Context-Free Grammars

Natalie Parde
UIC CS 421

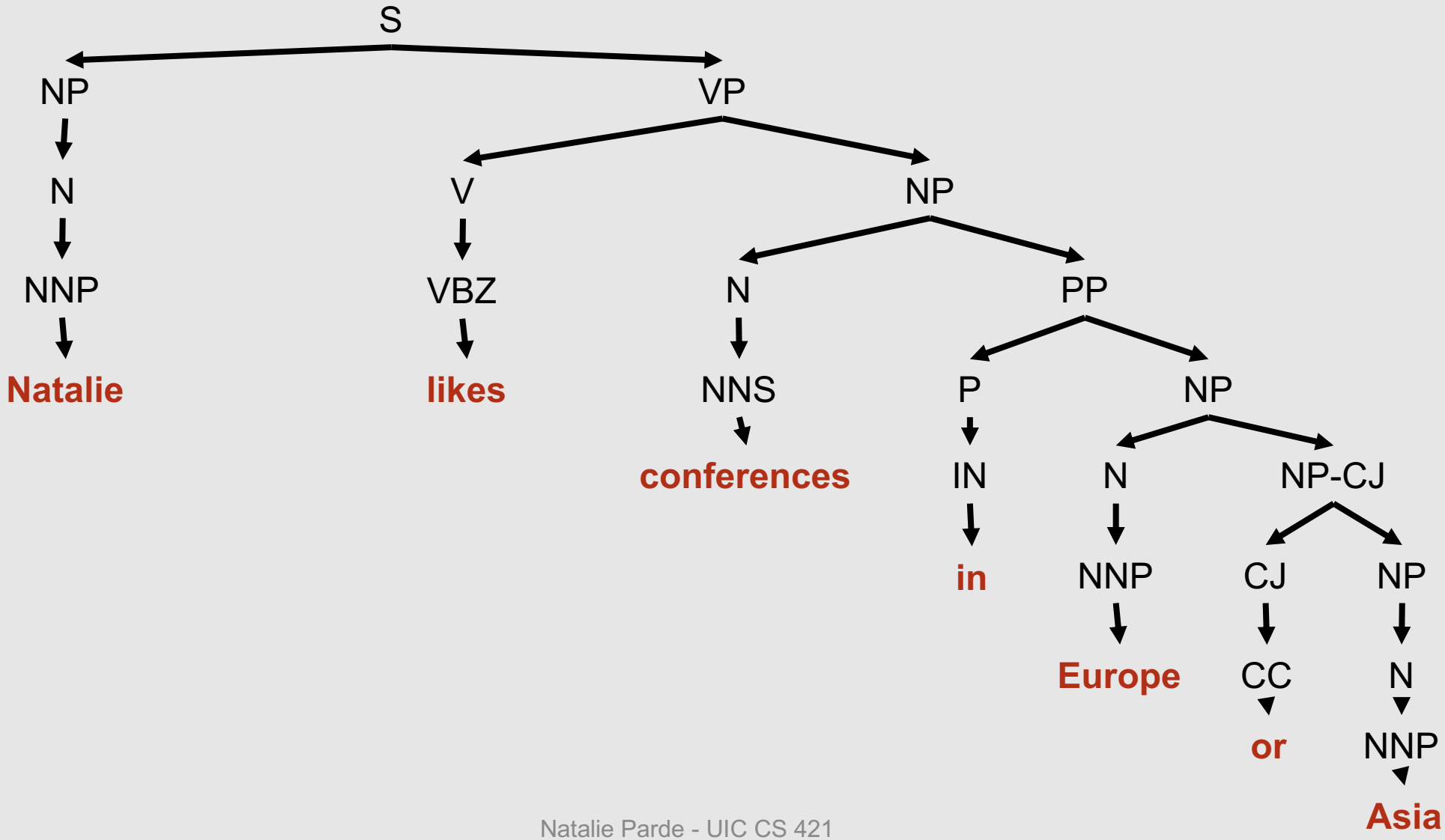
Sentences Form a Hierarchy

- A sentence consists of words that can be grouped into phrases (**constituents**)
- **Sentence structure** defines dependencies between these constituents

We can use trees to model this hierarchy.

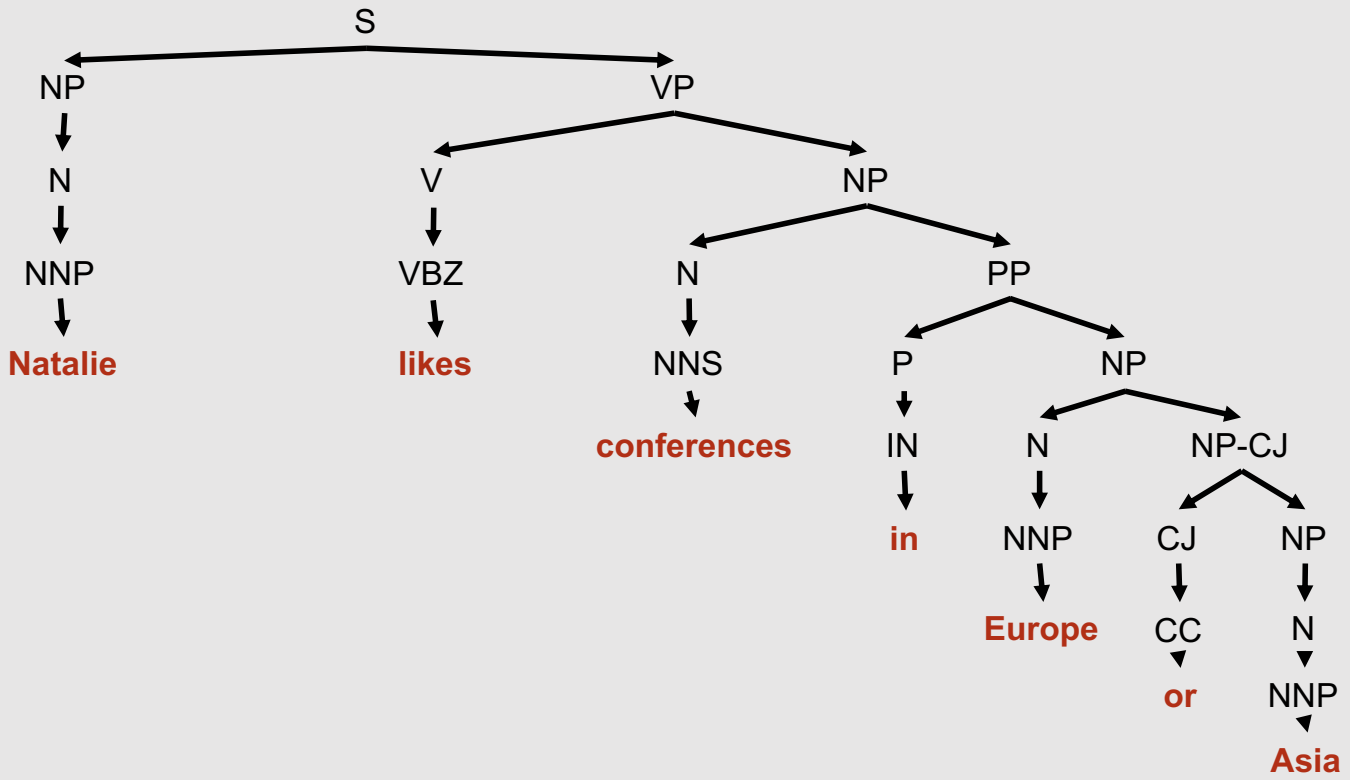
- Formal trees will usually have **internal (non-terminal) nodes** and **outer (terminal) leaves**
- **Nodes: Elements of sentence structure**
  - Constituent type
  - POS type
- **Leaves: Surface wordforms**
- The nodes and leaves are connected to one another by **branches**

What does this look like?



Trees can grow to be quite complex!

However, they can be reduced to simple subtrees defining underlying syntactic constituents



The grammars defining these hierarchical trees are context-free grammars.

- **Context-Free Grammar (CFG):** A mathematical system for modeling constituent structure in natural language.
- Also called **Phrase-Structure Grammars**
- CFGs can describe all regular languages
- Why is it called context-free?
  - A subtree can be replaced by a production rule independent of the greater context (other nodes in the hierarchy) in which it occurs.

CFGs are defined by productions that indicate which strings they can generate.

- **Production:** Rules expressing the allowable combinations of symbols (e.g., POS types) that can form a constituent
- Productions can be **hierarchically embedded**
  - Noun Phrase (NP) → Determiner Nominal
  - Nominal → Noun | Nominal Noun

Production rules determine how constituents can be combined.

- **Constituent:** A group of words that behaves as a single unit.
  - Noun Phrase: the woman, the woman with red hair, the last conference of the year
  - Prepositional Phrase: with red hair, of the year
  - Verb Phrase: drinks tea, likes going to conferences
- Phrases contain **heads** and **dependents**
  - **Heads:** the **woman** with red hair, the last **conference** of the year
  - **Dependents:** **the** woman **with red hair**, **the last** conference **of the year**

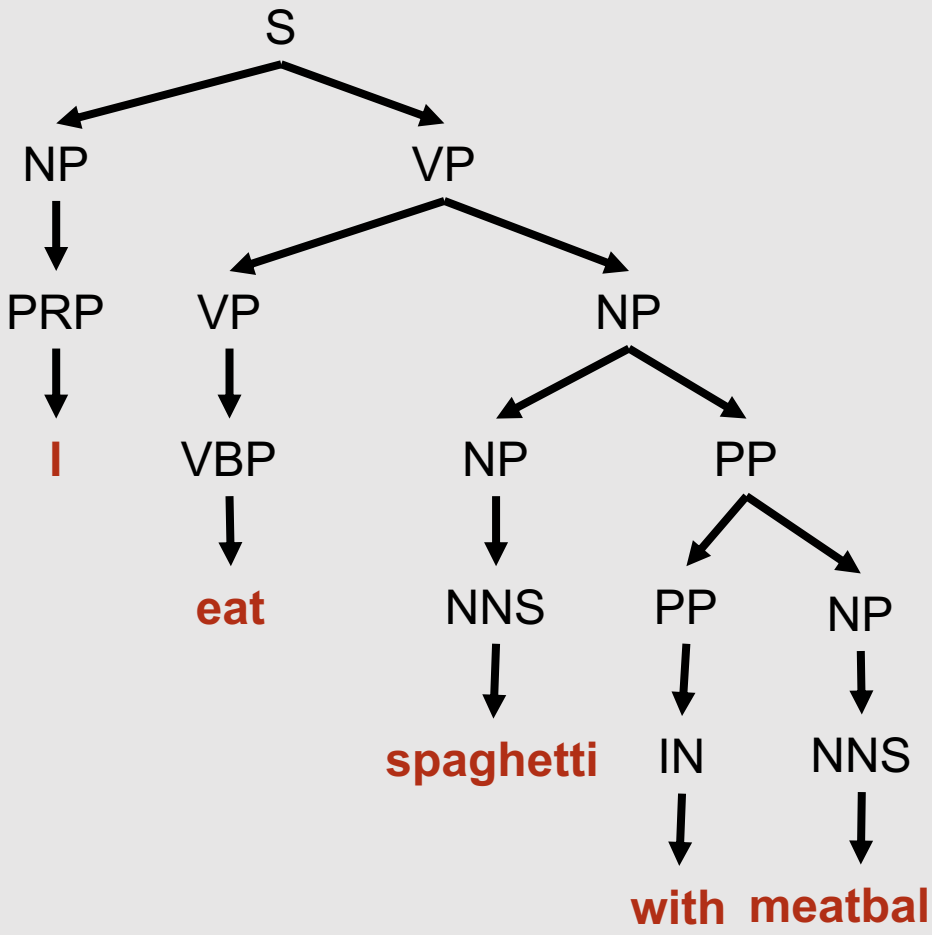
A Little More About Dependents

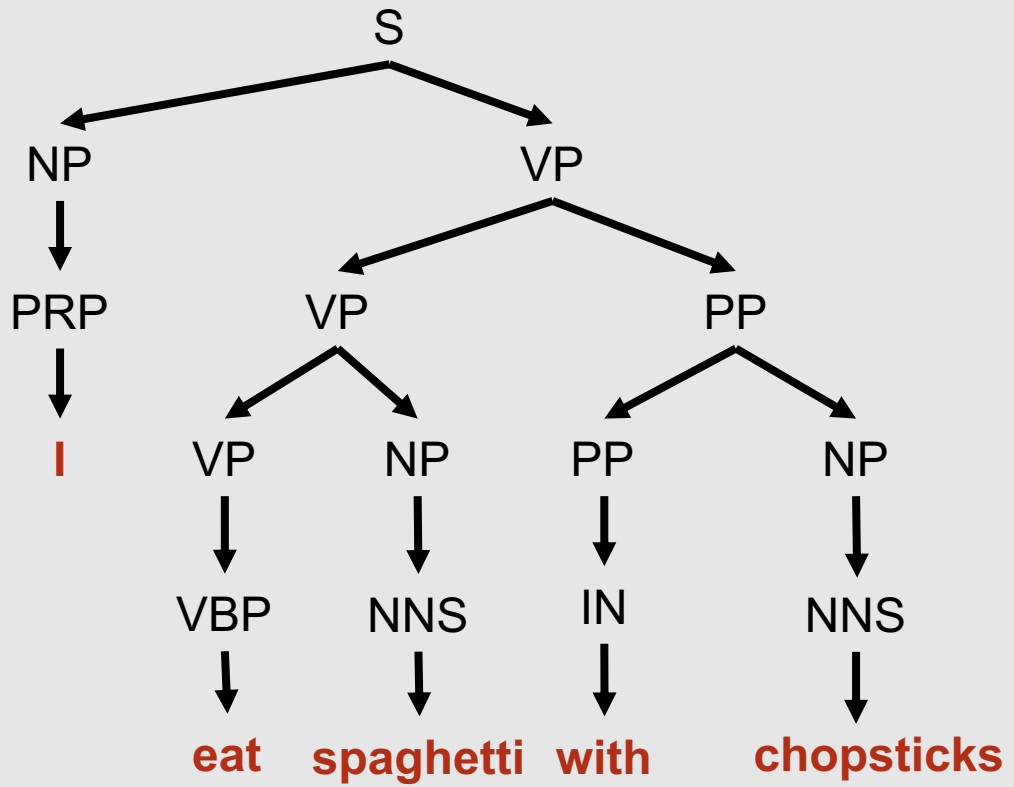
- Dependents can be arguments or adjuncts
- Arguments are **obligatory**
  - Natalie likes *conferences*. 🙂
  - Natalie likes. 😐
- Adjuncts are **optional**
  - Natalie drinks *tea*. 🙂
  - Natalie drinks. 🙂

Properties of Constituents

- **Constituents can be substituted with one another** in the context of the greater sentence
  - **The woman with red hair** rolled her eyes as lightning immediately struck the man's house.
  - **The unicorn** rolled her eyes as lightning immediately struck the man's house.
- **A constituent can move around** within the context of the sentence
  - **The woman with red hair** rolled her eyes as lightning immediately struck the man's house.
  - Lightning immediately struck the man's house as **the woman with red hair** rolled her eyes.
- **A constituent can be used to answer a question** about the sentence
  - Who rolled her eyes? **The woman with red hair**.

The structure of constituents in a tree corresponds to their meaning.





Case Example

- Draw a constituent tree for the sentence:
  - **Time flies like an arrow.**

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	

Case Example

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	

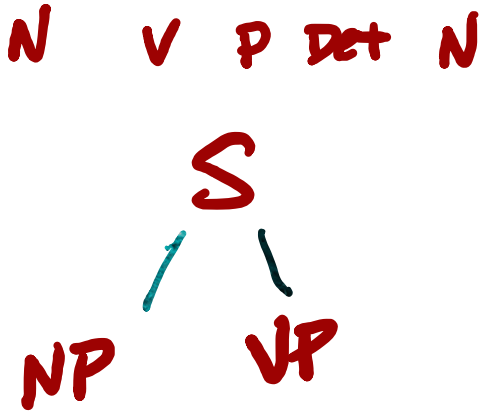
Time flies like an arrow

N V P Det N

Case Example

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	

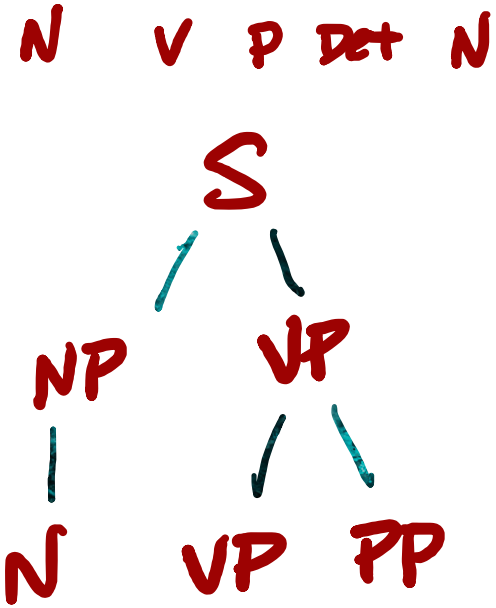
Time flies like an arrow



Case Example

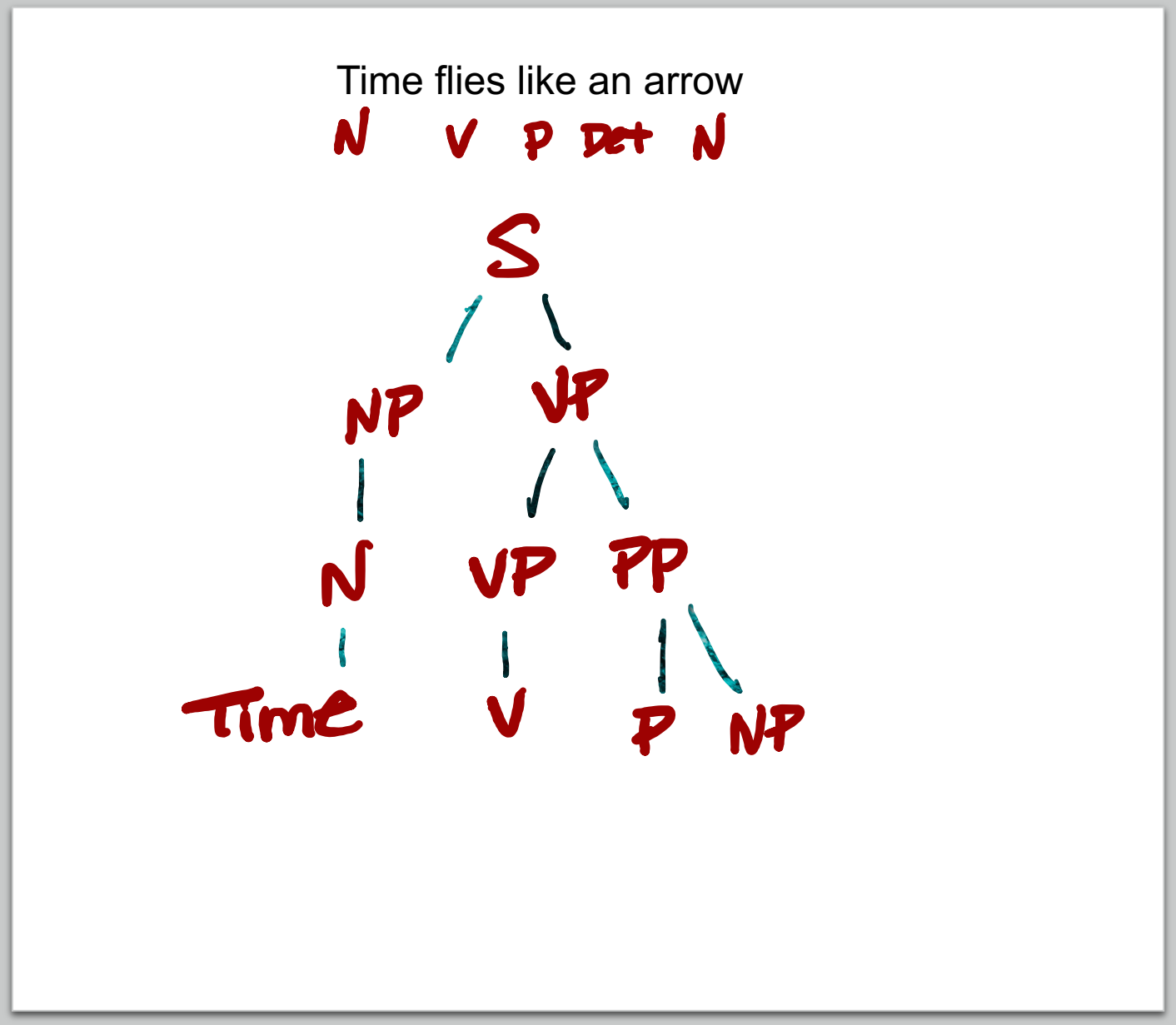
Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	

Time flies like an arrow



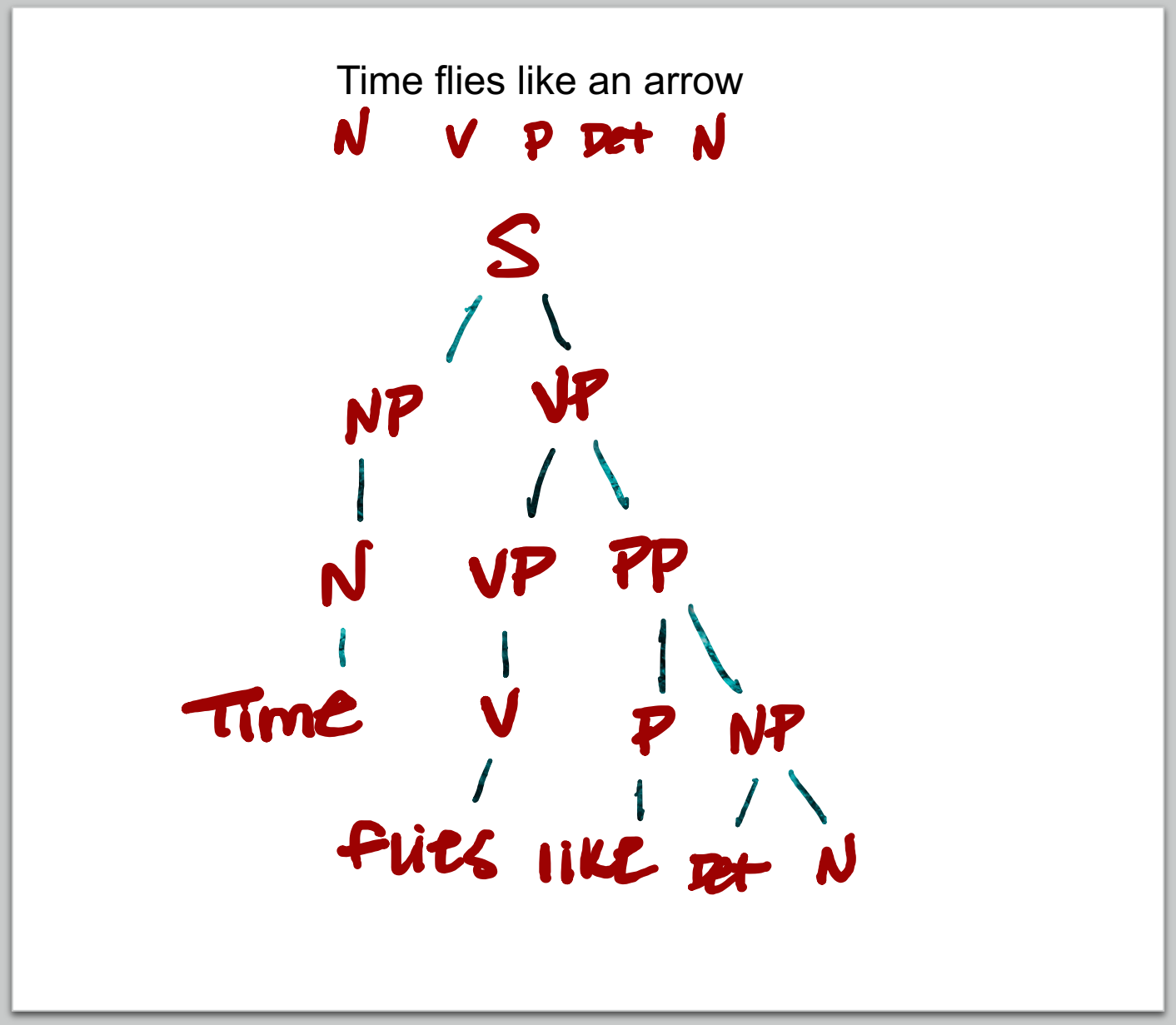
Case Example

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	



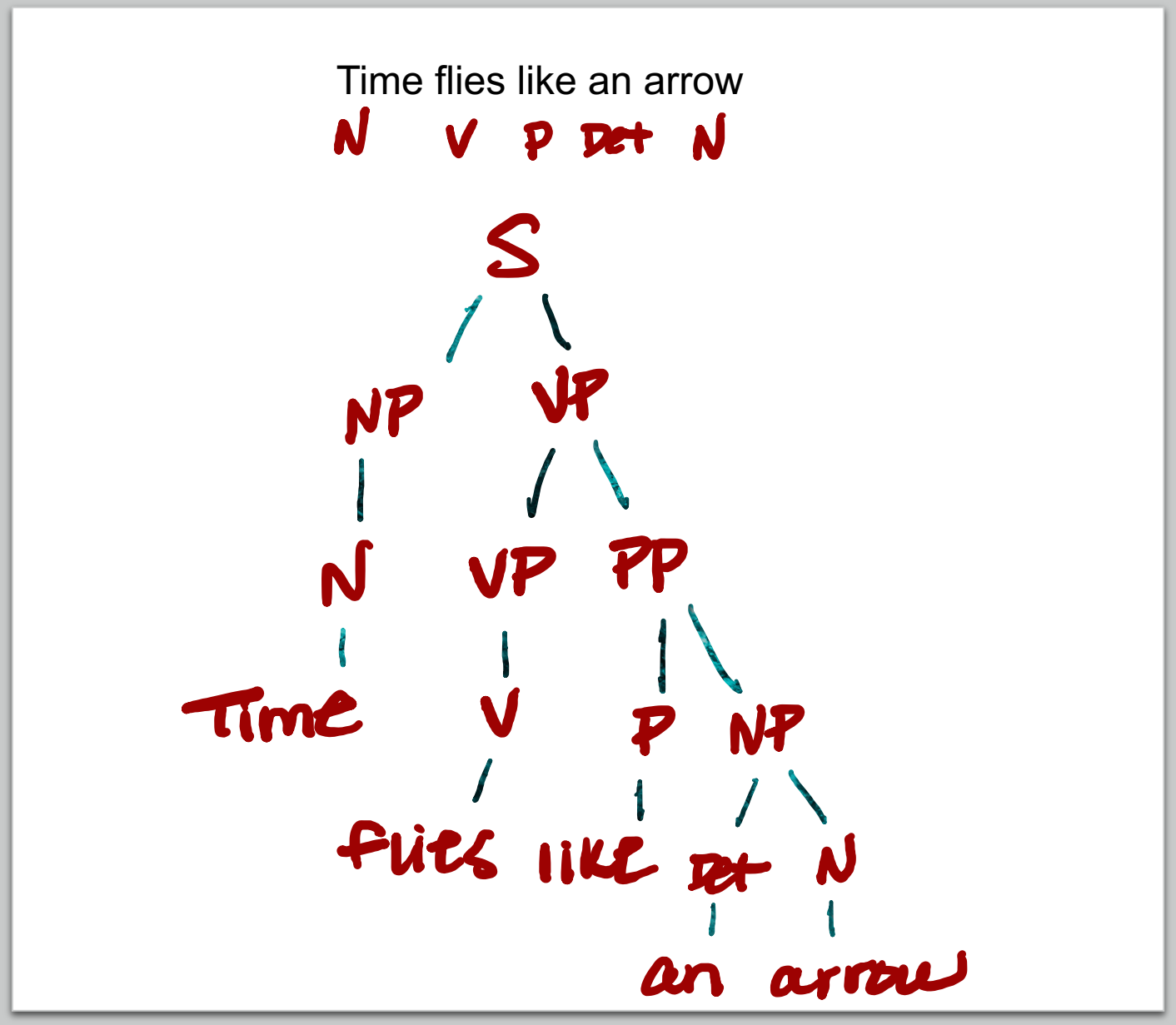
Case Example

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	



Case Example

Production Rules	
S ! NP VP	PP ! P NP
NP ! DET N	PP ! P
NP ! N	P ! like
NP ! N N	V ! flies | like
VP ! VP PP	DET ! a | an
VP ! V NP	N ! time | fruit | flies | arrow | banana
VP ! V	



Formal Definition

- A CFG is a 4-tuple $\langle N, \Sigma, R, S \rangle$ consisting of:
  - A set of non-terminal nodes N
    - N = {S, NP, VP, PP, N, V, …}
  - A set of terminal nodes (leaves) Σ
    - Σ = {time, flies, like, an, arrow, …}
  - A set of rules R
  - A start symbol $S \in N$

Which sentences are grammatically correct?

- Any sentences for which the CFG can construct a tree (all words in the sentence must be reachable as leaf nodes) are accepted by the CFG.

What about really complex sentences?

Natalie knew a lot. 🙂

The zebra **that Natalie knew** knew a lot. 😕

The unicorn **that the zebra that Natalie knew knew** knew a lot. 🤯

CFGs and Center Embedding

- Formally, these sentences are all grammatical, because they can be generated by the CFG that is required for the first sentence:
  - S → NP VP
  - NP → NP RelClause
  - RelClause → that NP ate
- However, very few humans would consider the last sentence to be grammatically correct!

CFGs and Center Embedding

- **CFGs are unable to capture bounded recursion** (e.g., embedding only one relative clause)
- So, linguists acknowledge that formal grammaticality is not perfectly equivalent to human perception of grammaticality
  - They additionally consider human grammatical knowledge, as well as processing and memory limitations
- In the context of this class, we'll just assume that if something is accepted by a CFG, it is grammatically correct