Ling 566
Oct 1, 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

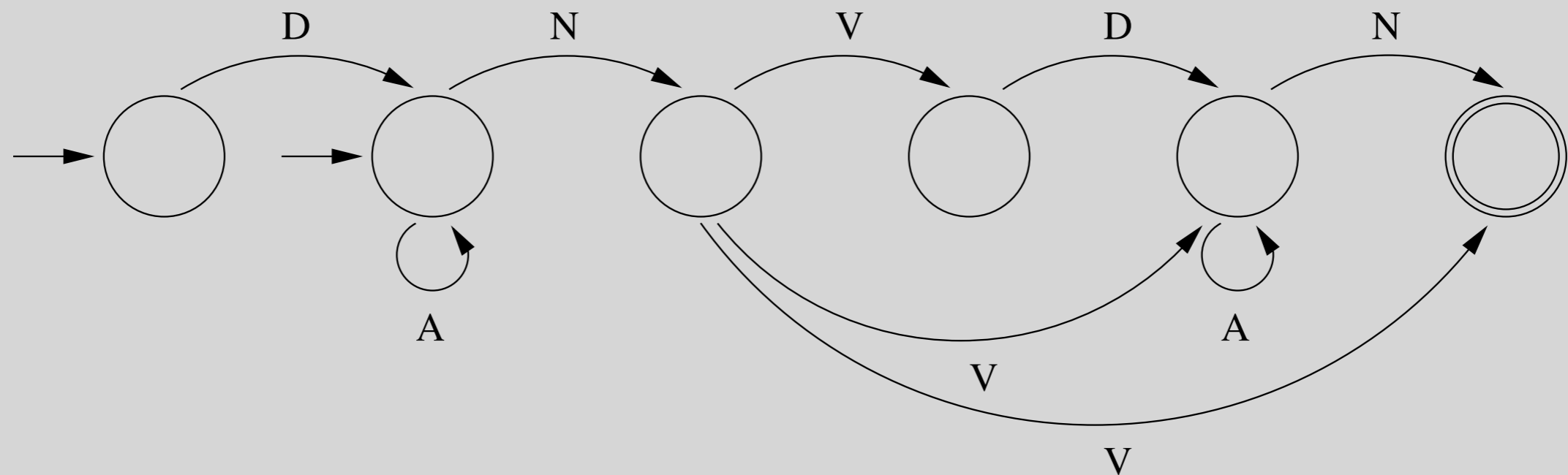 D   A   N   V

- the noisy dogs chased the innocent cats

  D   A   N   V   D   A   N

- a* = {ø, a, aa, aaa, aaaa, ... }

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

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

4

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

Type 0 Languages

Context-Sensitive Languages

Context-Free Languages

Regular Languages
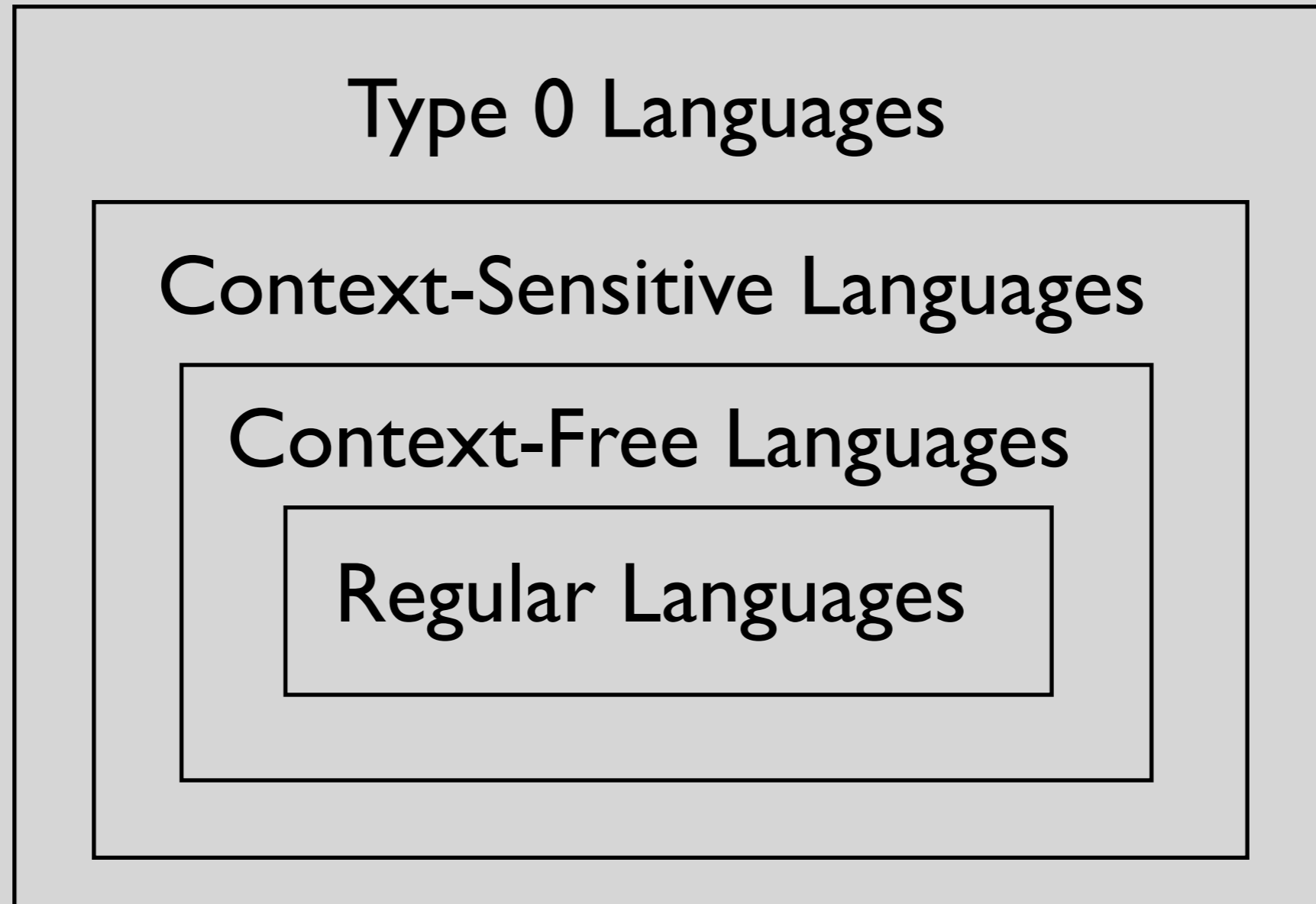
Context-Free Grammar

- 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 $\longrightarrow$ 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
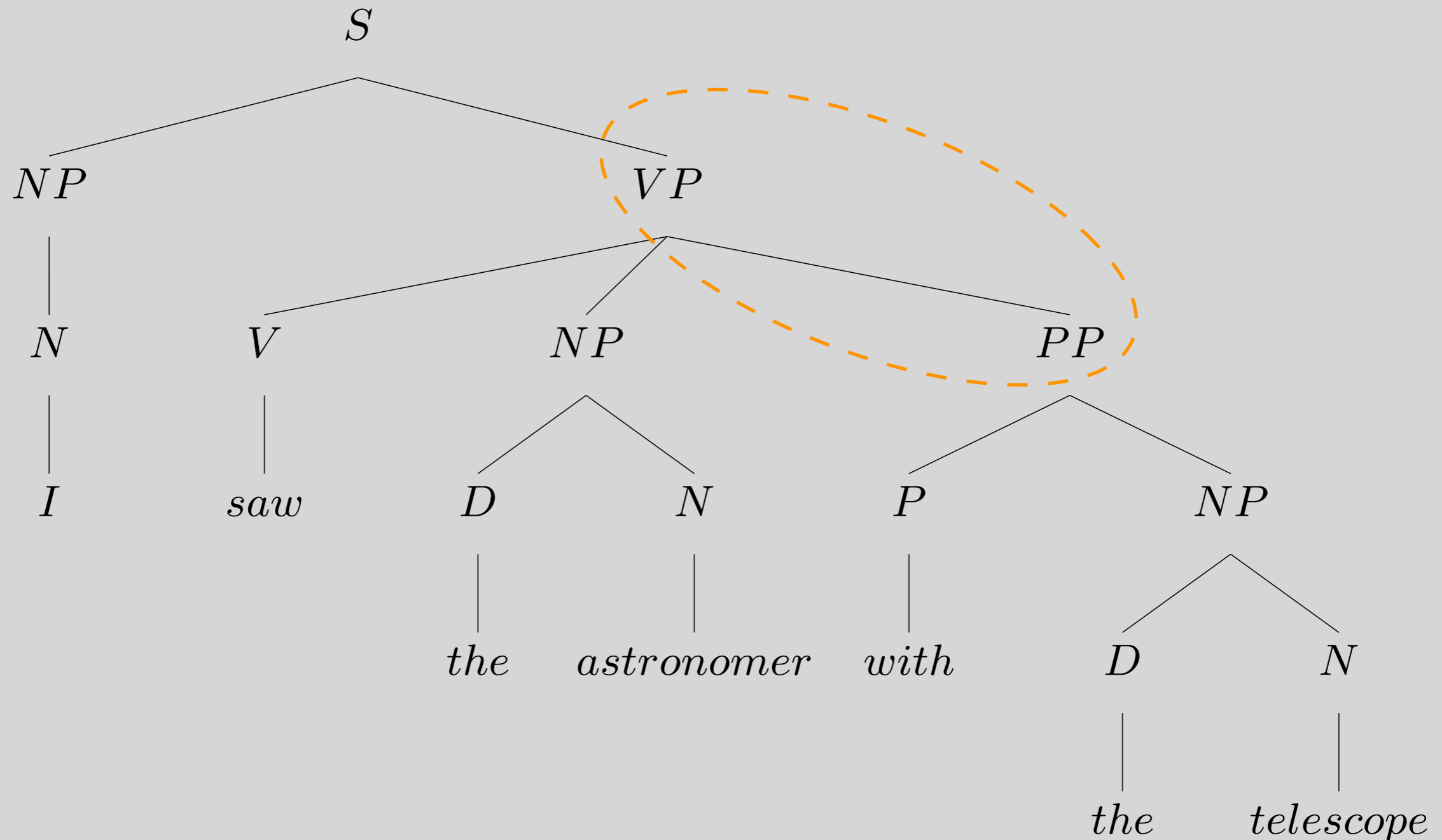
N: birds, fleas, dog, hunter, I

V: attack, ate, watched

P: for, beside, with
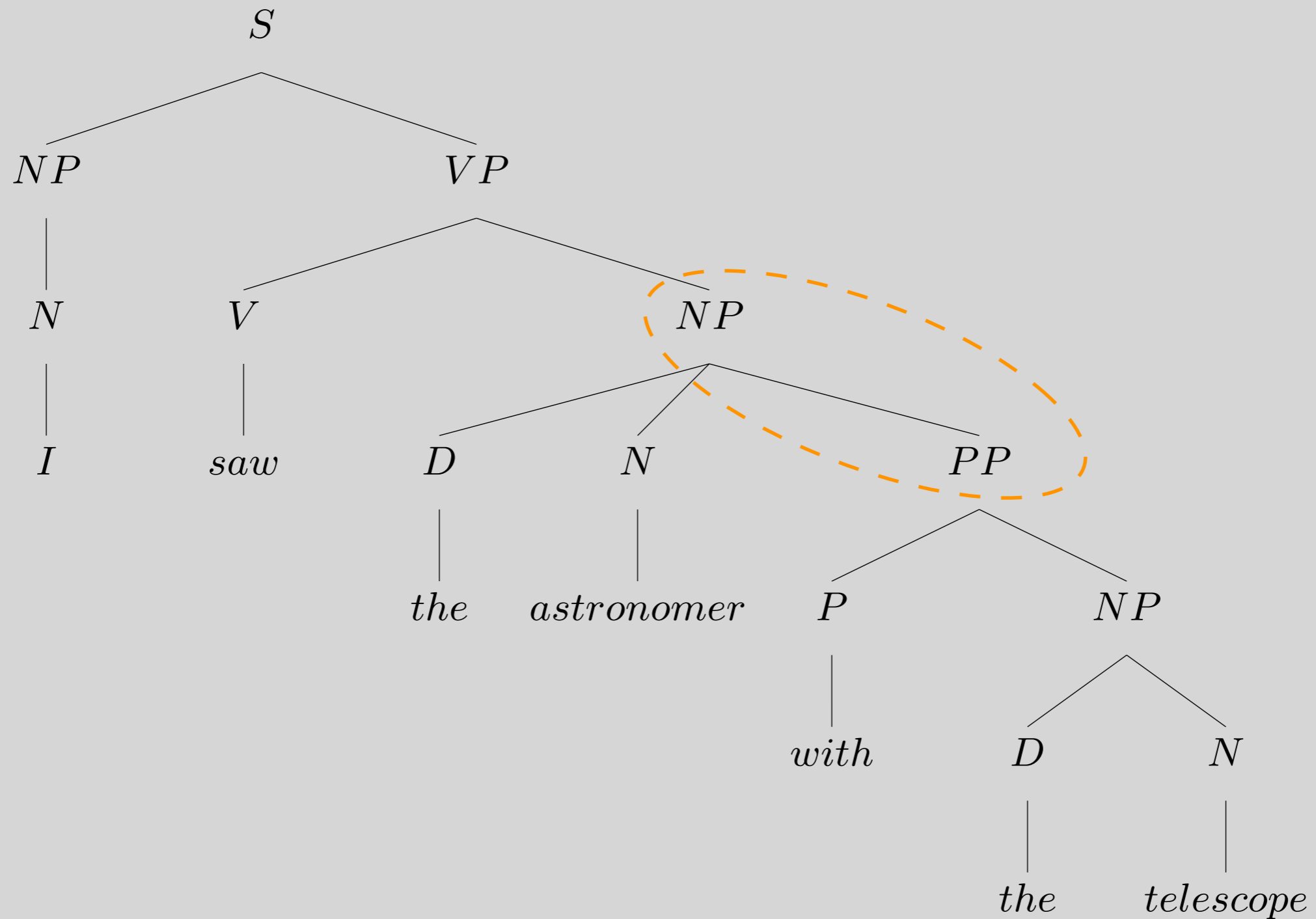
Structural Ambiguity

I saw the astronomer with the telescope.

Structure 1: PP under VP

S

NP   VP

N

V   NP   PP

I   saw   D   N   P   NP

the   astronomer   with   D   N

the   telescope

Structure 1: PP under NP

S
- NP
  - N
    - I
- VP
  - V
    - saw
  - NP
    - D
      - the
    - N
      - astronomer
    - PP
      - P
        - with
      - NP
        - D
          - the
        - N
          - telescope

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*.

19

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

C_0

C_1 ... C_n

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

C_1 , ... , C_n 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*

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

$$NP \longrightarrow D \ N \qquad\qquad VP \longrightarrow V \ NP$$

```
        NP                    NP                        VP
       /  \                  /  \                      /    \
      D    N                D    N                    V      NP
      |    |                |    |                    |     /  \
     the  dog              the  cat               chased  D    N
                                                          |    |
                                                         the  cat
```

$S \longrightarrow NP \ VP$

S

NP VP

D N V NP

the *dog* *chased* D N

the *cat*

24
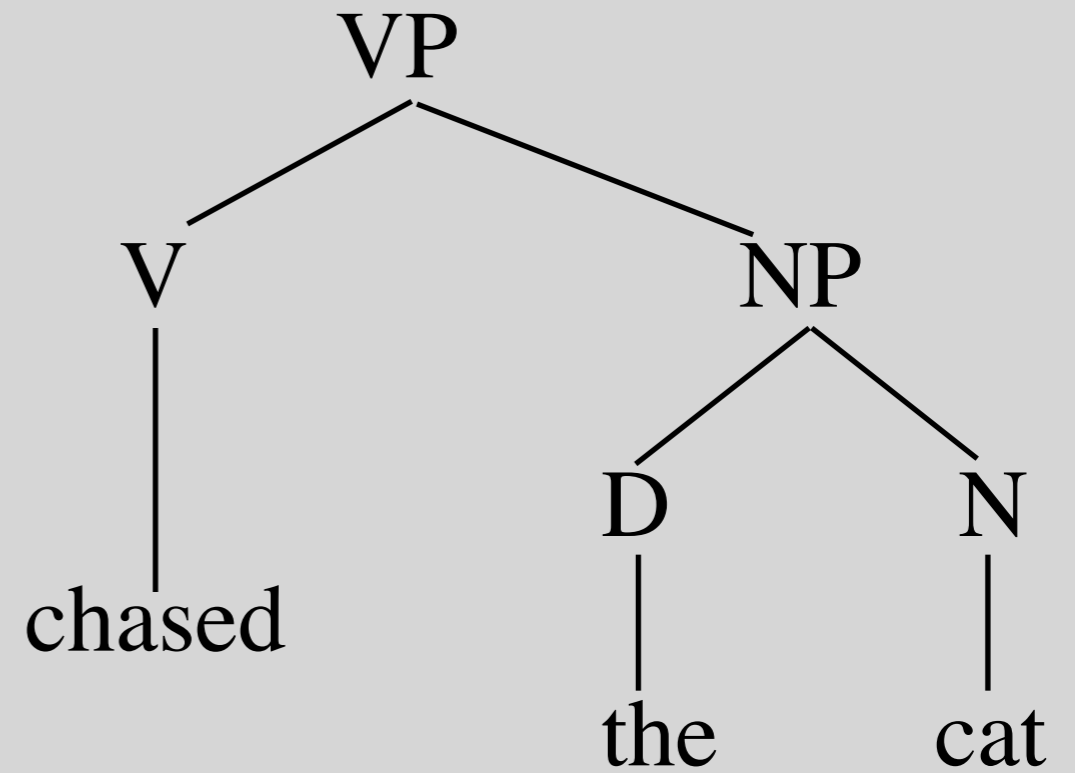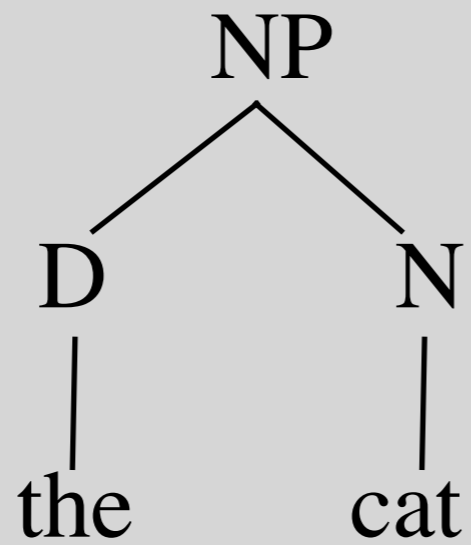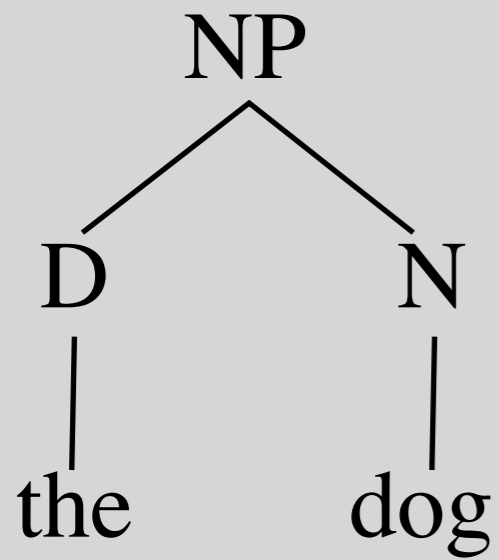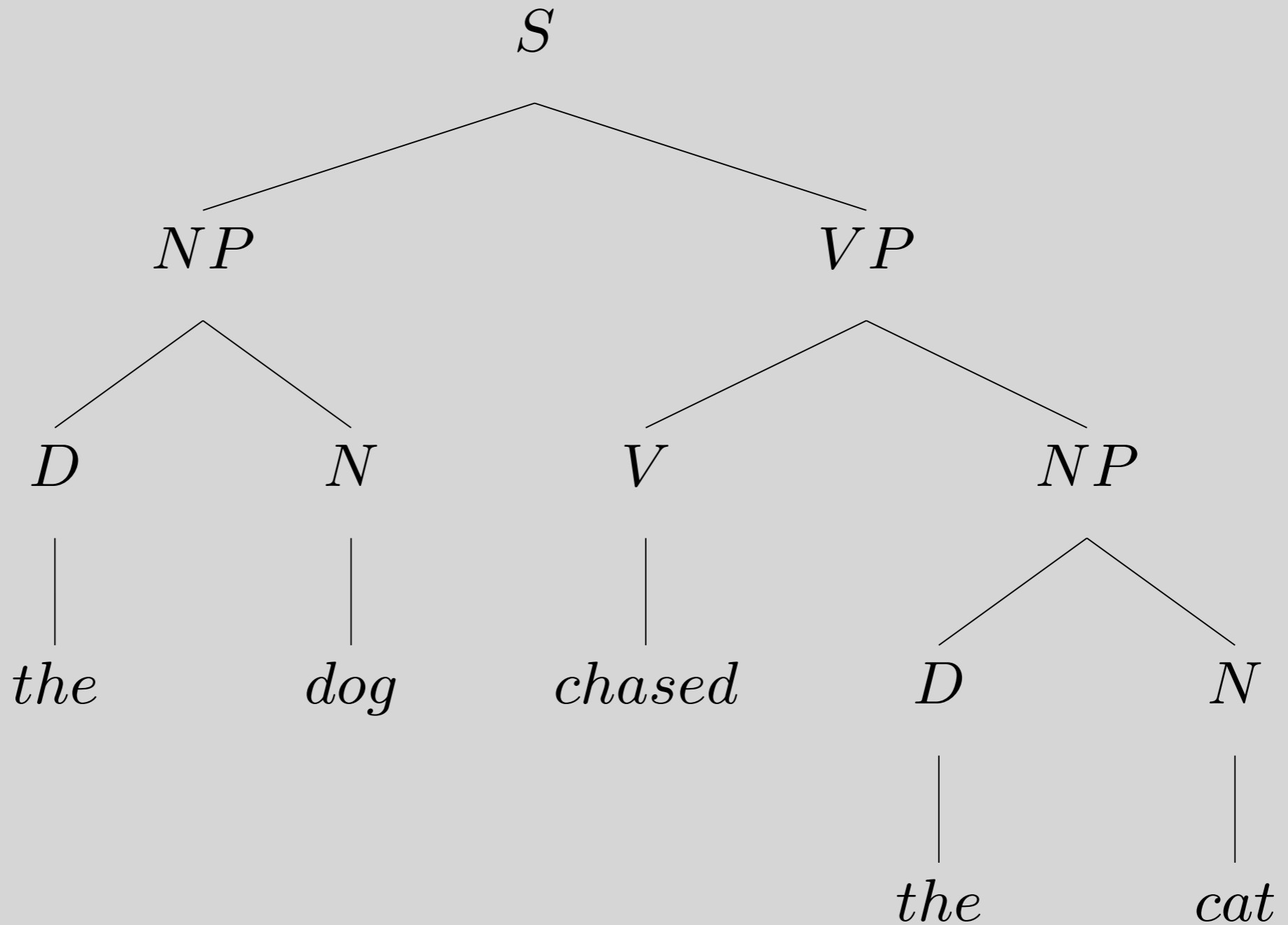
Top-down Tree Construction

S ⟶ NP VP   NP ⟶ D N   VP ⟶ V NP

```
       S                      NP                     VP
      / \                    /  \                   /  \
    NP   VP                 D    N                 V    NP
                          (twice)
```

$$S$$

$$NP \qquad VP$$

$$D \qquad N \qquad V \qquad NP$$

$$D \qquad N$$

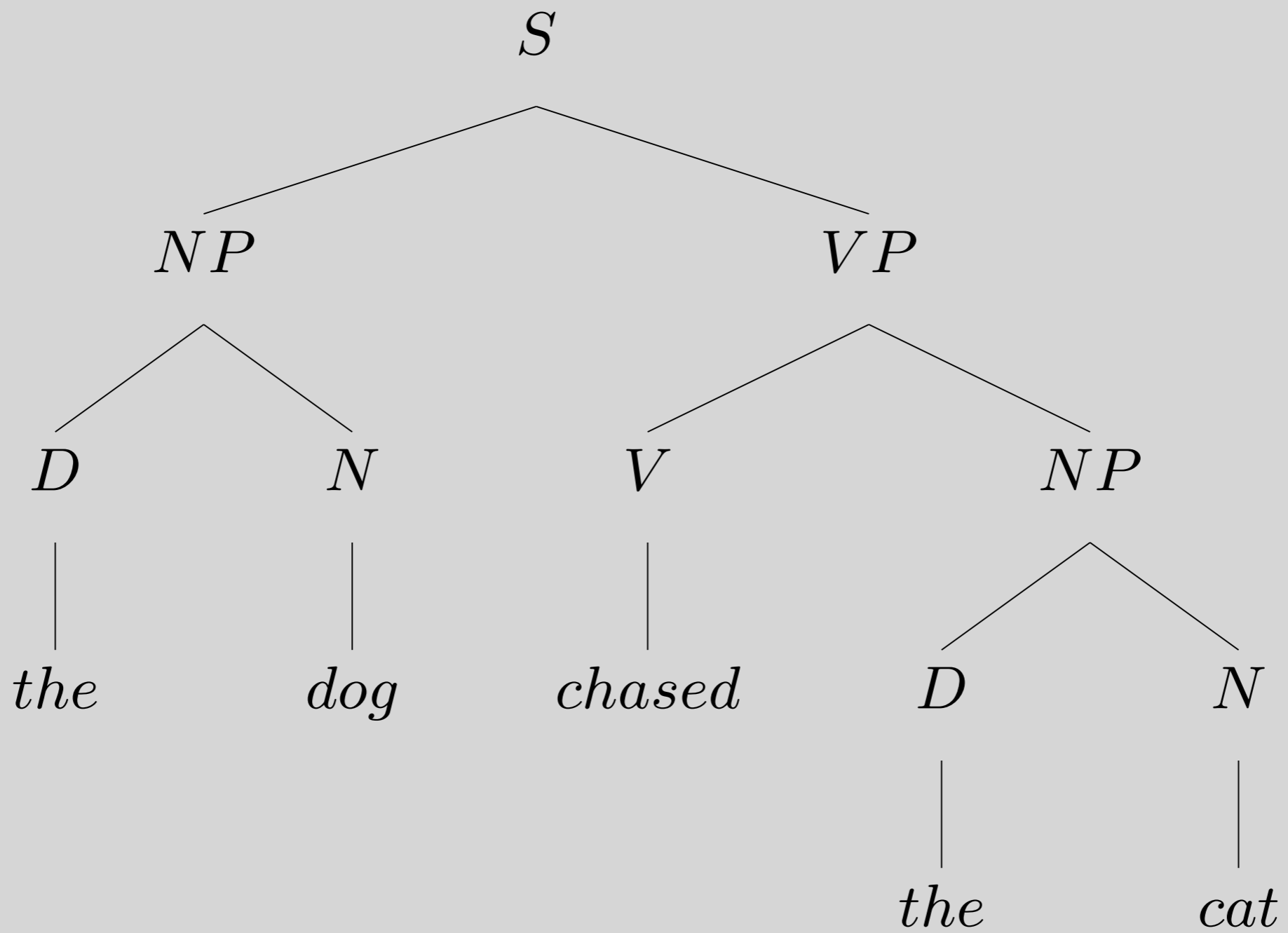D   V   N   N
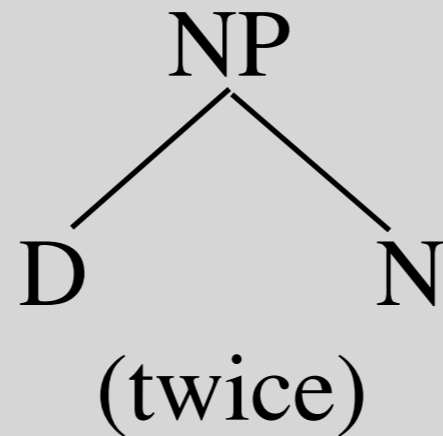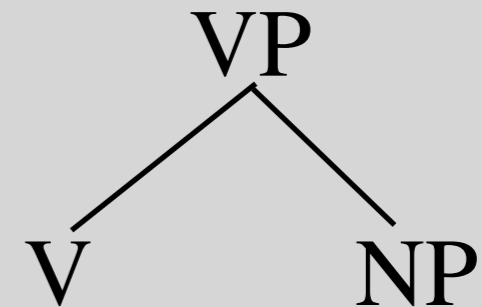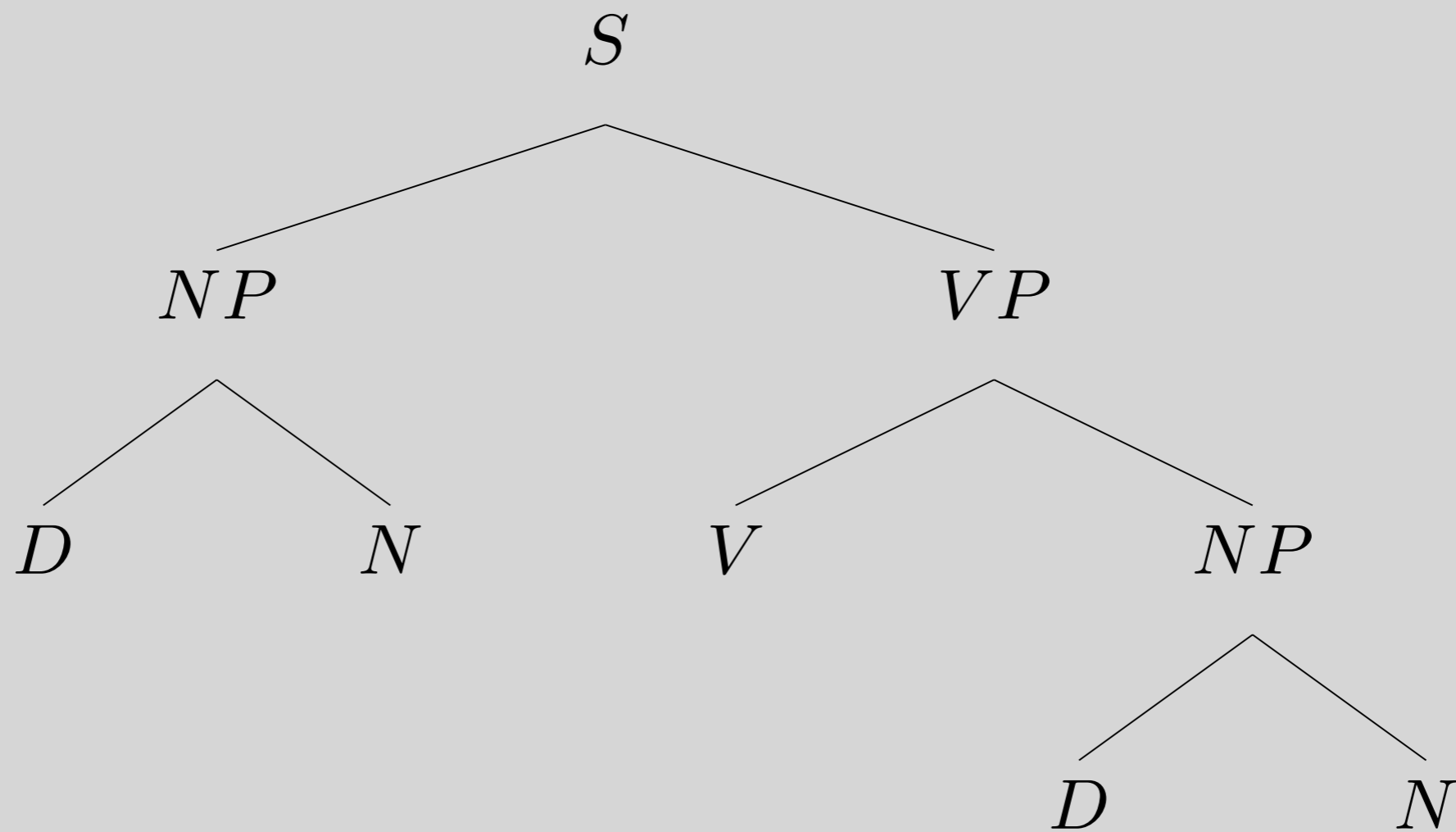
the   chased   dog   cat

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 $f(\mathrm{SG})$:

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

- Let r be the regular language $wa^*b^*xc^*d^*y$

- $f(\mathrm{SG}) \cap r = wa^m b^n x c^m d^n y$

- $wa^m b^n x c^m d^n y$ is not context free.

- But context free languages are closed under intersection.

- $\therefore f(\mathrm{SG})$ (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

- When creating a grammar is the goal to build something that can work across the largest most representative text sample? Make a grammar that is intuitive, elegant and easy to understand? Make a grammar that best produces results in downstream tasks? In diving into this matter its clear how we go about making grammars but not how we select one grammar as better. What kind of metrics are used?

Reading Questions

- What is headedness (as opposed to head)?

- "The notion of headedness is a problem for CFG because it cuts across many many different phrase types" (p. 37) I'm not sure if I understand this. Is this referring to the fact that in CFG heads must be defined for each phrase structure? Must there be more generalization in a grammar?

- How does this work in other languages?

Reading Questions

- Why use clefting for a test if NOM can't be clefted / why do we still think NOM is a constituent?

- What's the difference between NP and NOM?

- Why is example (22) not possible without NOM? Why wouldn't a NP category be sufficient?

  - *No painting by Miro or drawing by Klee was displayed.*

- What exactly does "NOM" do that a rule such as "NP -> NP PP" would not allow?

- Is NOM a subcategory of NP?

39

Reading Questions

- In the grammar described in Section 2.4.2 (23), why can't we collapse some of the rules and just say NP -> (D) NOM PP* ? Can we not use Kleene * anymore, like the set of rules in the grammar in Section 2.3 (11)?

- Is it possible for something to be a constituent in one formal grammar of English, but not another? Or is constituency something that's a set property of a language, where all grammars of that language should agree?

Reading Questions

- What is the value of constituency parsing for NLP?

- Constituency tests are a useful heuristic to determine how to analyze certain sentences but are there other tools we can use to determine constituency besides constituency tests? How do folks in industry determine constituency in their systems (if they do so at all)?

Reading Questions

- Although the Chomsky Hierarchy refers more to formal languages rather than natural, where might English fit in if we had to put it somewhere?

- Could some human language be context free?

- "The most convincing arguments for employing a grammatical formalism richer than regular expressions, however, have to do with the need to capture generalizations," (p25): does this mean that depending on the degree of generalizations needed for specific grammar, we will need context free grammar and even context sensitive grammar to help formalize the grammar? What else do we need to capture except for generalizations?

Reading Questions

- Where does TG fit into the Chomsky Hierarchy?

- I'm not entirely clear on the functional difference between transformational and non-transformational grammars. I'm mostly wondering if the two approaches are functionally equivalent in terms of capabilities, but just different techniques that just vary in their efficiency/clarity in describing certain things. Or if one has limitations relative to the other in what it can produce.

Reading Questions

- Since all that syntactic investigation has to work with is the surface form of an utterance, what sort of evidence is used to justify the idea that a transformation is occurring? A related question, in my mind, is why would we want to posit that declarative/interrogative or active/passive sentence pairs are derived from the same underlying phrase structure? Additionally, does this question matter if making those assumptions creates a more elegant or complete theory of grammar?

- What makes Transformational Grammar so successful? Is there a relationship between Transformational Grammar and Universal Grammar?