Chapter 2, sections 2.3-2.7:

Context-free Grammars

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

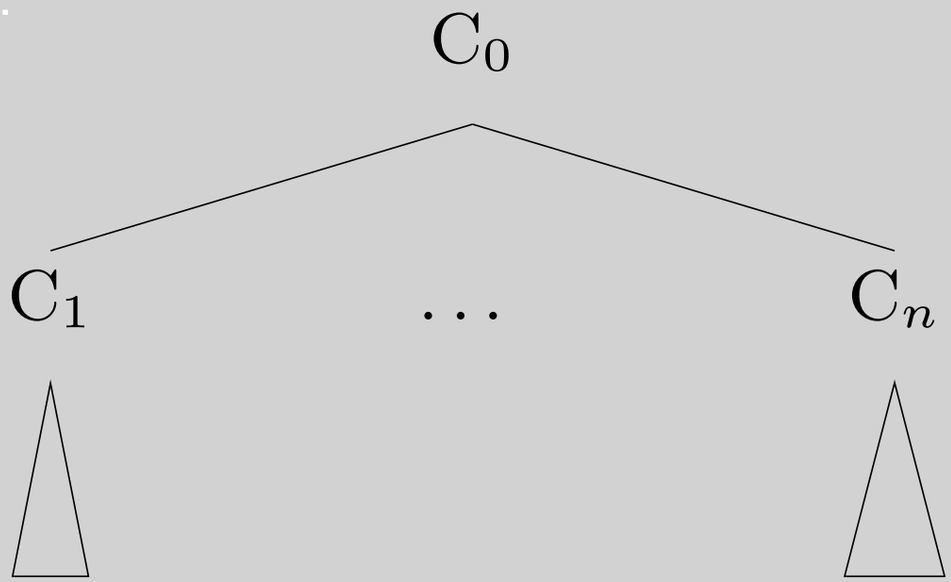
VP $\longrightarrow$ V NP

D: *the*

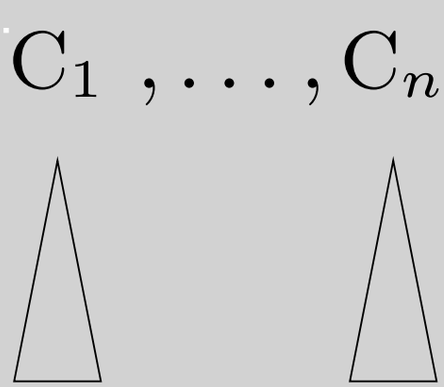
V: *chased*

N: *dog, cat*

Trees and Rules



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

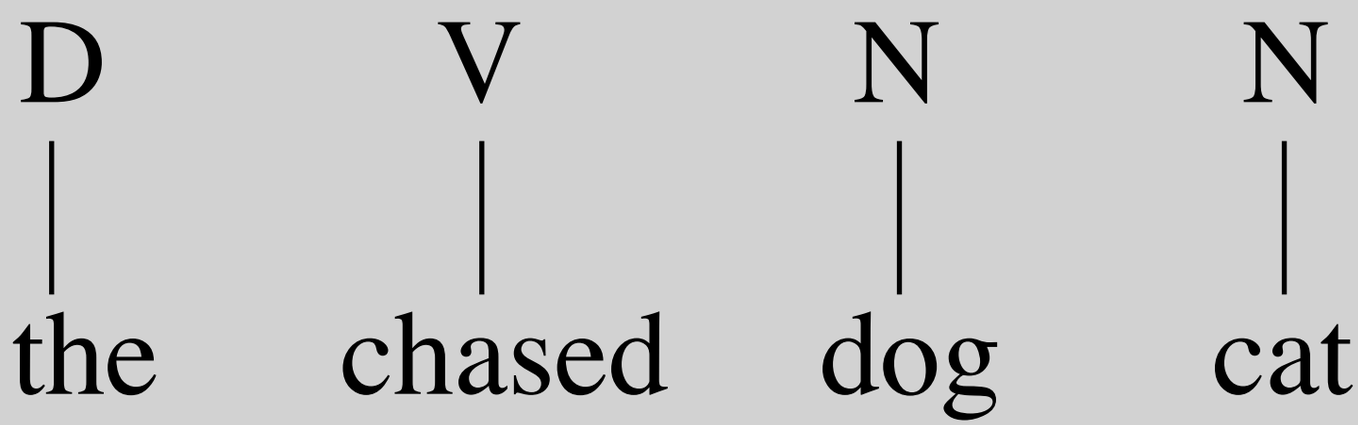


are well-formed trees, and

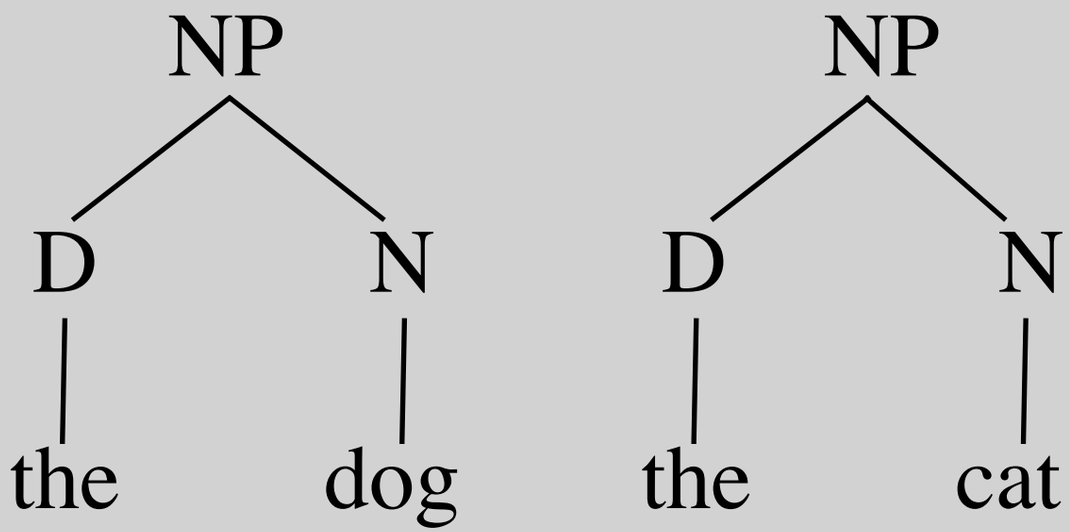
$C_0 \rightarrow C_1 \ldots Cn$ is a grammar rule.

Bottom-up Tree Construction

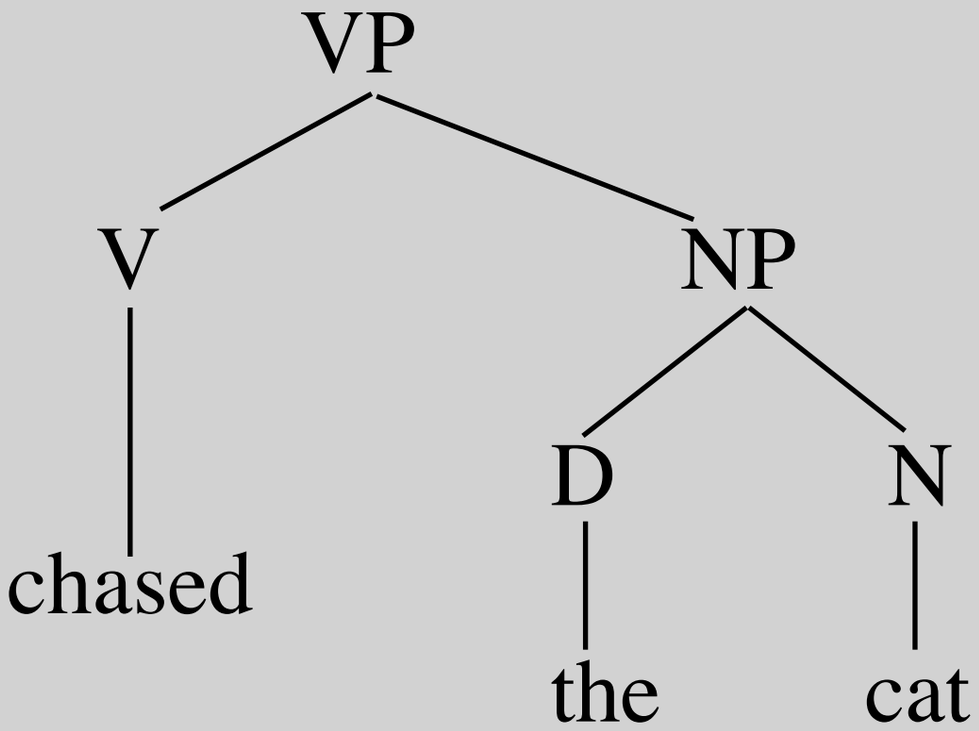
D: *the*
V: *chased*
N: *dog, cat*



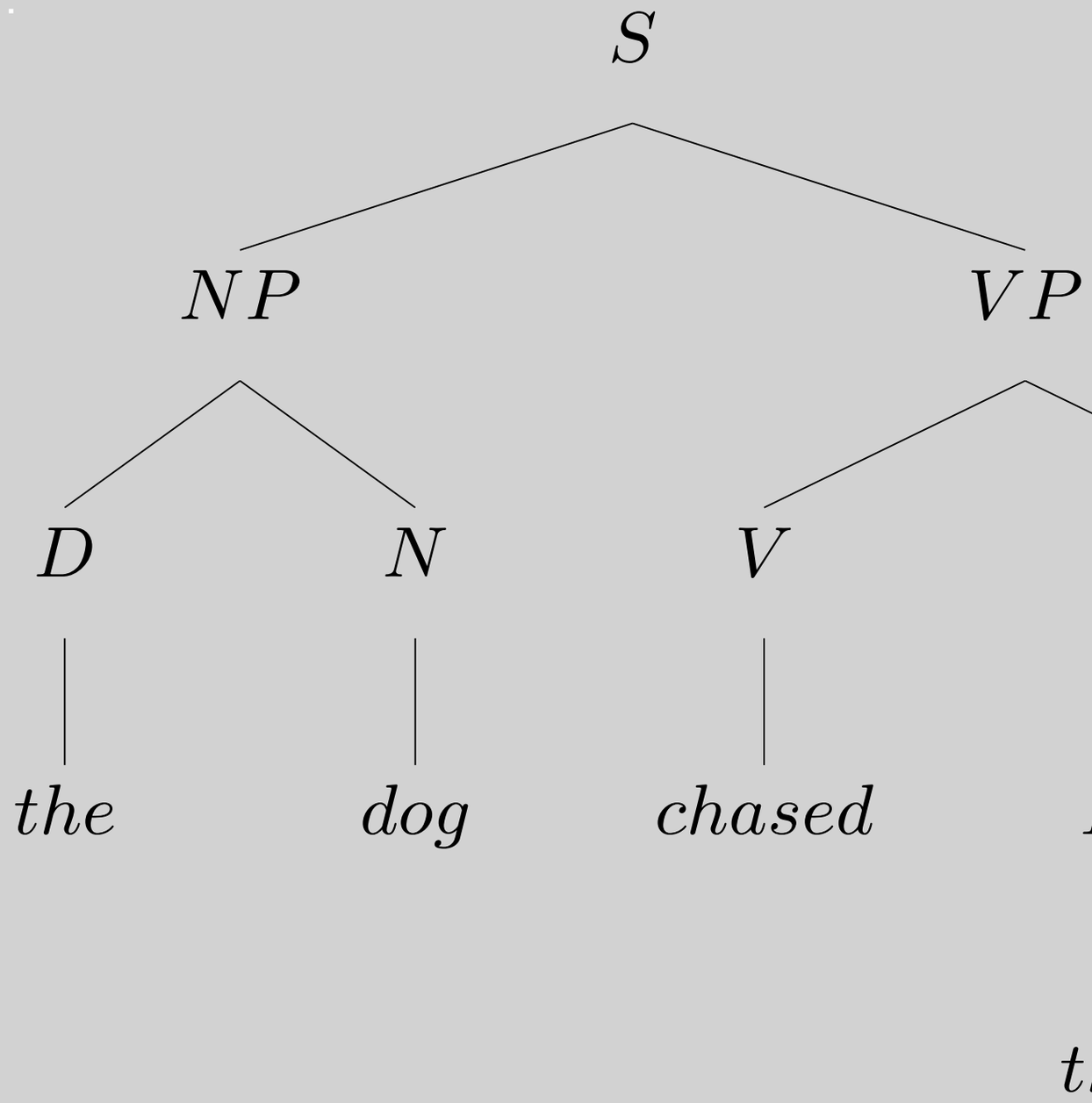
NP ⟶ D N



VP ⟶ V NP

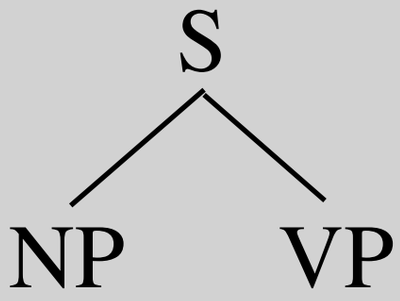


$$\mathrm{S} \longrightarrow \mathrm{NP}\ \ \mathrm{VP}$$

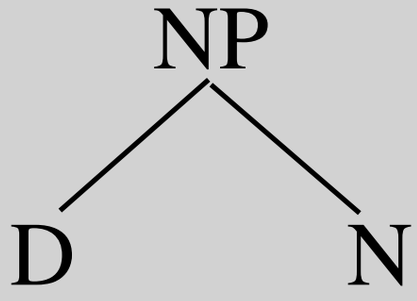


Top-down Tree Construction

S ⟶ NP VP

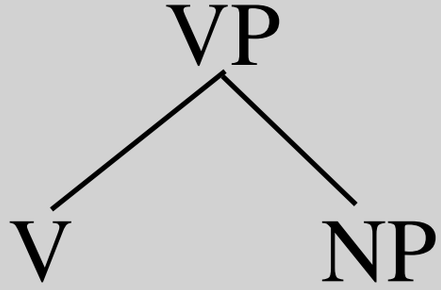


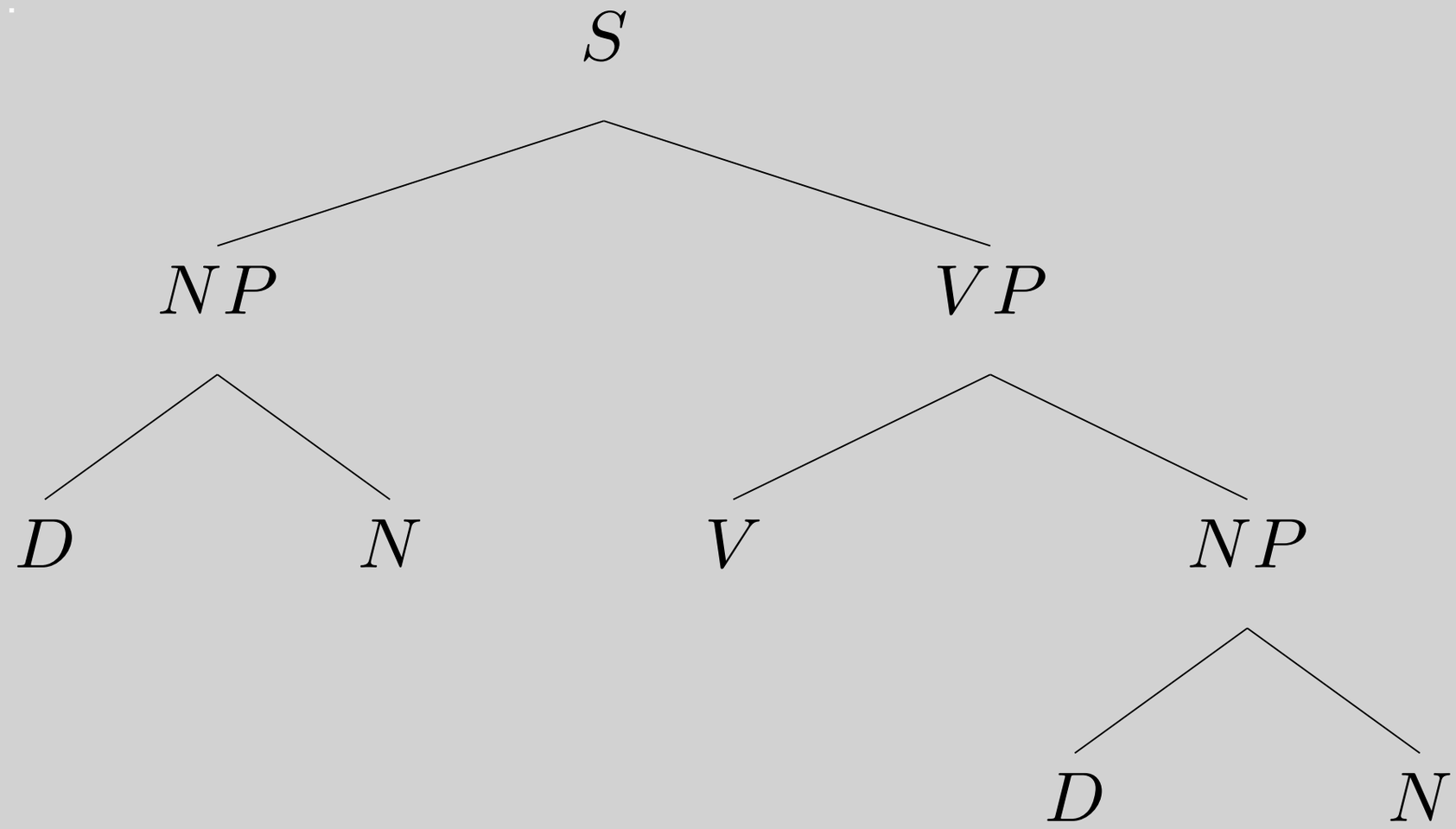
NP ⟶ D N

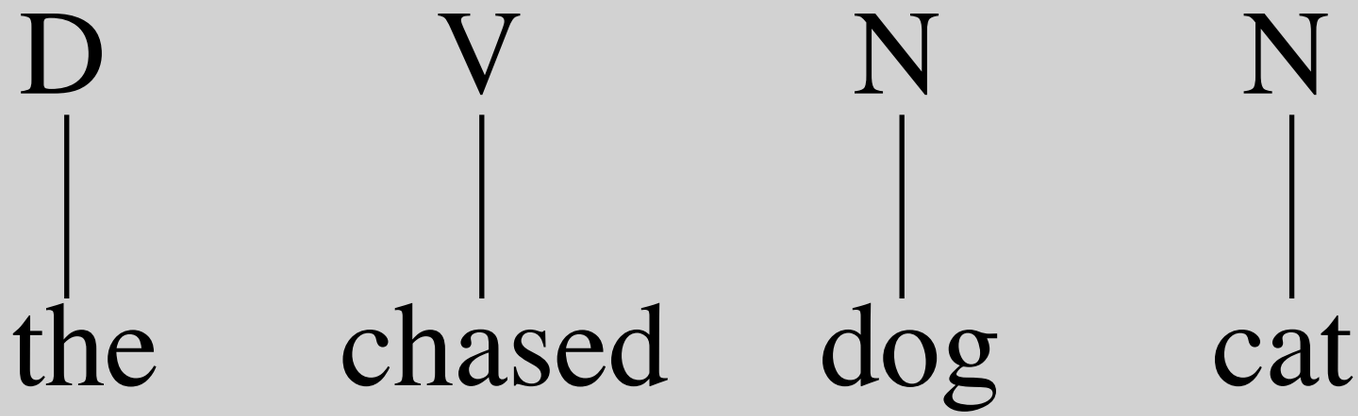


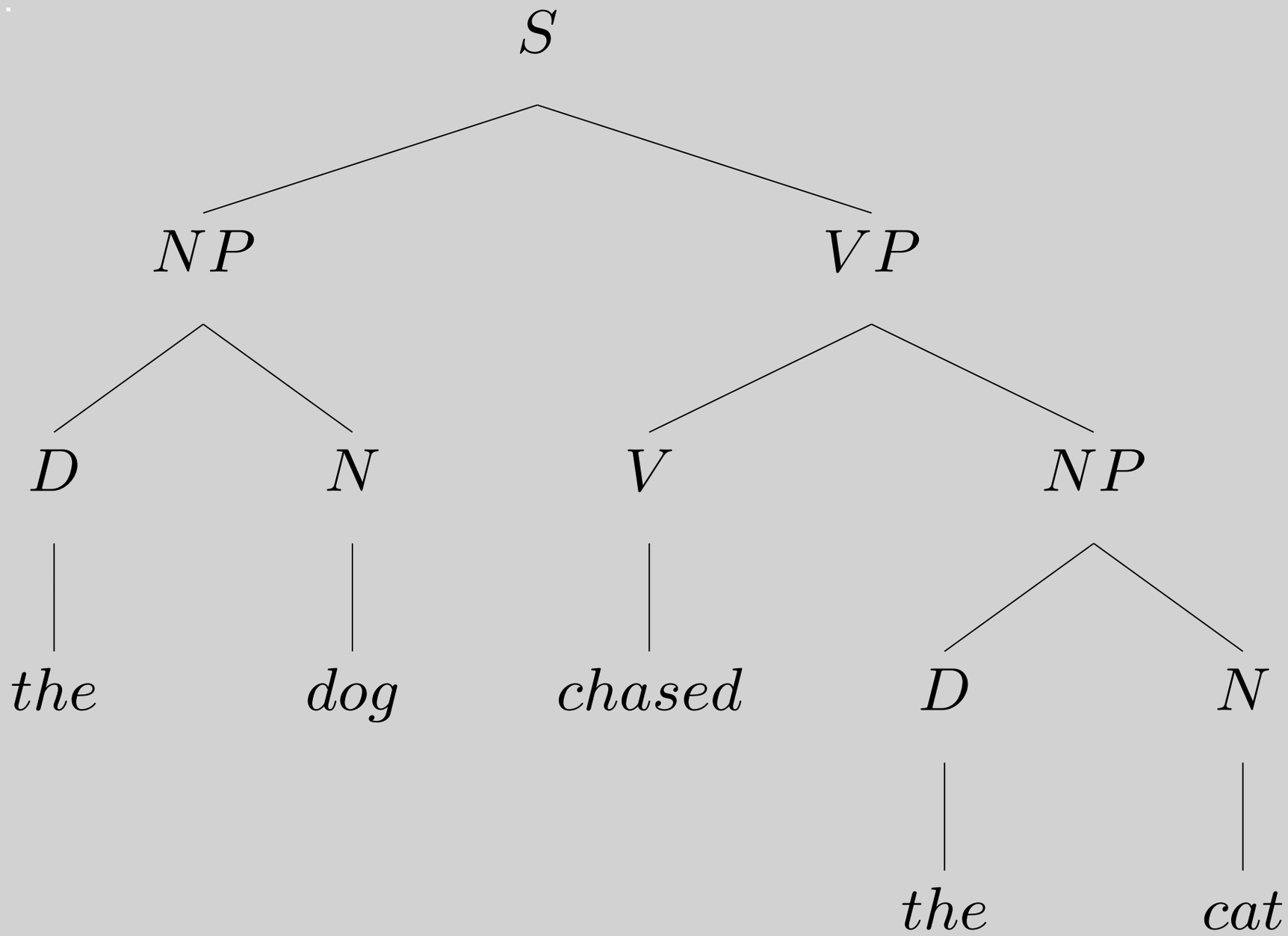
(twice)

VP ⟶ V NP









Bottom-up and top-down approaches are equivalent for CFG, but can differ for more complex types of grammars:

Rules

S ⟶ A B

A ⟶ C D, in the environment __E.

B ⟶ E F, in the environment D__.

Lexicon

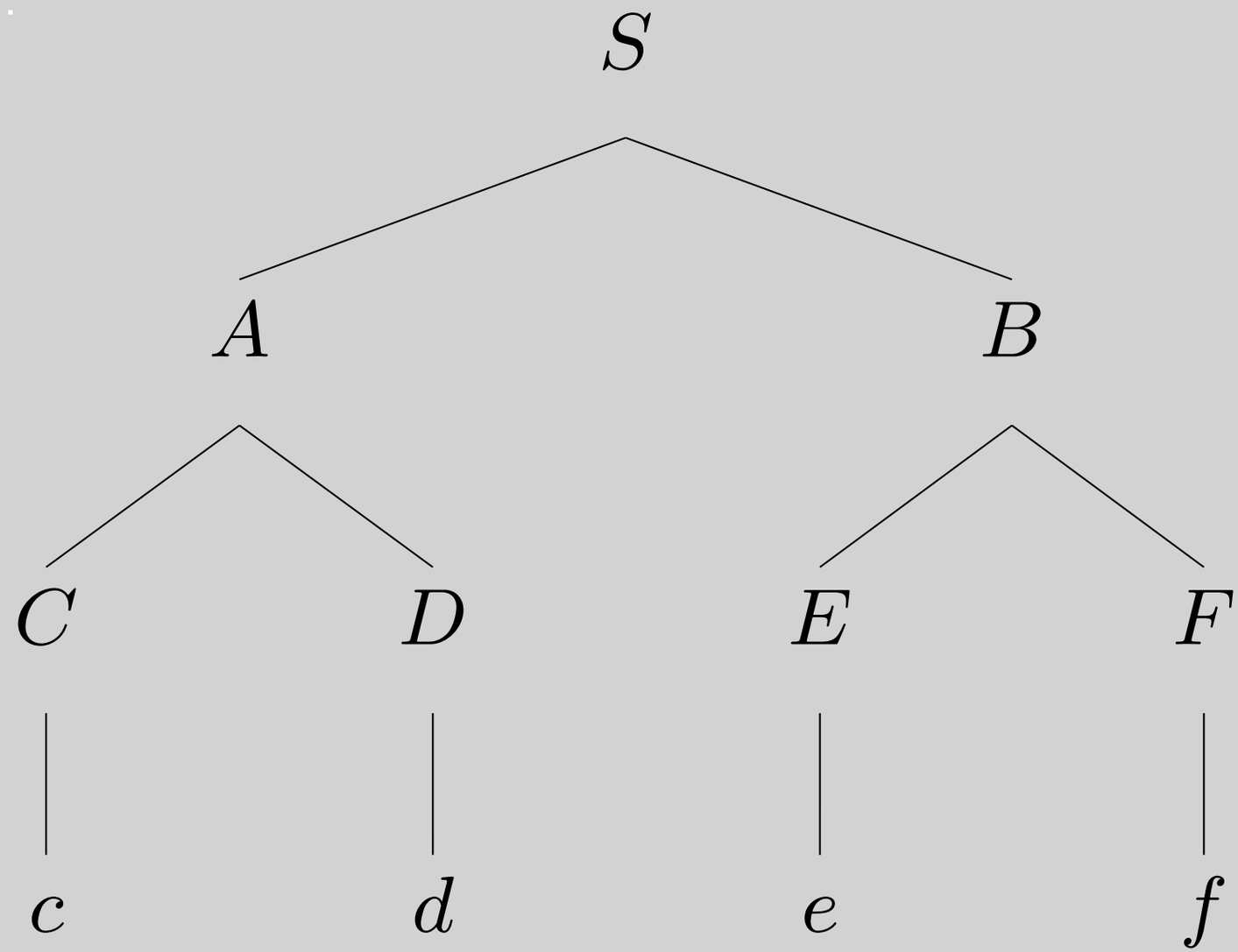
C: c

D: d

E: e

F: f

This tree is licensed bottom-up,
but not top-down:



Weaknesses of CFG

- It doesn't tell us what constitutes a linguistically natural rule

  VP P NP

  NP 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.

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.