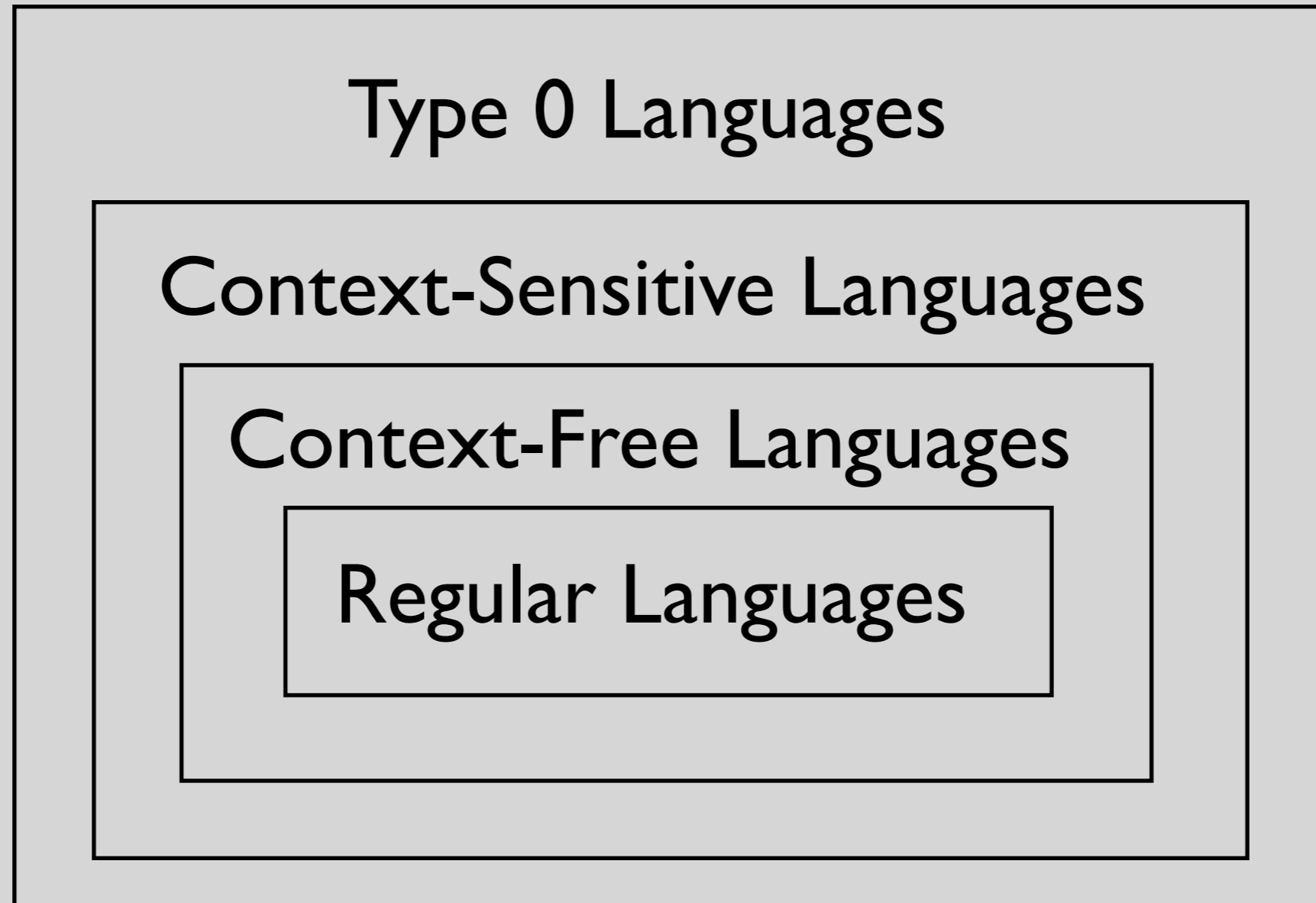
# What does a theory do?

- 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



# Context-Free Grammar

- A quadruple:  $\langle C, \Sigma, P, S \rangle$ 
  - $C$ : set of categories
  - $\Sigma$ : 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

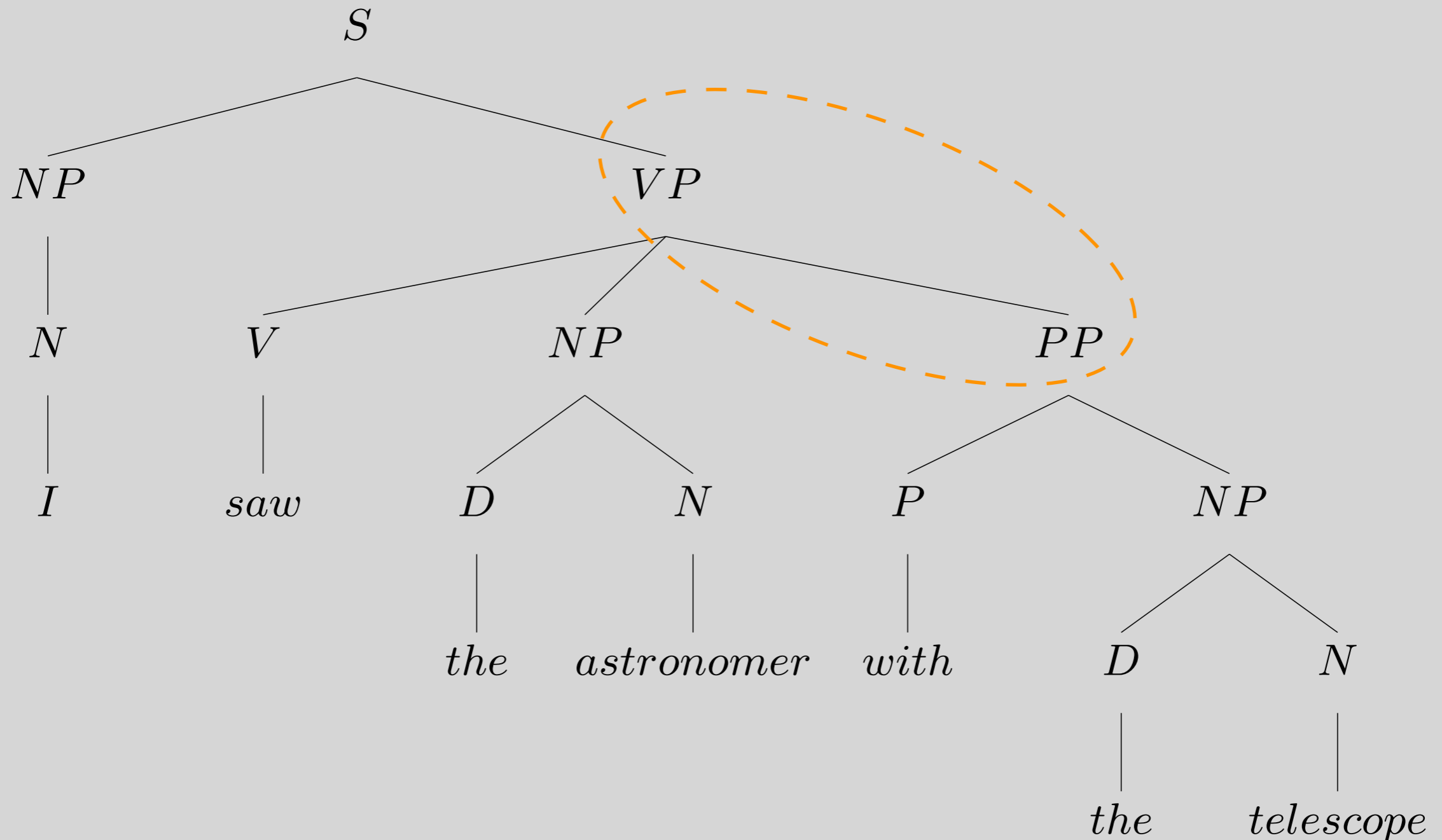
V: attack, ate, watched

P: for, beside, with

# Structural Ambiguity

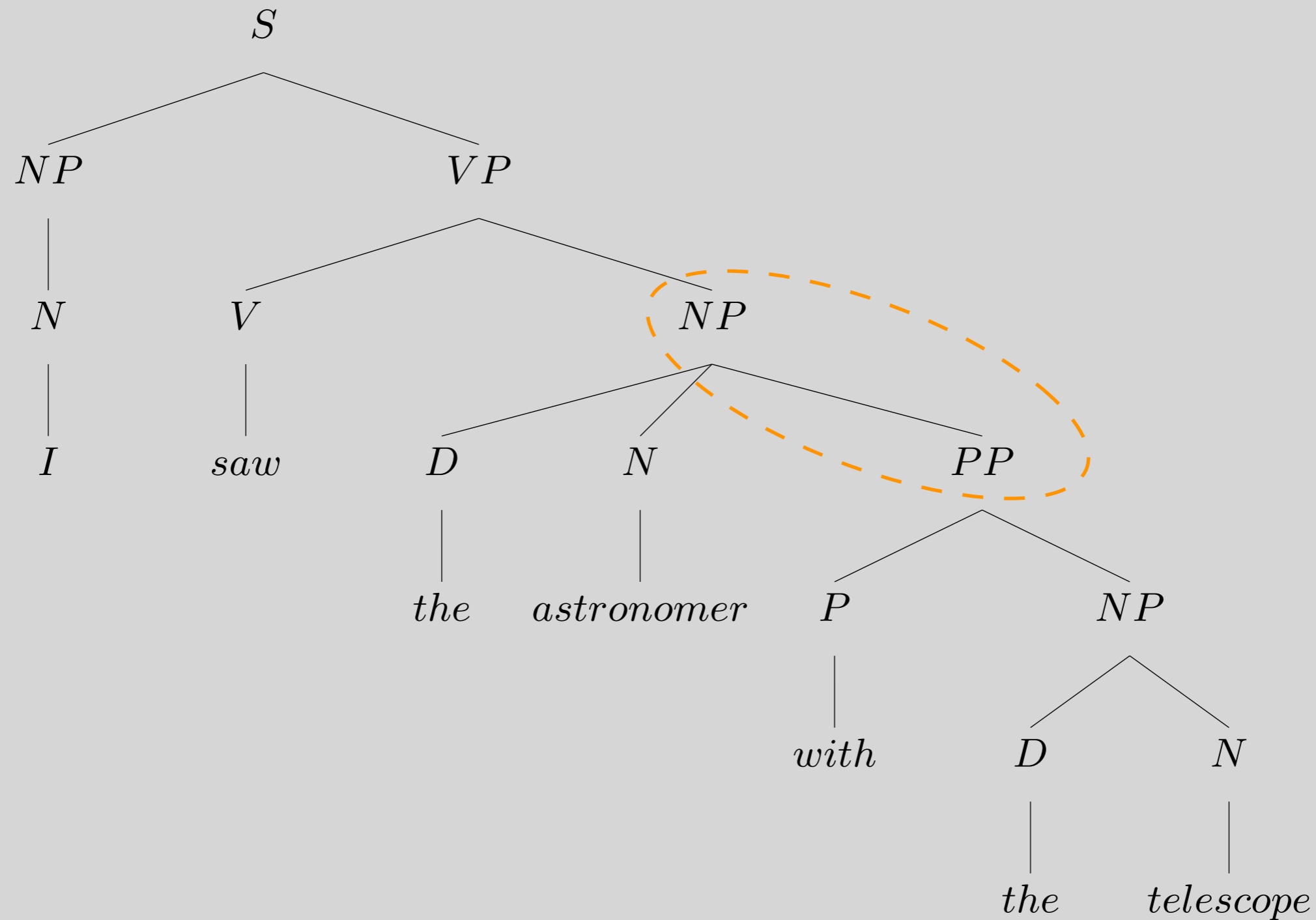
I saw the astronomer with the telescope.

# Structure 1: PP under VP





# Structure 1: PP under NP



# Reading Questions

- Why are ambiguities relevant for syntax?  
Isn't the function of syntax just grammaticality?

🌐 When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



## W Select all that you find true:

Rules of syntax determine permissible strings

Rules of syntax determine sentence meaning

Structural ambiguity (multiple trees) helps model semantic ambiguity

All ambiguities involve multiple trees

None of the above

# 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.*
- *They arrive this morning.*

# A Trivial CFG

$S \rightarrow NP VP$

$NP \rightarrow D N$

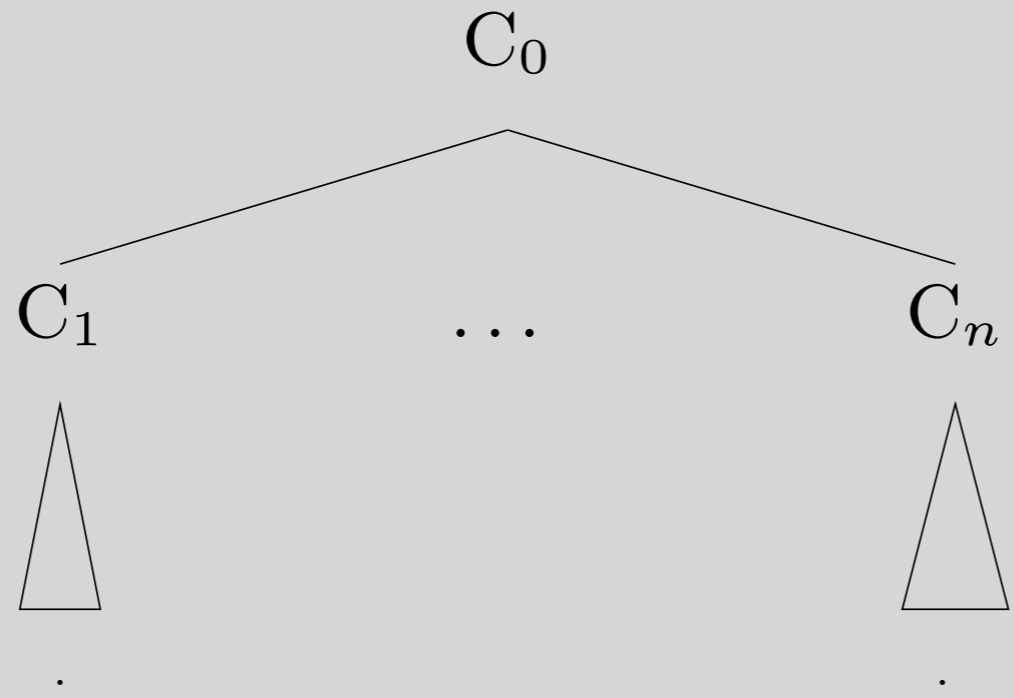
$VP \rightarrow V NP$

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

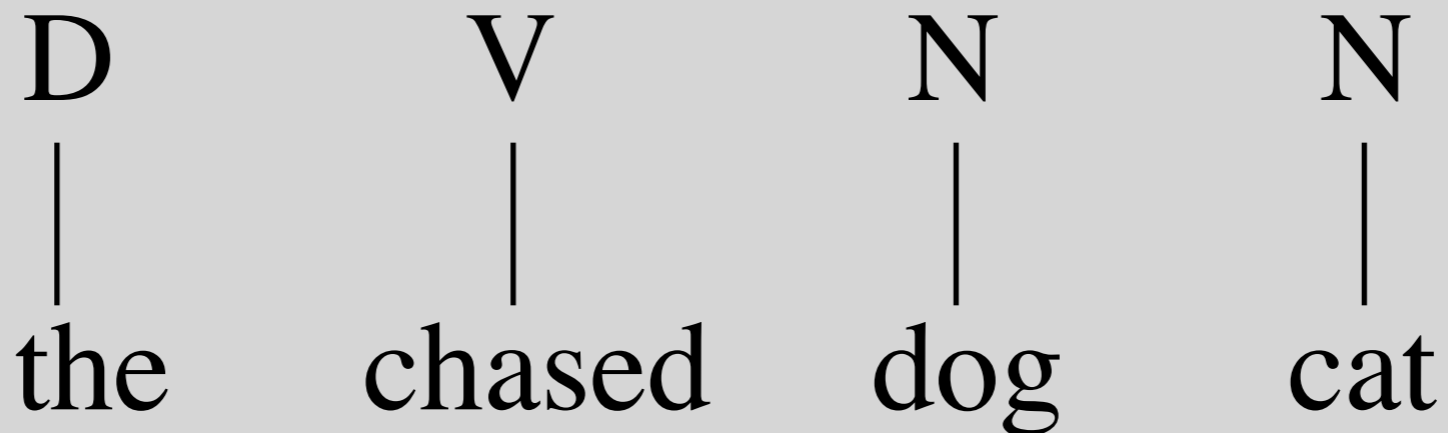
$C_0 \rightarrow C_1 \dots C_n$  is a grammar rule.

# Bottom-up Tree Construction

D: *the*

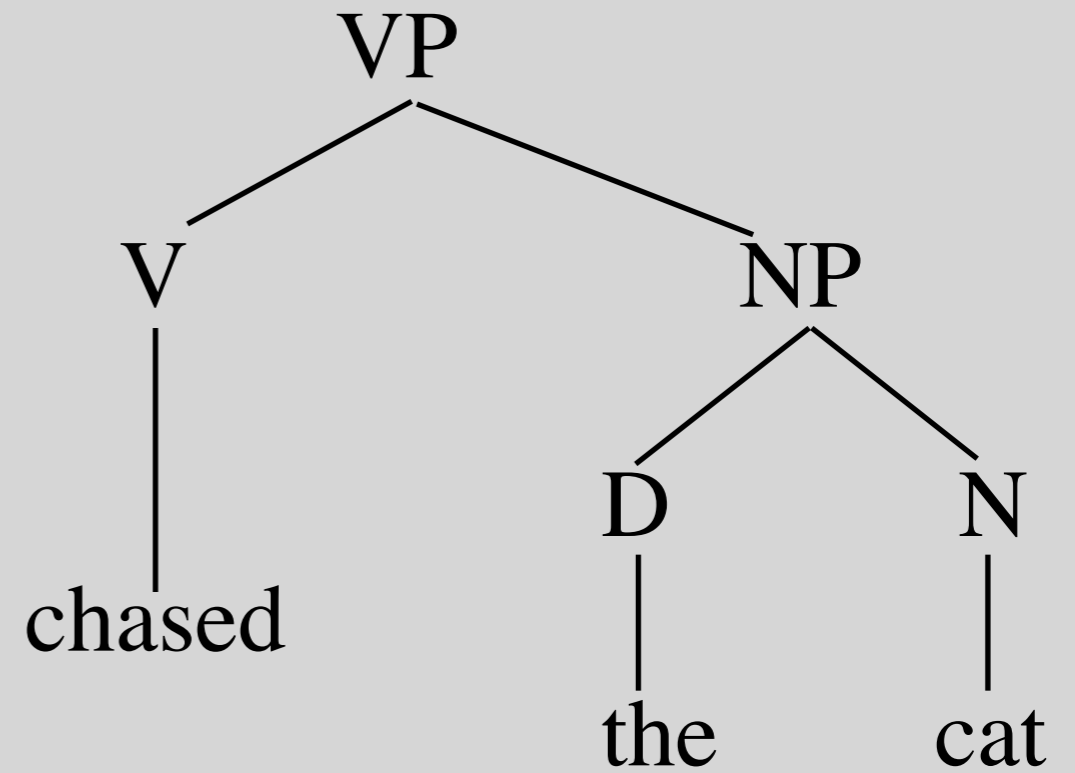
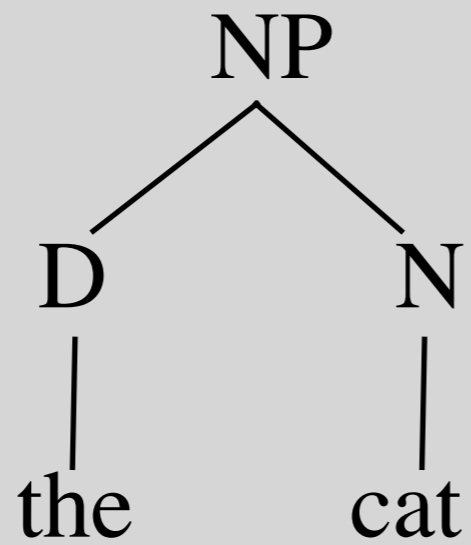
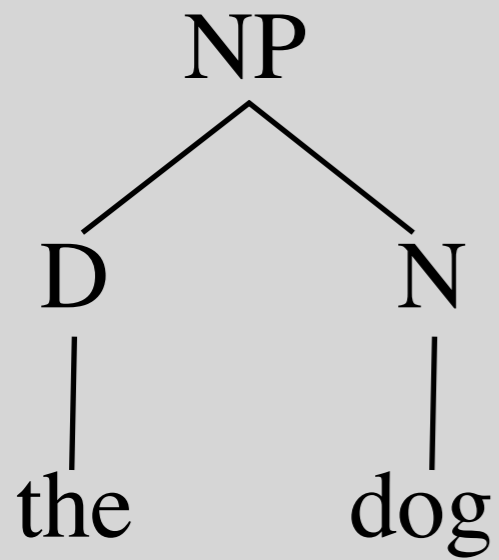
V: *chased*

N: *dog, cat*

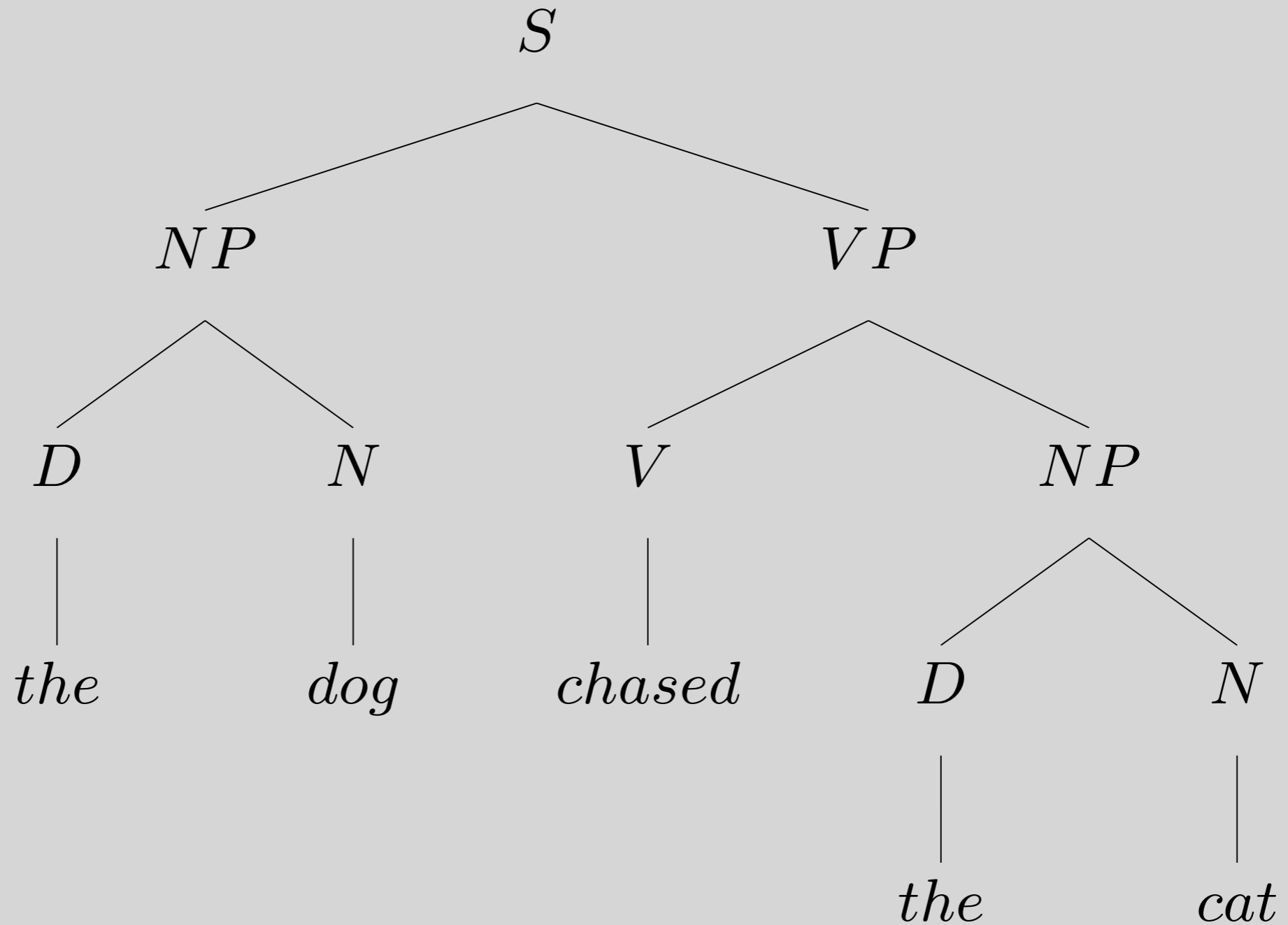


NP  $\rightarrow$  D N

VP  $\rightarrow$  V NP

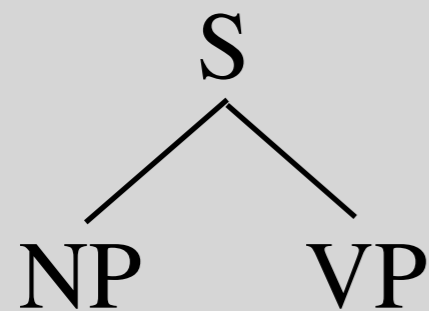


$S \rightarrow NP VP$

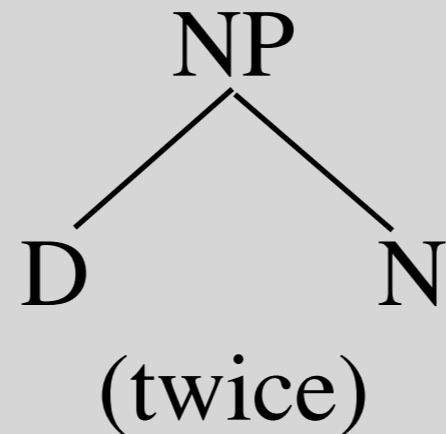


# Top-down Tree Construction

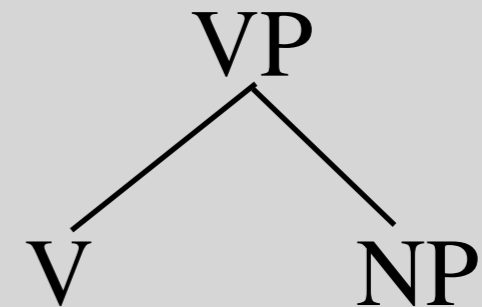
$S \longrightarrow NP VP$

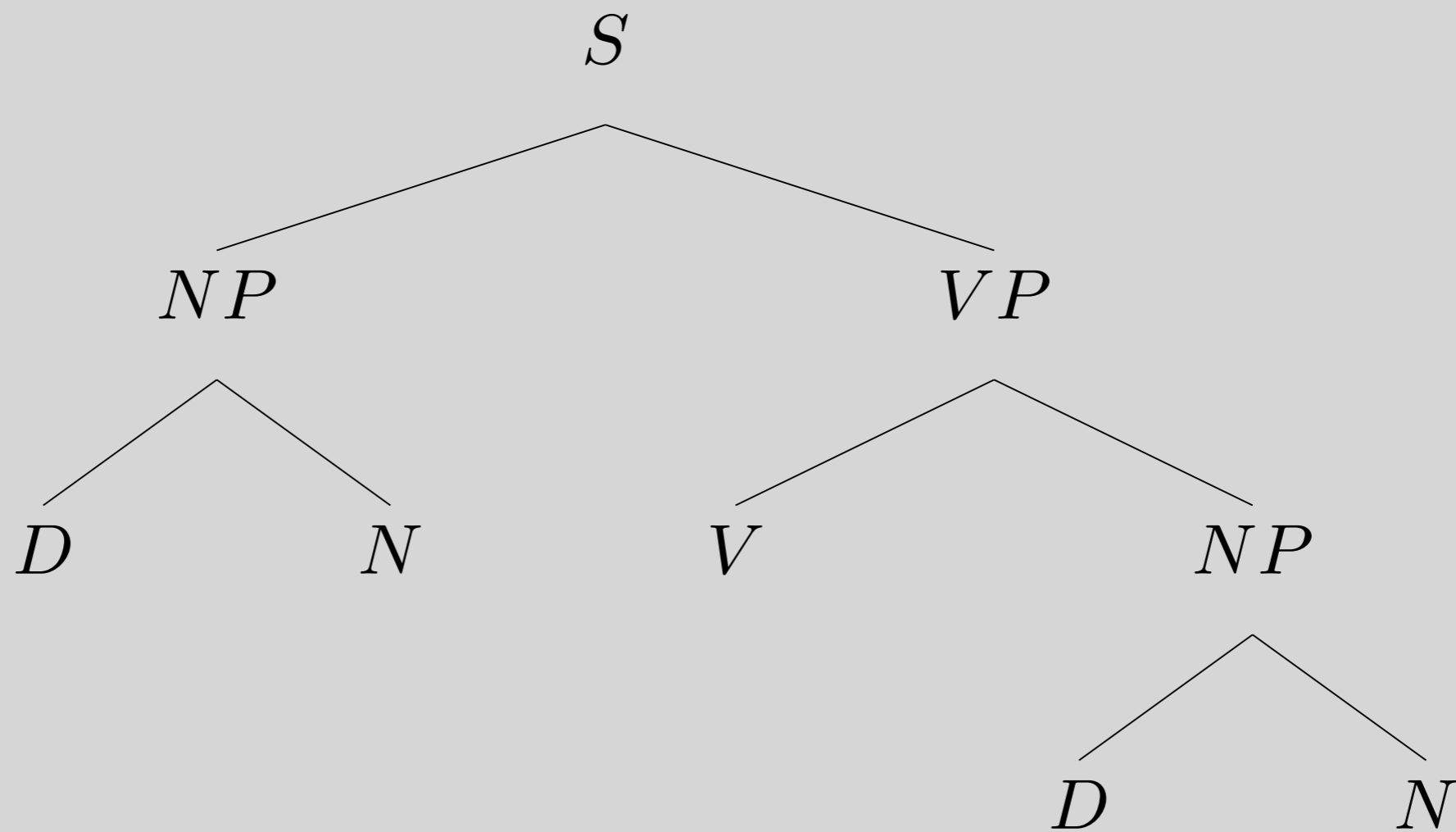


$NP \longrightarrow D N$



$VP \longrightarrow V NP$





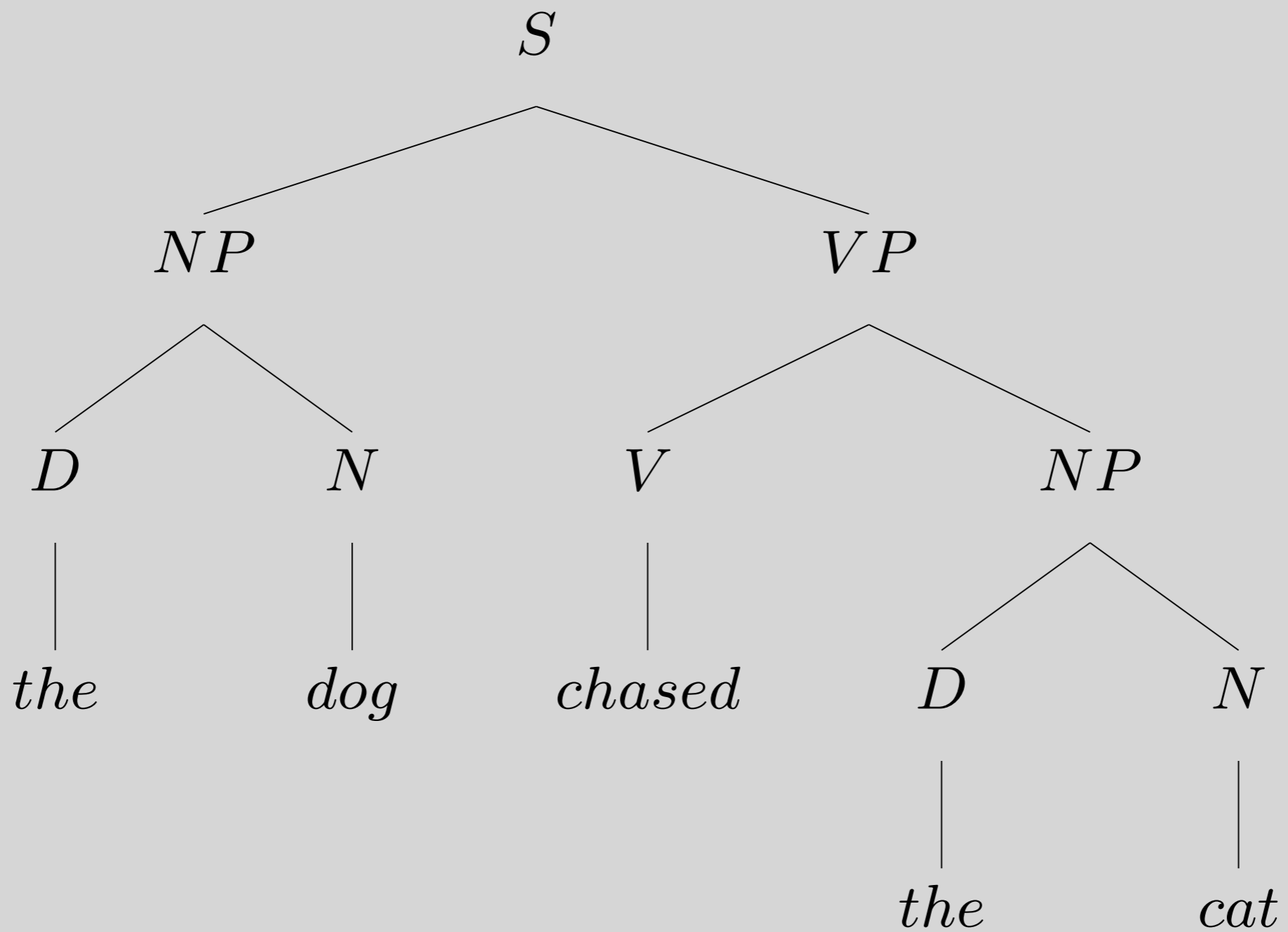


D  
|  
the

V  
|  
chased

N  
|  
dog

N  
|  
cat



🌐 When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



# **W** For parsing, which style feels most intuitive to you?

Bottom up

Top down

Left to right

Something else

None of the above

# 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 w/reg languages
- $\therefore f(\text{SG})$  (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 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

# Reading Questions

- Why do we need NOM?

🌐 When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



# **W** Which rules license "no no no dog in the house barked"? (Together with PP -> P NP)

NP -> D NOM;  
NOM -> NOM PP

NP -> NP PP

None of the  
above

When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



# Which rules license "no painting by A or drawing by B was displayed"? (Together with PP -> P NP)

NP -> NP PP

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

NP -> (D) N (PP)

None of the above

# Reading Questions

- Why is coordination a constituency test?
- What are the implications of constituency tests on syntactic models?

# Reading Questions

- How can CFG be used to make a grammaticality judgement when encountering a word has not been listed in the lexicon?
- Do the existing grammar rules cover all sentence structures that do not exist yet but might one day? Are we able to be sure of this? Are they mostly concerned with language that is not casual or colloquial? Who decides to formalize these rules, and how do they choose which ones to formalize? How would we update these rules?



# Reading Questions

- How do we account for changes in language over time, using CFG as instruments to describe grammar? Are there any measures to hint at what might be a good predictor of which rules might never be violated and which might be violable (prone to evolution of language)? For eg: are there qualities such as some head priority or depth of phrase in the parse tree which might be helpful in predicting violability of rules?

# Reading Questions

- Towards the end of Chapter 2, the issue that no two speakers have the same linguistic knowledge is brought up. I have been introduced to various constraints that account for issues such as those concerning ambiguity, anaphors, thematic relations, and more. However, I have yet to see constraints that could account for regional differences other than a footnote stating that some people might disagree on a given example sentence's acceptability. How can any generalized model of grammar be completely accurate if two different native speakers could look at the same sentence and disagree on whether or not it is acceptable?

# Reading Questions

- How can CFG deal with transitive/intransitive ambiguity?
- Why are transitive and intransitive verbs subcategories of something and not different categories?

# Reading Questions

- Can I say CFG (and other theories mentioned in this chapter) is a collection of prescriptive grammar in the sense that it defines certain rules of what is allowed (hence we know what is not allowed), only CFG is more abstract?

When poll is active, respond at [pollev.com/emb](https://pollev.com/emb)

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



# W Is CFG a type of prescriptive grammar?

Yes

No

Depends on what  
you're using it for

Not sure

# Reading Questions

- In transformational grammar, I didn't understand how a transformation would look exactly. Would it be a list of rules that move/replace certain branches in the tree? If yes, how would these rules look like?
- I'd like a bit more clarification about the different grammatical theories and in what ways they each influence what we'll be learning in the course.
- How has the existence of multiple relevant frameworks (HPSG, GPSG, TG, etc..) affected the development of linguistics? Are there other languages where linguistic studies are instead dominated by a singular, universally accepted framework?

# Reading Questions

- What kind of formalisms aren't order-independent?
- What is the state-of-the-art grammar check technology used as of today in the industry?  
Is it based off CFG?

# Reading Questions

- Are there any known examples of sentences in English language that require context-sensitive grammar to model them? Or conversely, is context free grammar sufficient to model English grammar?



# Reading Questions

- Why does syntax need to worry about agreement? Why isn't this just morphology?
- Why is number an intrinsic property of nouns, not verbs?

# Reading Questions

- It looks like from the trees in this chapter that ternary+ trees are permitted. What are the advantages and drawbacks of binary vs. ternary vs. and so on trees? Is it an elegance consideration, ease of implementation consideration, etc. ?

# Reading Questions

- On pg. 23: "a grammar motivated largely on the basis of considerations of parsimony seems to be a good candidate for a psychological model of the knowledge of language that is employed in speaking and understanding." I find the use of "parsimony" as basis of evaluating grammars interesting. Is it delegating too much importance on the ease of cognitive processing, rather than focusing on producing a theoretically sound framework of grammar?

# Reading Questions

- How do linguists handle punctuation?