Ling 566
Oct 6, 2020

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

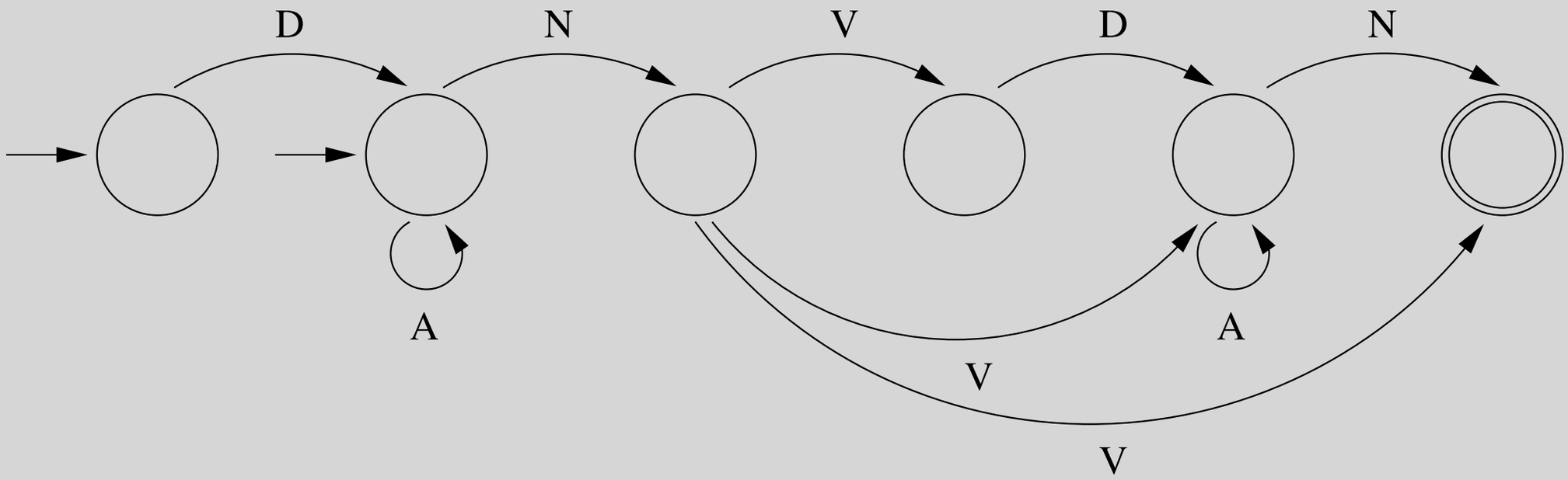
D A N V

- 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

Reading Questions

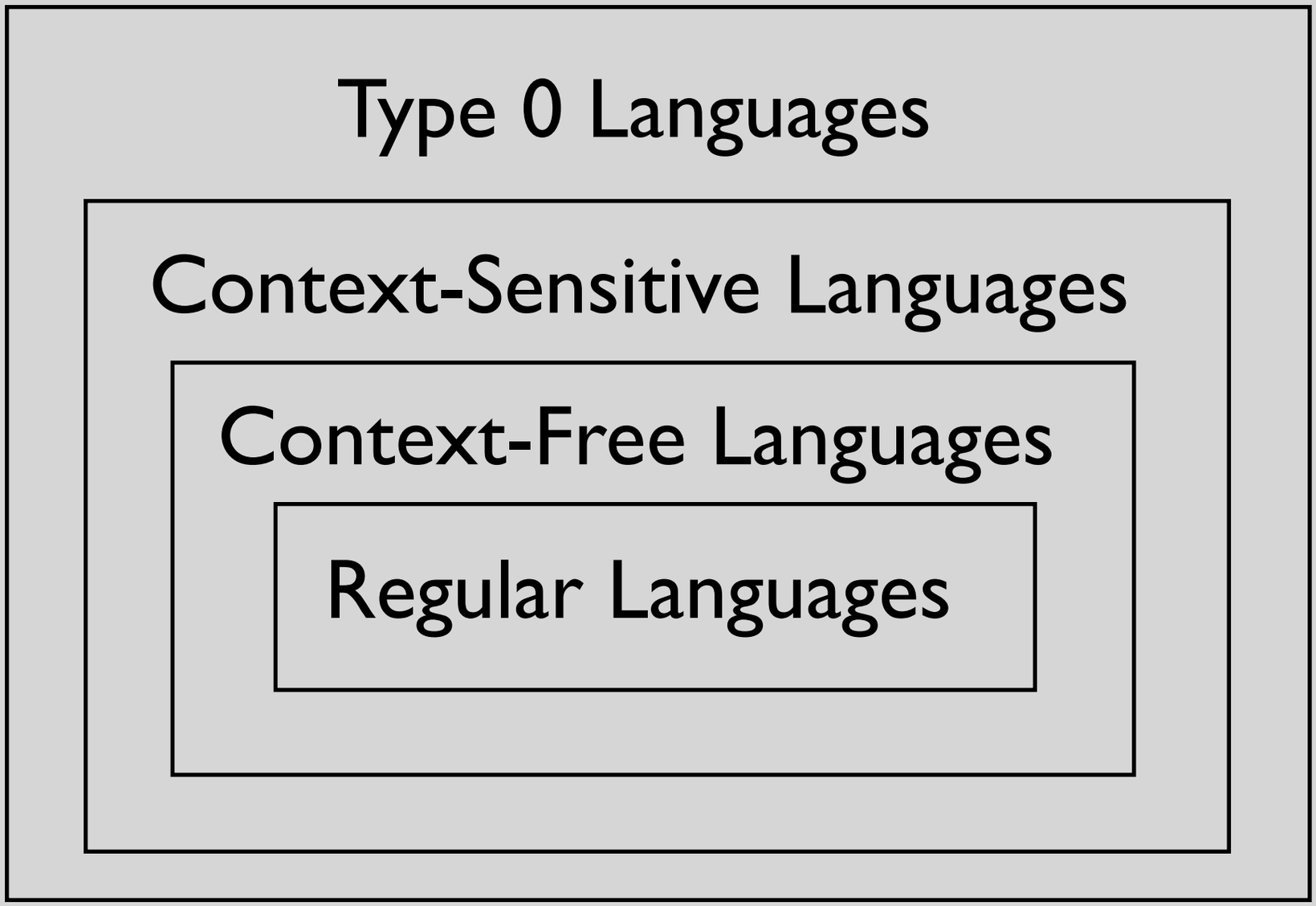
- It was stated that Chomsky and other linguists find the object of the study of syntax as the mental representation of individual speakers. It seems to me that trying to use syntax independently of semantics to understand individuals' mental representation is an ironic endeavor. This is because without understanding what the words mean, it's hard to determine if a syntactic expression is correct. e.g., "I am in deep waters." vs. " *I am on deep waters". Further, in many slangs people don't necessarily speak with "acceptable" syntax and yet they still manage to communicate their mental representation rather effectively.

- Is the framework that aims to get at the actual mental representation of linguistic knowledge necessarily limited to the individual? Isn't Chomsky's main point that we share a common mental representation? Is it understood here that the social context of language is located not in the mind but elsewhere, and therefore is left unconsidered when we look only at the mind? Is it located in the mind, but not in the language center?

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

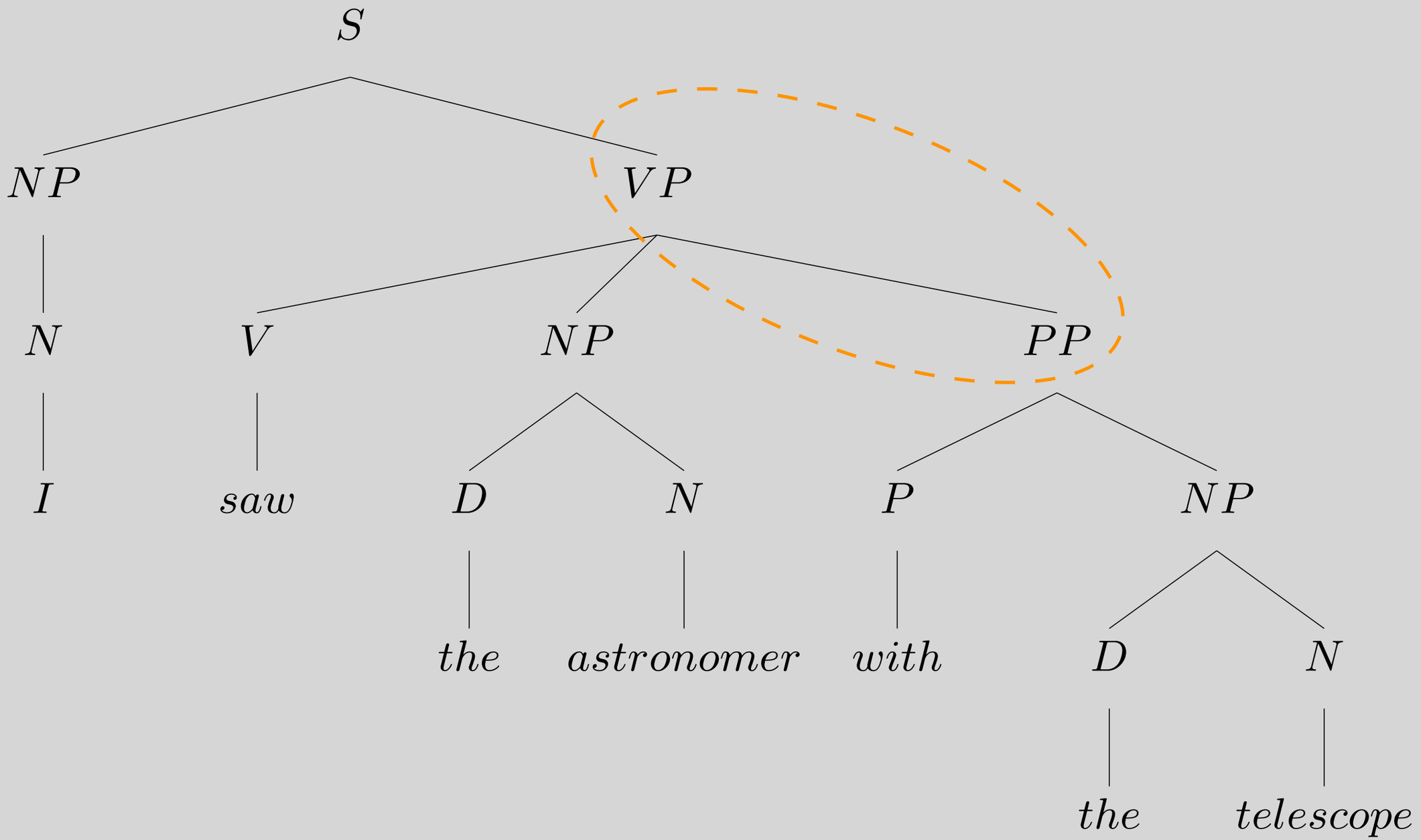
V: attack, ate, watched

P: for, beside, with

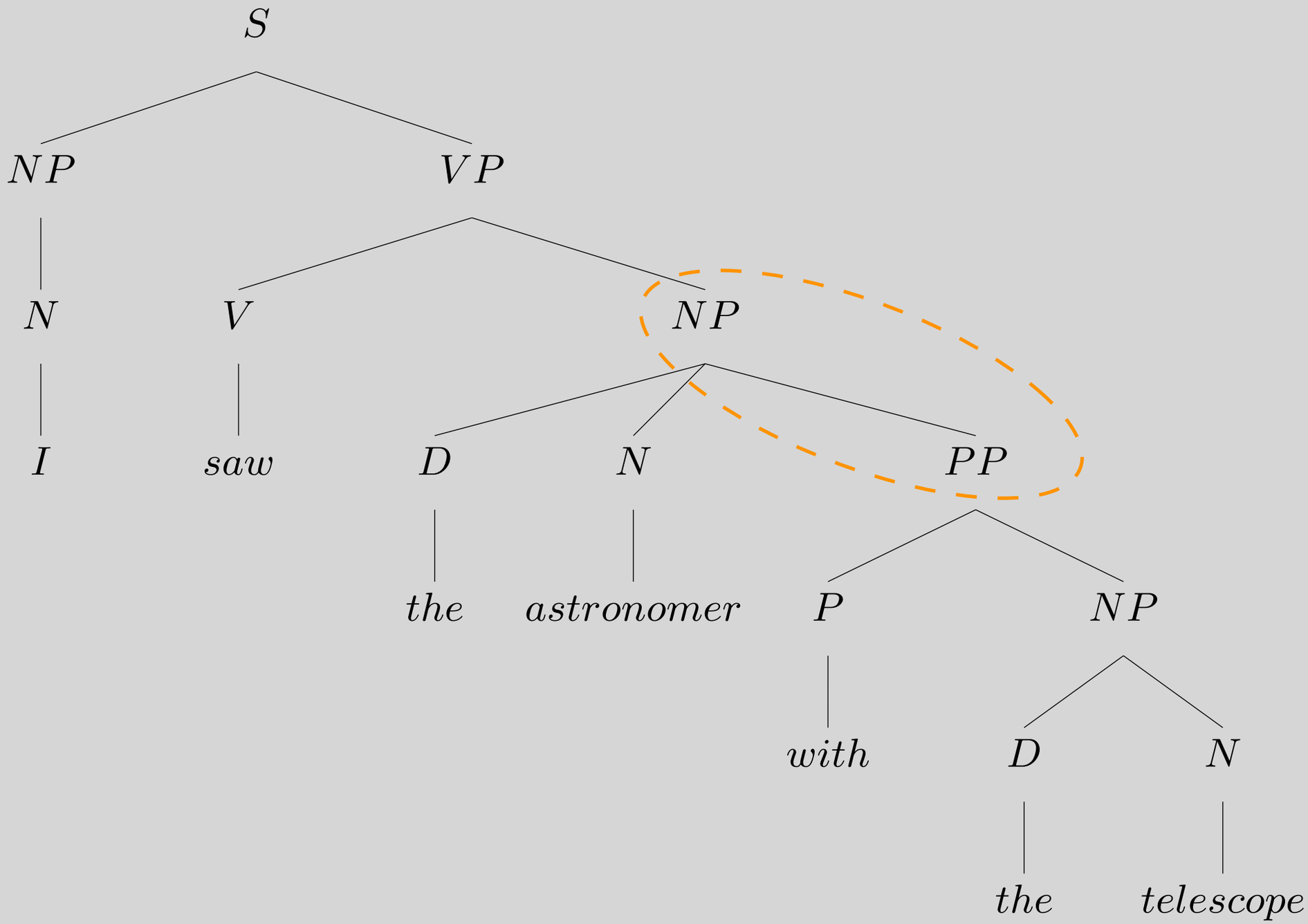
Structural Ambiguity

I saw the astronomer with the telescope.

Structure 1: PP under VP



Structure 1: PP under NP



Reading Questions

- My previous syntax classes have been very strict about trees being binary branching only. The trees in this chapter were sometimes more than binary branching -- will that continue to be the case as we build the grammar, and if so, why is that preferable to binary? How does that affect our definitions of constituency?

- I wonder how/whether word order plays a part in the architecture of a tree. Usually when we talk about trees like a decision tree, the direction in which a branch bifurcate doesn't really matter. But in the context of generating a sentence under the tree paradigm, it seems like there is an implicit assumption about word order being embedded in a tree.

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

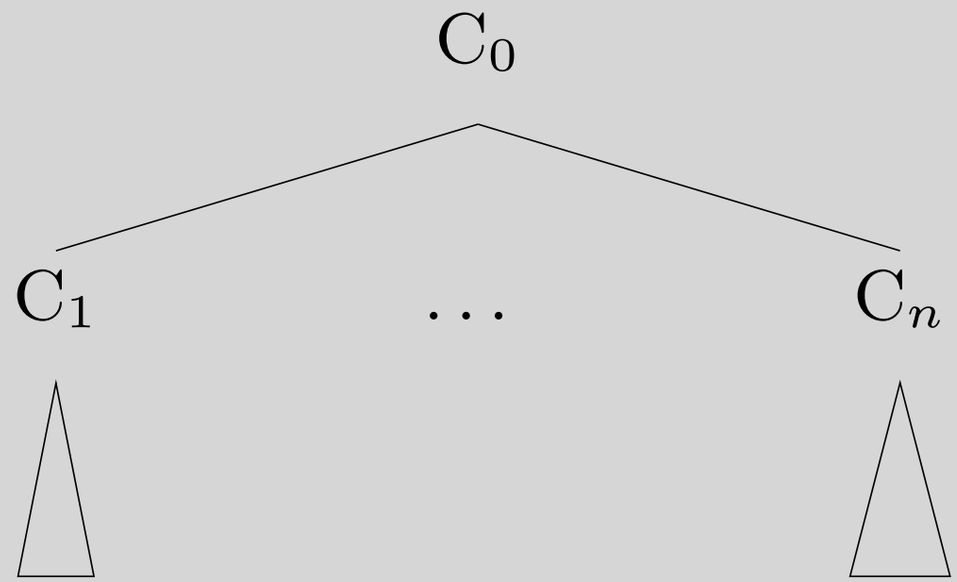
VP $\longrightarrow$ V NP

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



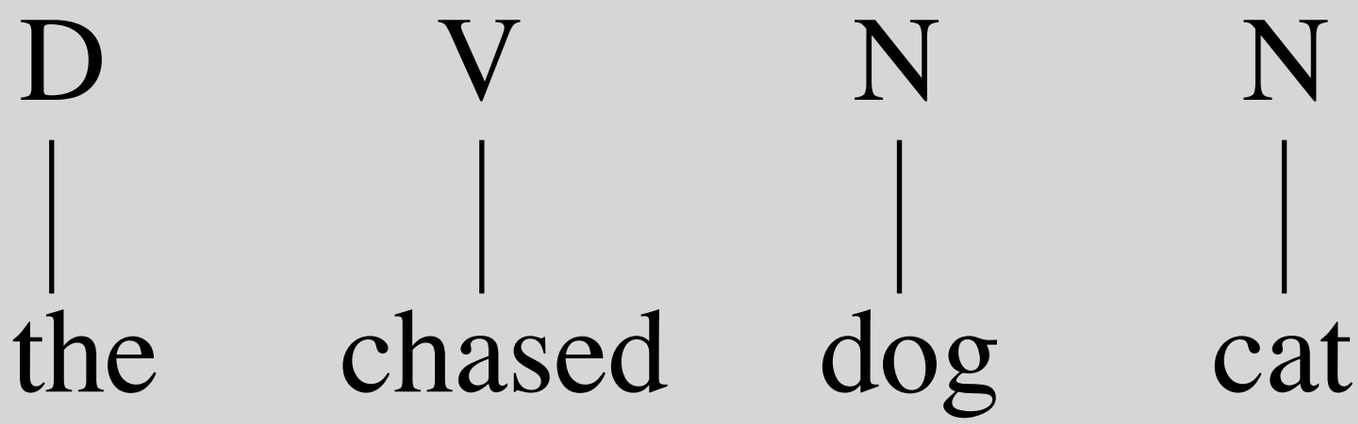
is a well-formed nonlexical tree if (and only if)

are well-formed trees, and

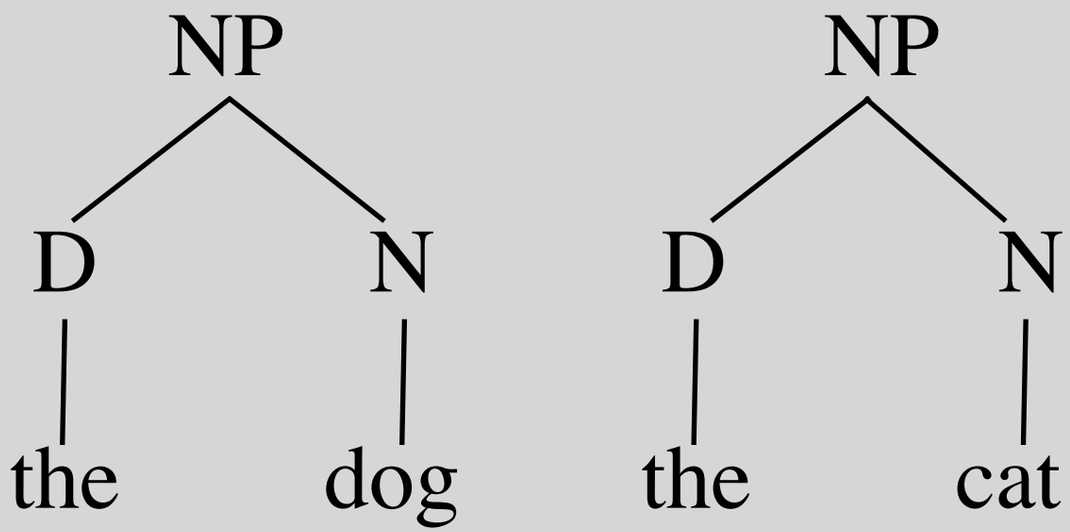
$C_0 \rightarrow C_1 \ldots Cn$ is a grammar rule.

Bottom-up Tree Construction

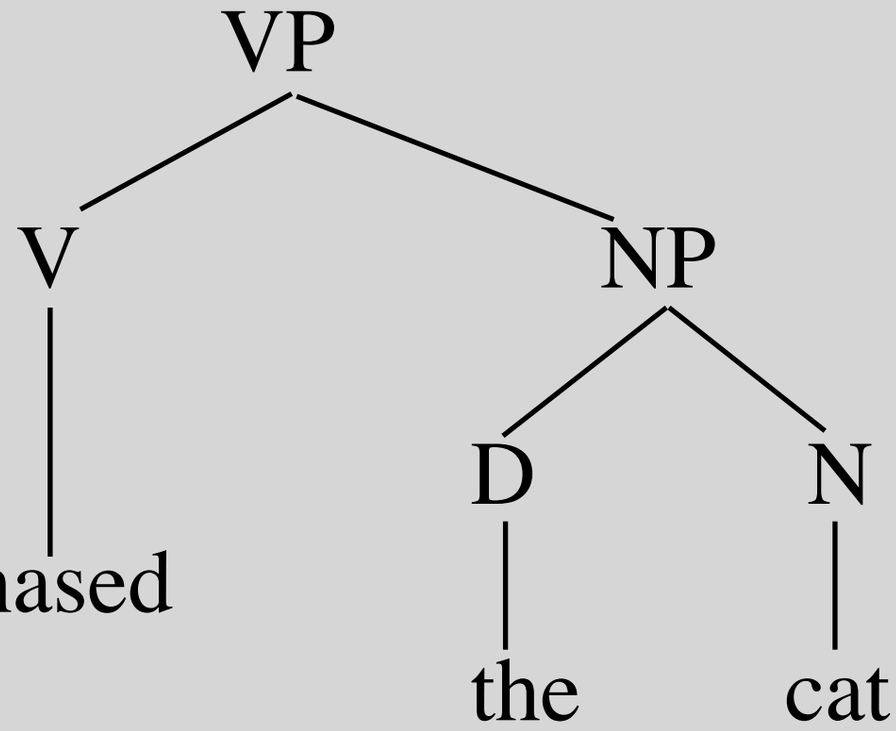
D: *the*
V: *chased*
N: *dog, cat*



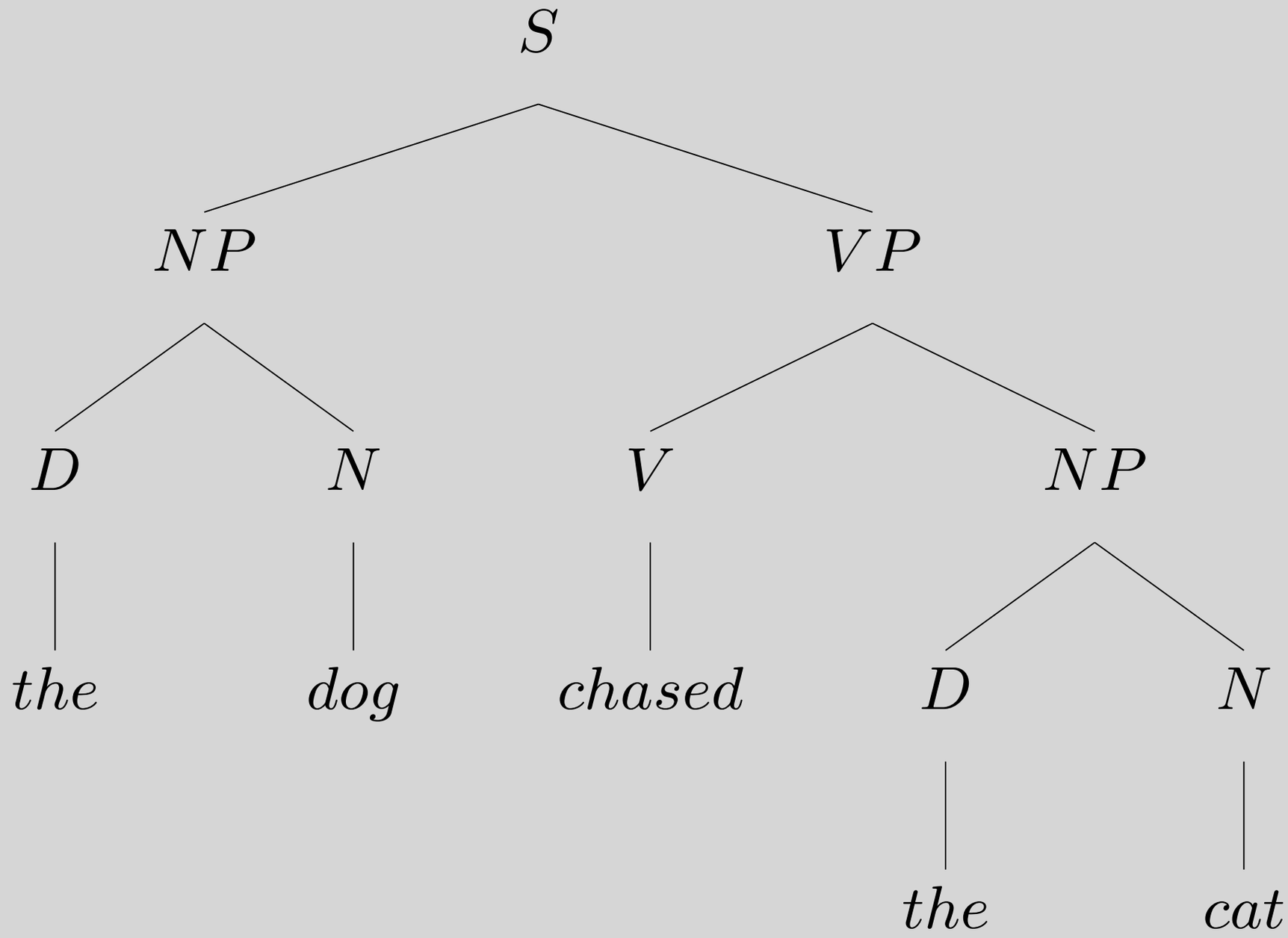
NP ⟶ D N



VP ⟶ V NP

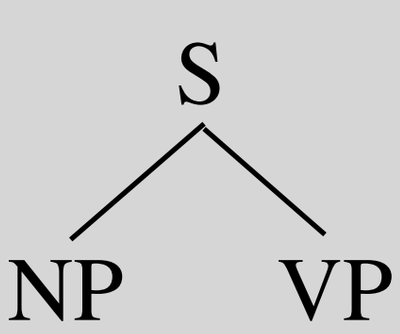


S ⟶ NP VP

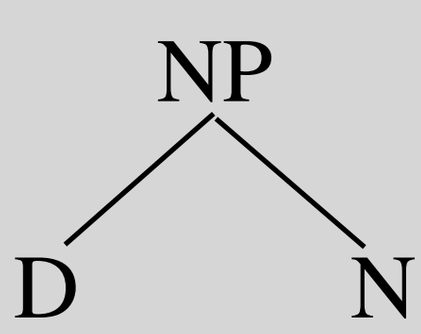


Top-down Tree Construction

$S \longrightarrow NP\ \ VP$

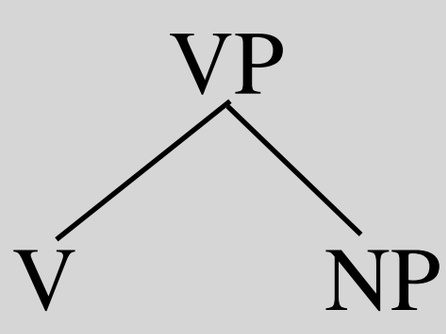


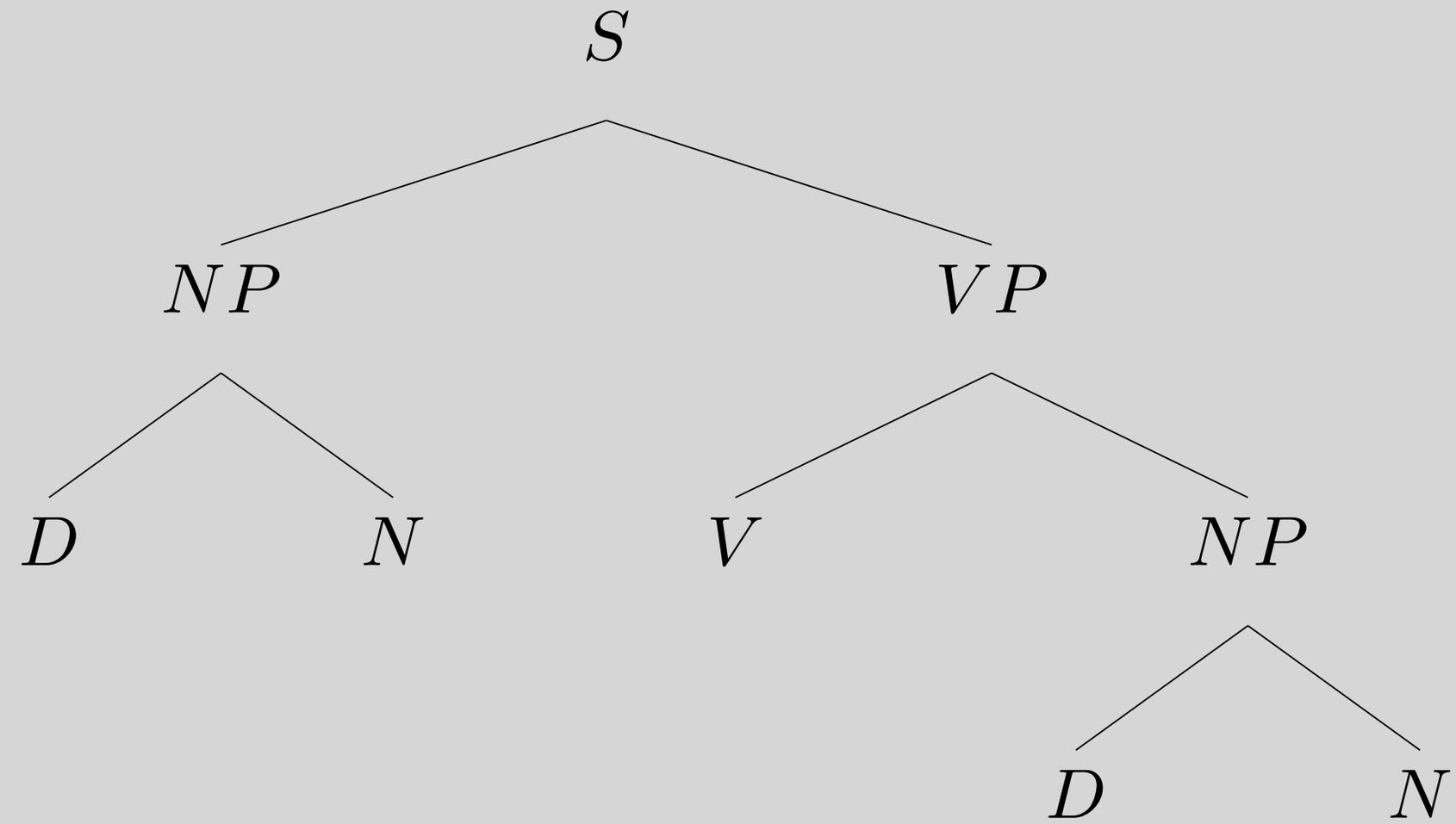
$NP \longrightarrow D\ \ N$



(twice)

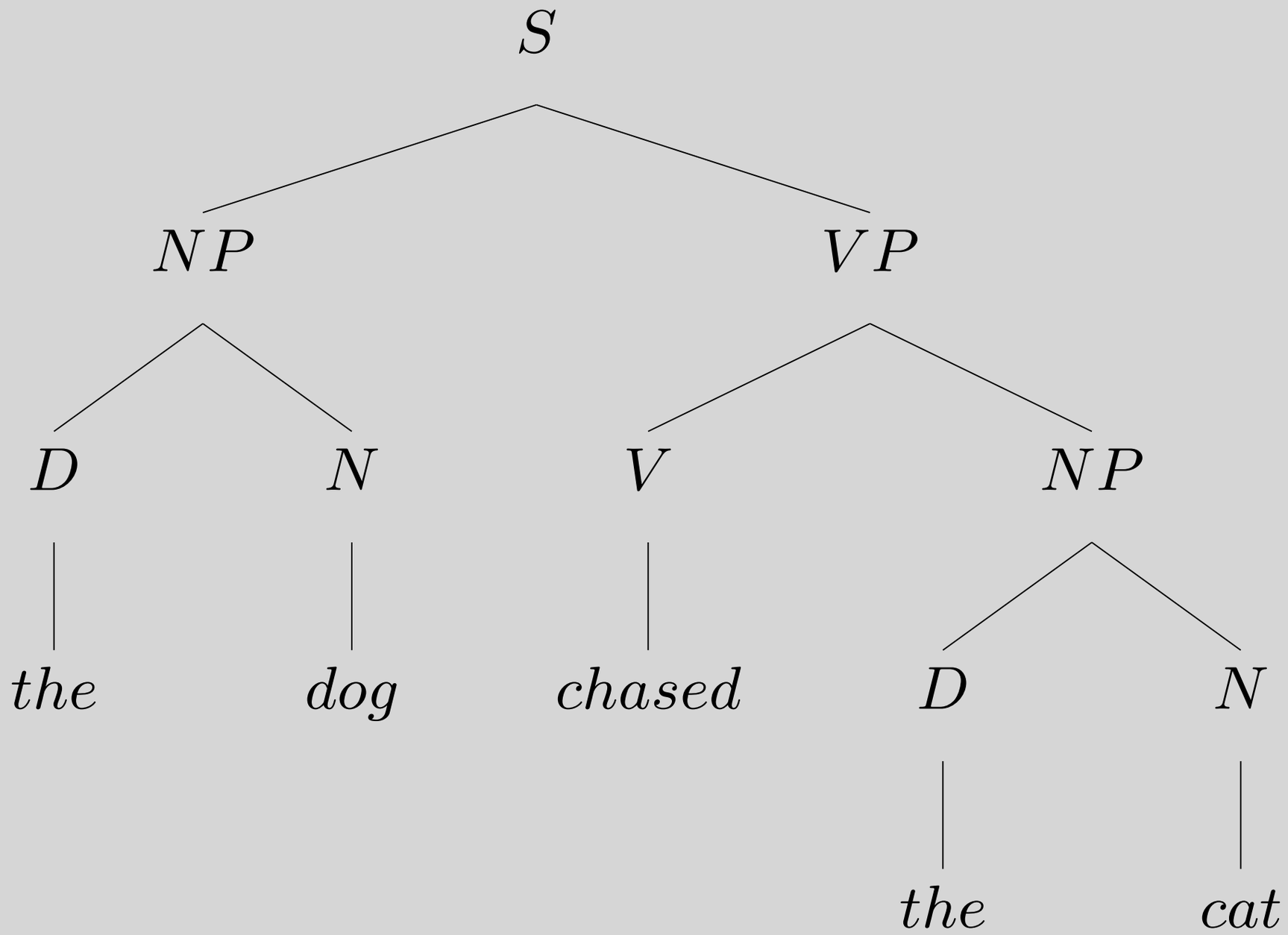
$VP \longrightarrow V\ \ NP$





D	V	N	N
the	chased	dog	cat

Poll!



Reading Questions

- Since a CFG is completely neutral w/ respect to top-down and bottom-up approaches when constructing trees/analyzing sentence structure, when this is transformed to a computer program, is there a general approach or is it also flexible based on how the programmer designs it?

- The chapter states that the examples are generating the sentences in a bottom-up approach, but the set of sentence generated are the same if written top-down. What are the advantages and disadvantages towards choosing bottom-up versus top-down and in what situations does choosing one over the other really matter?

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 $f(\text{SG})$:

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

- Let r be the regular language $wa^*b^*xc^*d^*y$
- $f(\text{SG}) \cap r = wa^m b^n x c^m d^n y$
- $wa^m b^n x c^m d^n y$ is not context free.
- But context free languages are closed under intersection.
- $\therefore$ $f(\text{SG})$ (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

- What's the difference?
- Why do we need NOM? Why not:

NP -> (D) NP
NP -> NP+ CONJ NP
NP -> N PP

NP -> (D) N (PP)
NP -> NP+ CONJ NP

- Key ex: *No painting by Miro or drawing by Klee*

Reading Questions

- Other than agreement transformation and passivization transformation, what are other common transformations which can help solidify the concept of transformational grammar?

- I didn’t understand if there are rules associated with transformations. And if there are, why are they separate from the original CFG productions? A follow up question is: do transformations successfully capture the richness of English that CFGs can’t?

- I'm new to the concept of transformational grammar and it definitely seems like a big topic. I am wondering if there is a general rule about which or how many constituents must carry over to the new phrase structure in order to be considered a transformation vs a completely new phrase (Or maybe there isn't a differentiation between those two?)

Reading Questions

- I would like to learn more about the applications of CFG to other language families such as Japanese which is SOV not SVO as English? Is CFG mostly applicable to English solely?

- "The conception of grammar we develop will involve general principles that are just as applicable to superficially different languages as they are to English. Ultimately, much of the outward differences among languages can be viewed as differences in vocabulary." What does "superficially" different from English mean here? Does this refer to languages that are just different varieties of English? Does it include languages that come from the same language family as English but exclude those that are from different language families?

Reading Questions

- How are L2 speakers of a language accounted for in the Universal Grammar theory? If grammar is a theory about the mental representation of linguistic knowledge, can we say L2 speakers of a language use a different form of grammar than the grammar used by native speakers of the same language?

Reading Questions

- Also, how would CFG represent sometimes ungrammatical usage in English such as slang?
- How do we account for the necessity of commas in conjunctions? Given that they differentiate a grammatical sentence from an ungrammatical one, do they have a formalizable syntactic role?

Reading Questions

- Chapter 2 introduces a few criteria for determining good grammar—namely, high coverage of valid sentences and low coverage of unwanted sentences. It seems that the development of better grammar usually involves case studies, as illustrated by many examples in chapter 2. My question is: are there other tools (for example, statistical evaluation) that linguists use to measure the progress in developing better grammar?