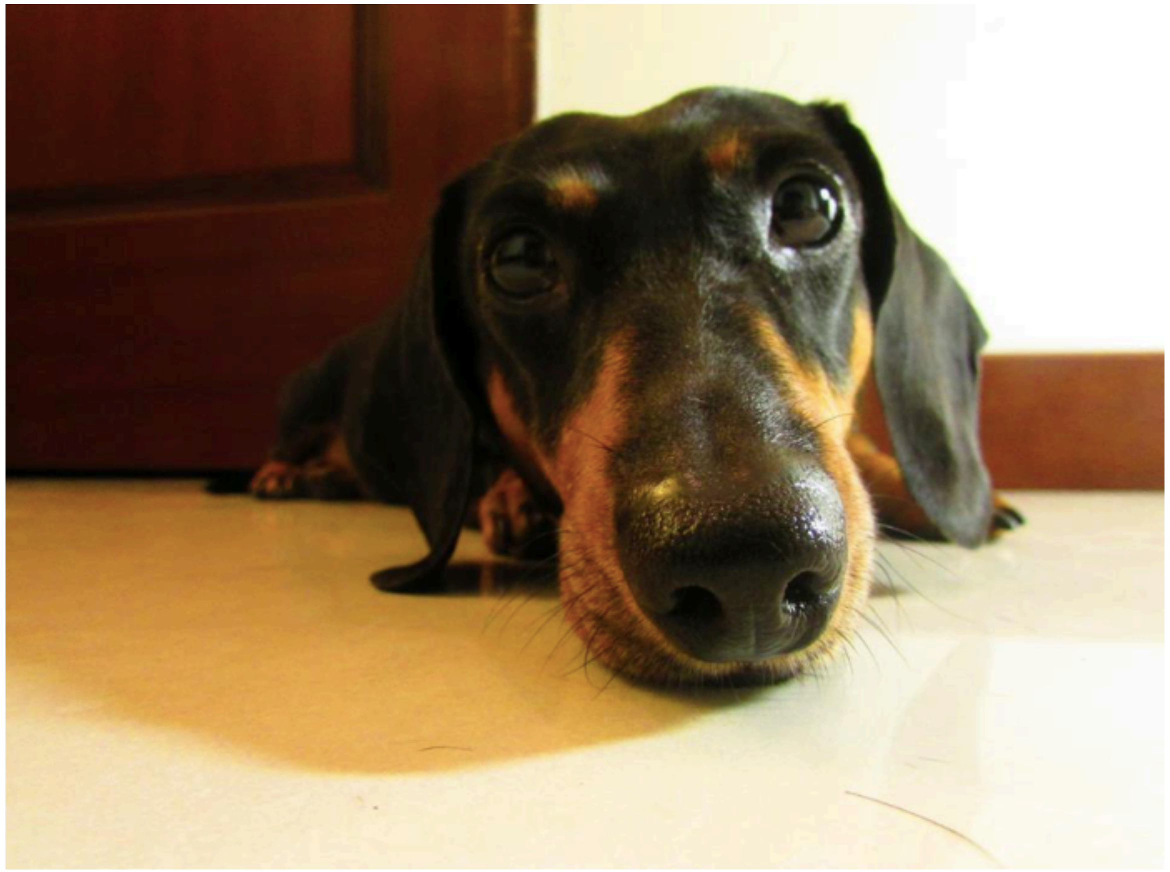
Modeling Grammaticality

Kevin Duh
Fall 2019

Today’s Outline

1. What does it mean to be grammatical?
2. Modeling grammaticality with Context-Free Grammars

Is this grammatical?

- 

a cute dog
a very cute dog
super cute puppy
adorable puppy looking at me
....

dog cute a
dog cut a very
puppy cute super
me at puppy looking adorable at
....

Grammar and Syntax

- Grammar: Formal rules, principles, or processes that determine valid and invalid structure in language
- Syntax: Grammar of sentences
  - (We'll focus on this today)

Prescriptive vs Descriptive

- Prescriptive Grammar
  - How you “ought” to speak. Otherwise, you’re ungrammatical!
  - e.g. Don’t split infinitives! (e.g. “to go”)
- Descriptive Grammar
  - Focus on describing the language as it’s used
  - e.g. “<u>To boldy go</u> where no man has gone before”
- In NLP, we do a bit of both, with probabilities

Grammaticality in perspective

- Dialect differences:
  - I didn’t eat dinner
  - I didn’t eat no dinner
- Changes in usage:
  - She said, “I want to go!”
  - She was, like, “I want to go!”

“Chinese has no grammar” — false!

- 他喝茶 (Literal: He drinks tea)
    - Grammar rule: Subject, Verb, Object
- 我有一件黑色襯衫 (I have a black shirt)
    - Grammar rule: modifier “black” comes before modified
- All languages have grammar!

Today's Outline

1. What does it mean to be grammatical?
2. Modeling grammaticality with Context-Free Grammars

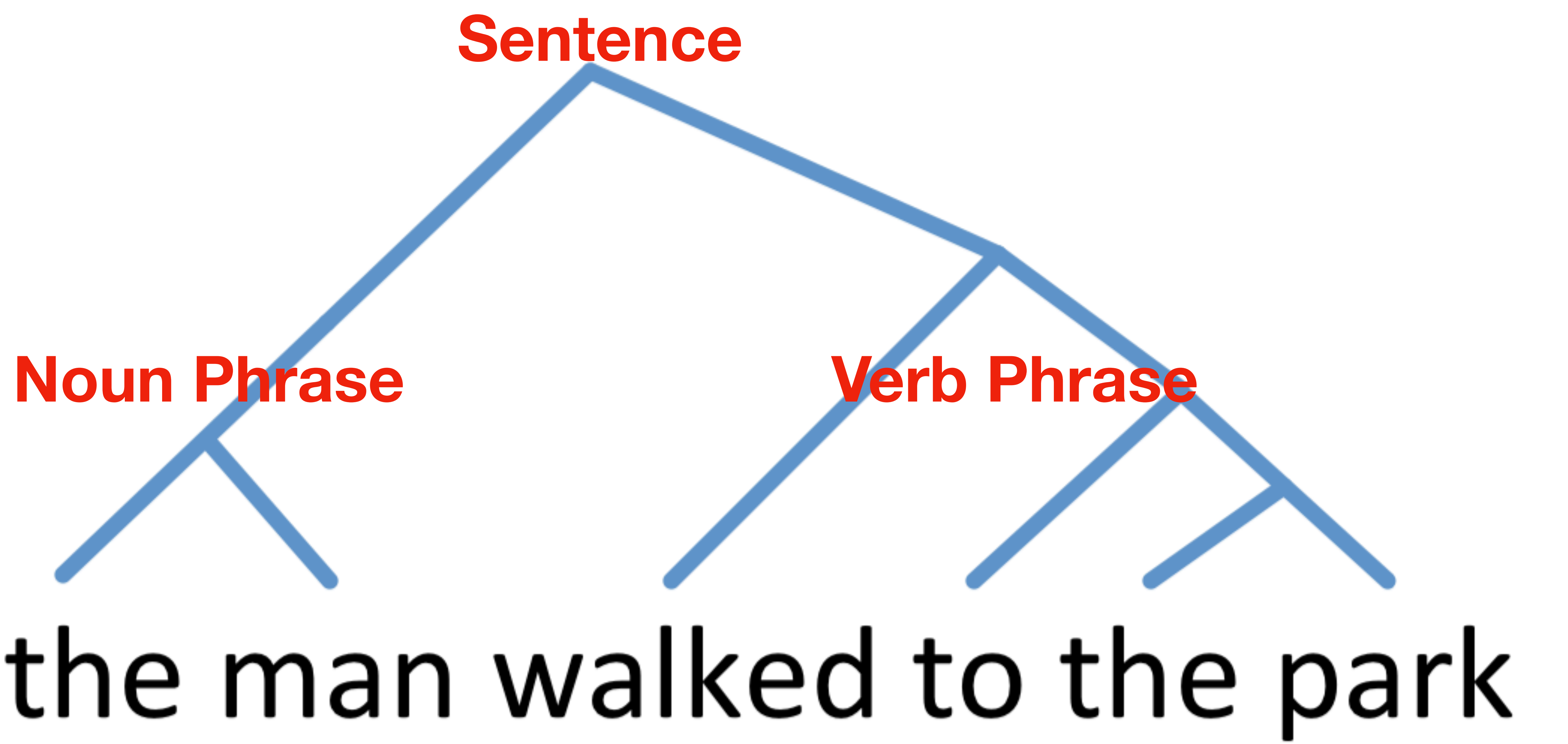
Goal of modeling grammaticality

- We need a way to mathematically or formally capture what is grammatical and what is not.
- There are many “formalisms” for doing so. We’ll cover:
  - Constituency grammar (today)
  - Dependency grammar (later)

Constituency grammar

(aka phrase structure grammar)

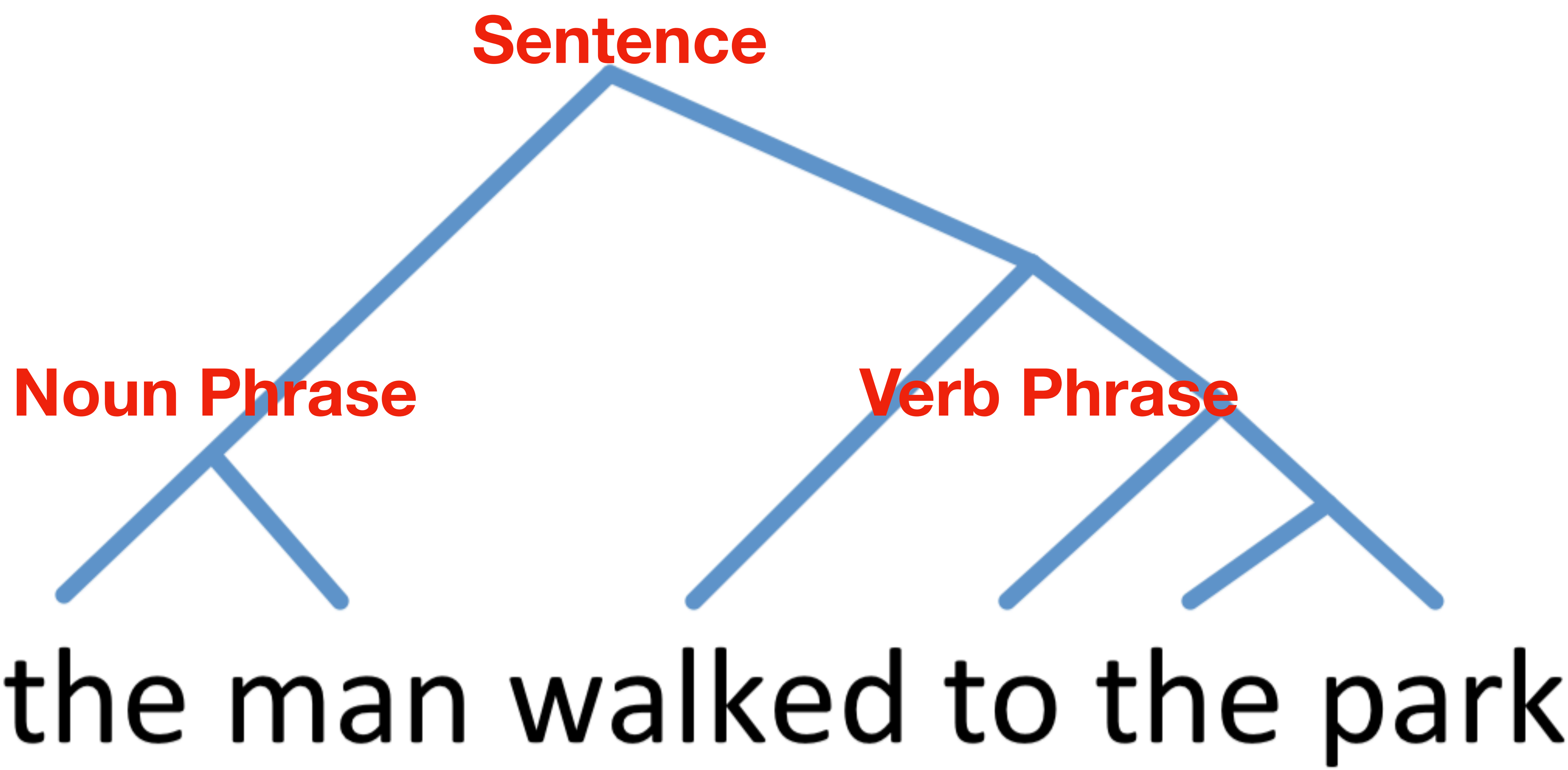
- Focuses on groups of words (constituent)
- A sentence (S) is made of:
  - subject, typically a noun phrase (NP)
  - predicate, typically verb phrase (VP)
- NP and VP are in turn made of groups of words



From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

Bracketing notation:

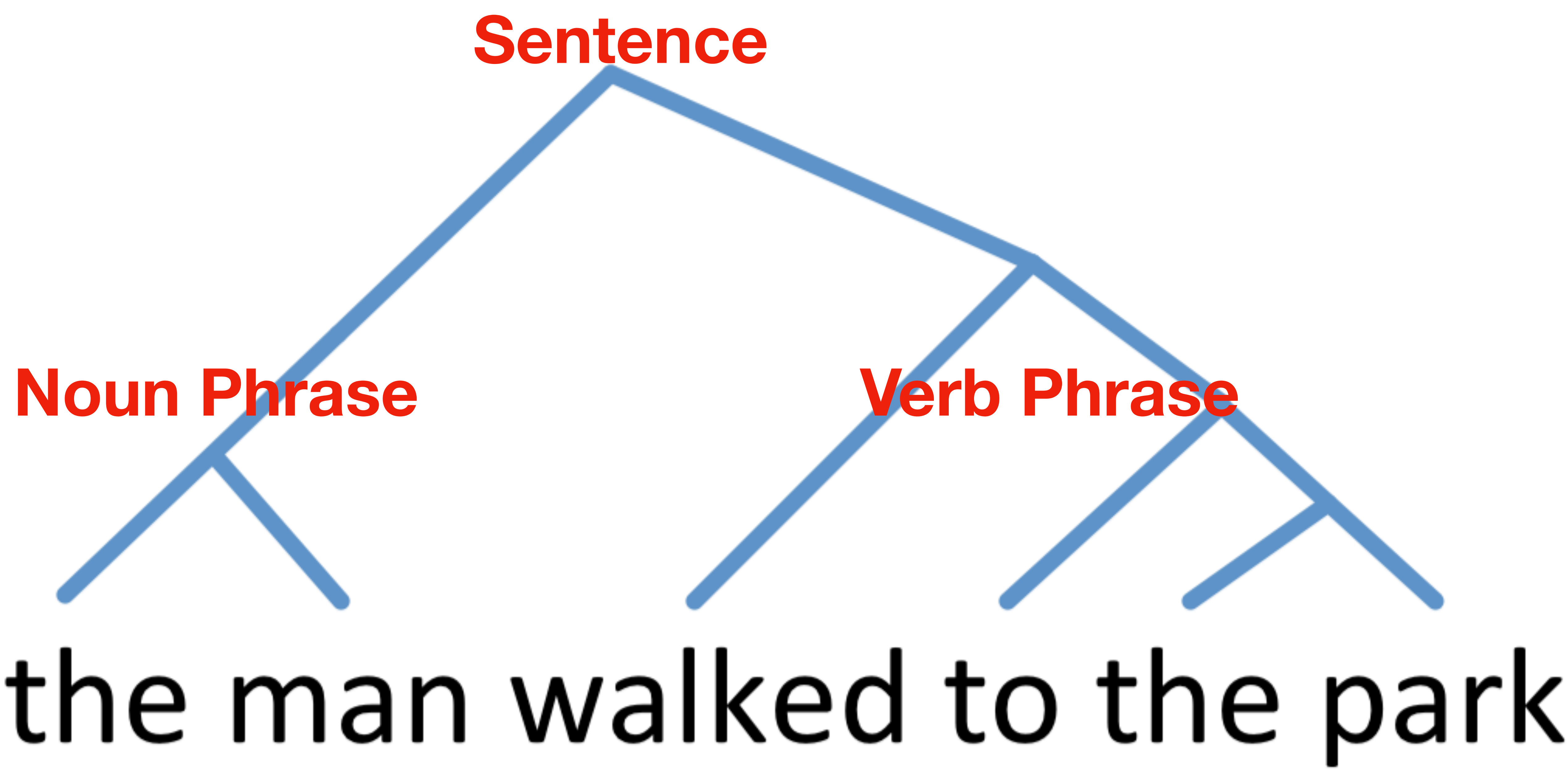
((the man) (walked to the park))



From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

Bracketing notation:

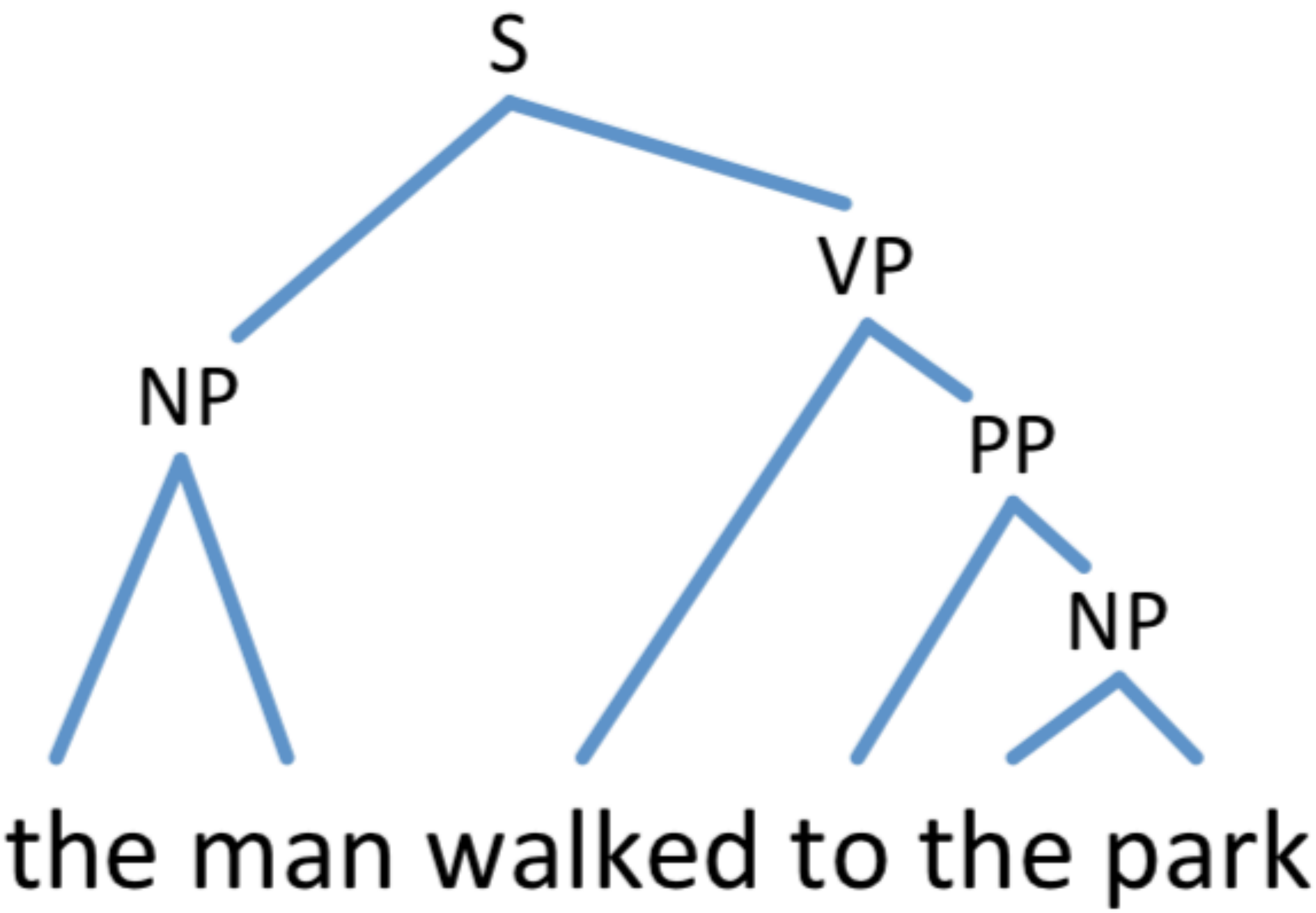
((the man) (walked (to (the park))))



From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

Add labels to each constituent

(S (NP the man) (VP walked (PP to (NP the park))))

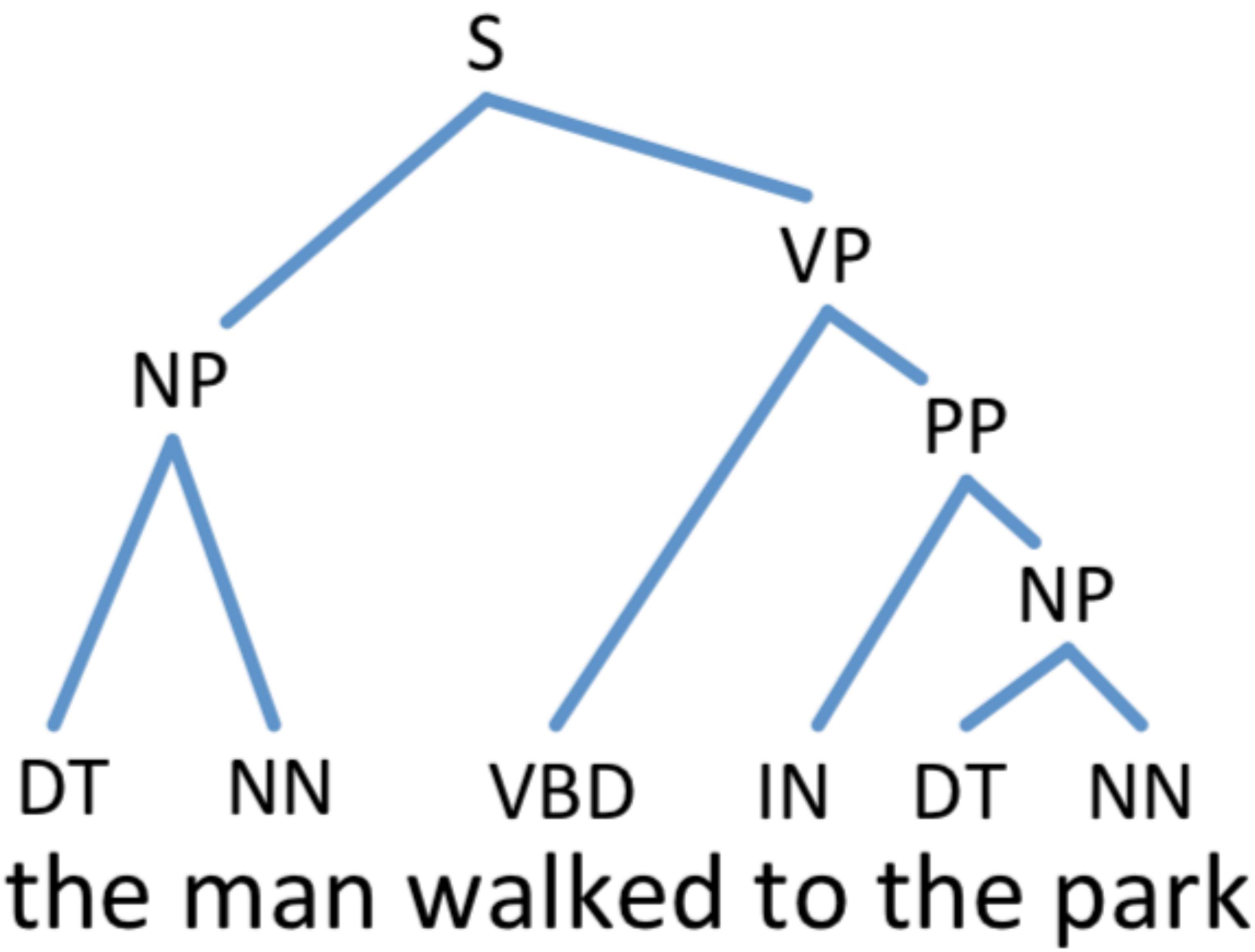


Key:
S = sentence
NP = noun phrase
VP = verb phrase
PP = prepositional phrase

From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

Include pre-terminals (part-of-speech labels)

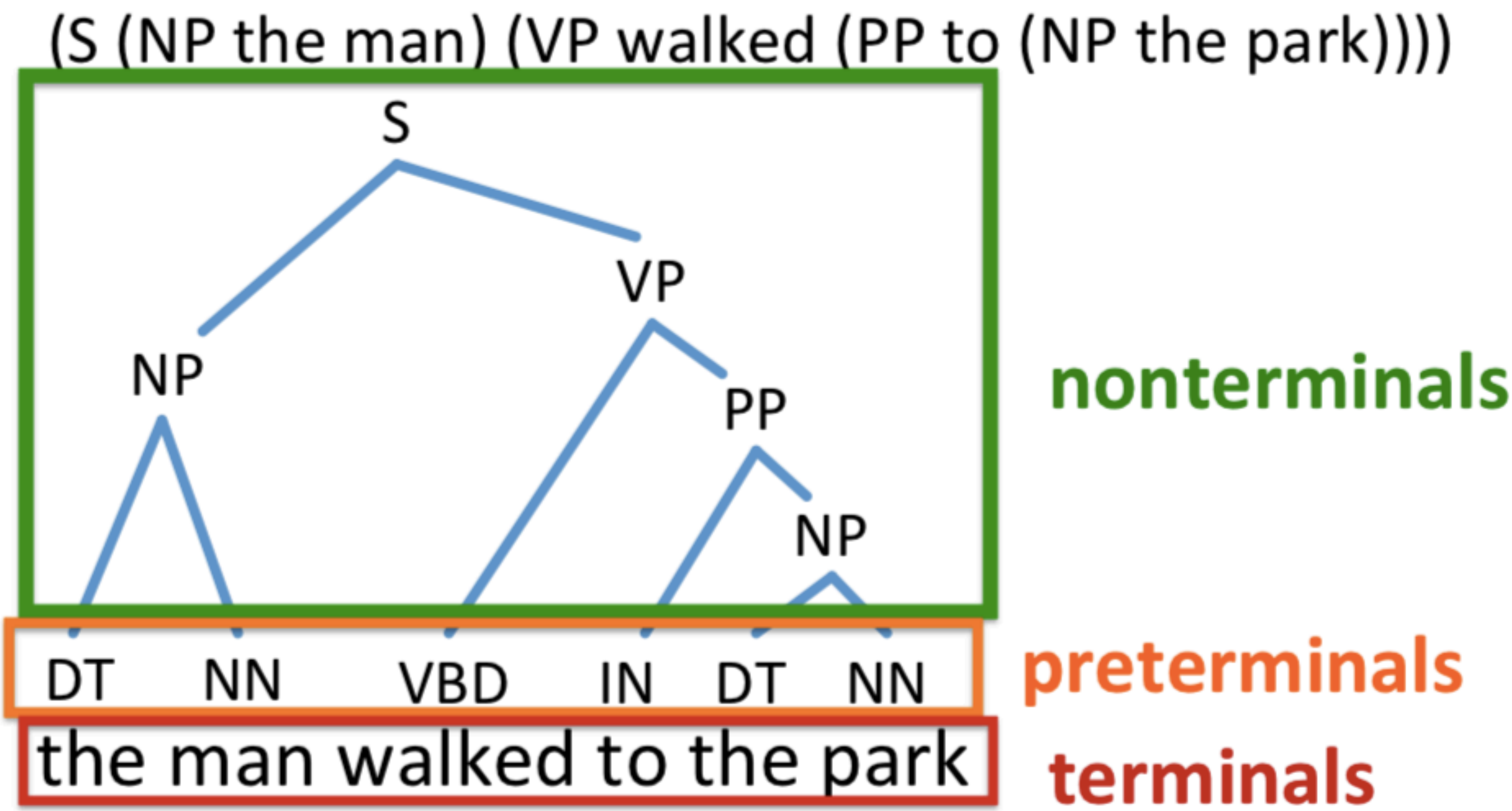
(S (NP the man) (VP walked (PP to (NP the park))))



Key:
S = sentence
NP = noun phrase
VP = verb phrase
PP = prepositional phrase
DT = determiner
NN = noun
VBD = verb (past tense)
IN = preposition

From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

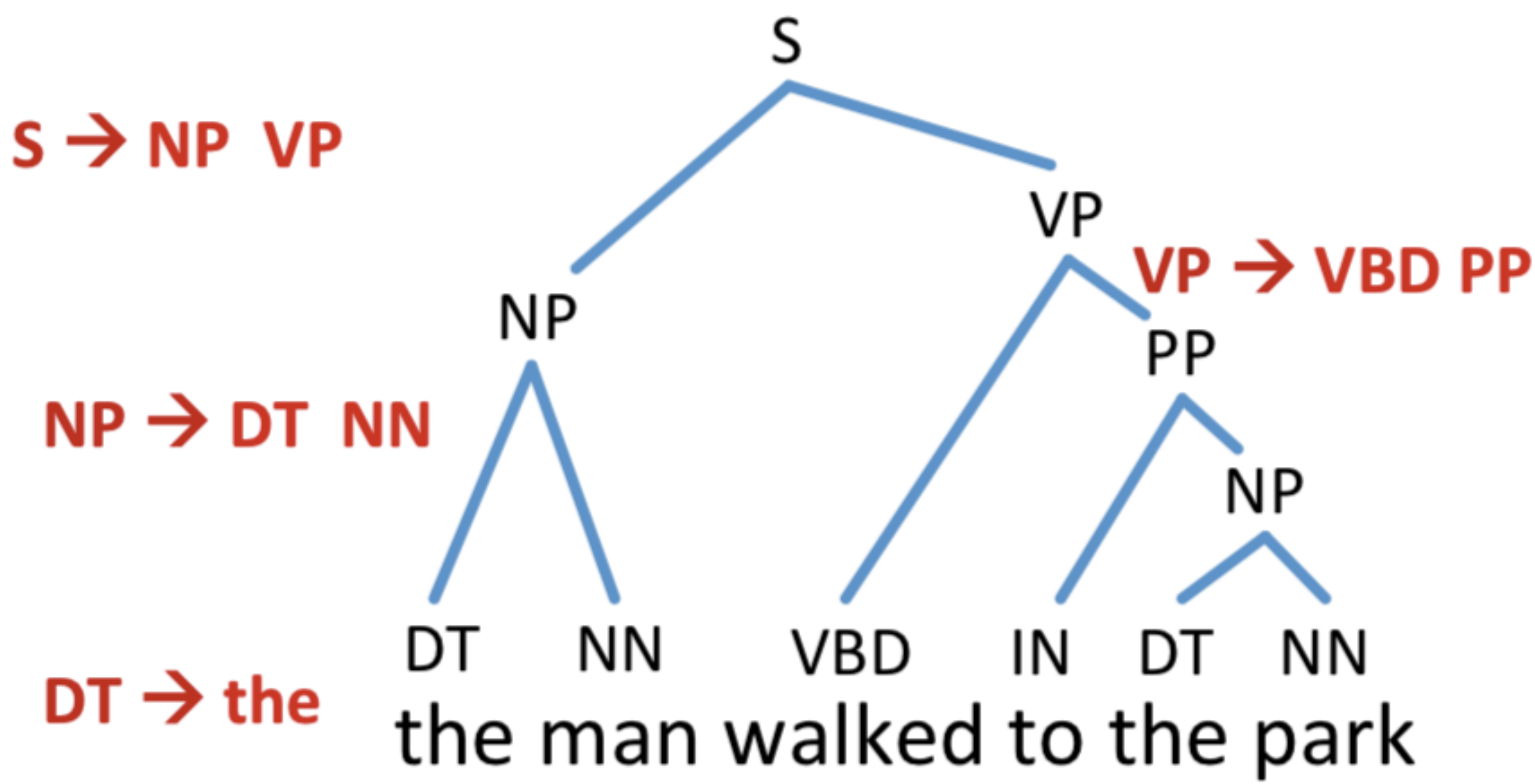
Now we have a constiuency tree!



From: Kevin Gimpel, https://ttic.uchicago.edu/~kgimpel/teaching/31190/lectures/12.pdf

Context-Free Grammar

- Syntactic Re-write Rules
  - S —> NP VP
  - NP —> DT NN
  - VP —> VBD PP
  - PP —> IN NP
  - etc
- Lexical Re-write Rules
  - NN —> man
  - DT —> the
  - VBD —> walked
  - IN —> to
  - NN —> park


S
S → NP VP
VP
VP → VBD PP
NP
PP
NP → DT NN
NP
DT
NN
VBD
IN
DT
NN
DT → the
the man walked to the park

Probabilistic CFG

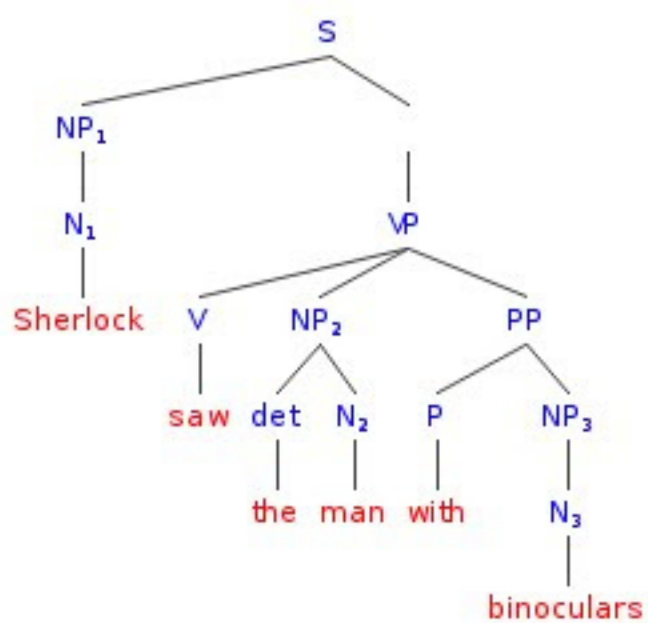
- Syntactic Re-write Rules
  - S —> NP VP **Probability=1.0**
  - NP —> DT NN **Probability=0.7**
  - VP —> VBD PP **Probability=1.0**
  - PP —> IN NP **Probability=1.0**
  - NP —> NNP **Probability=0.3**
  - etc

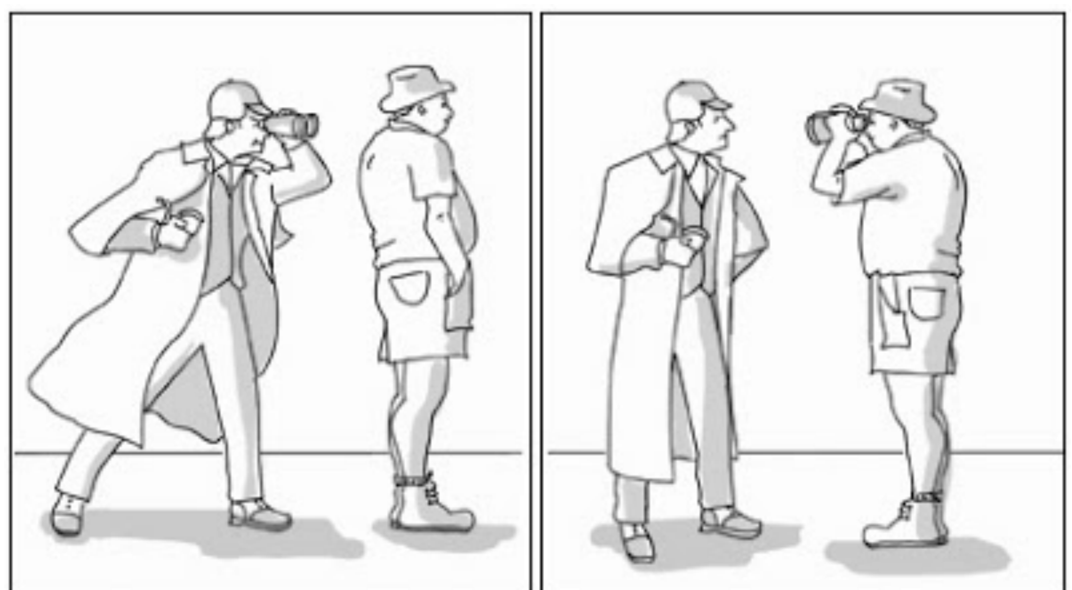
- Lexical Re-write Rules
  - NN —> man **Probability=0.4**
  - DT —> the **Probability=1.0**
  - VBD —> walked **Probability=1.0**
  - IN —> to **Probability=1.0**
  - NN —> park **Probability=0.4**
  - NN —> John **Probability=0.2**

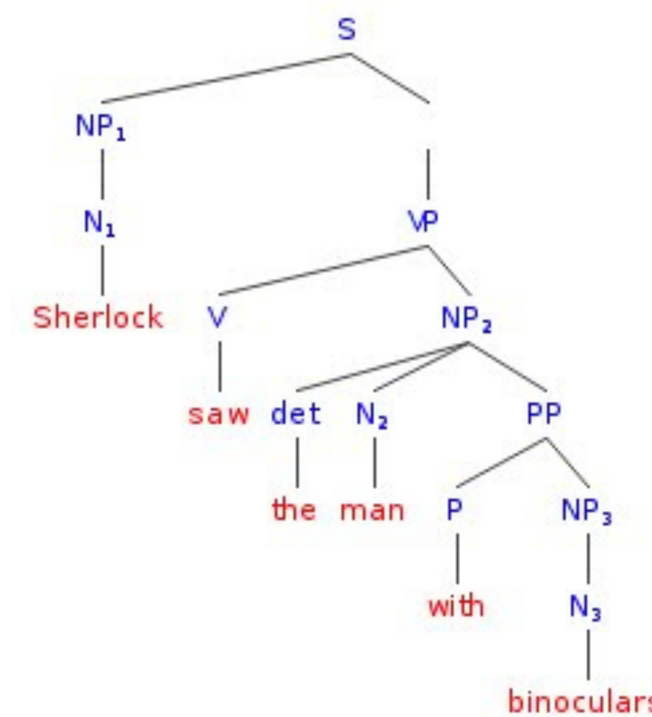
Top-down generation

Ambiguities - Prepositional Phrase (PP) Attachment

Sherlock saw the man using binoculars




Sherlock saw the man using binoculars. | Sherlock saw the man using binoculars.



From: http://walkinthewords.blogspot.com/2010/07/syntax-with-sherlock-sentence-ambiguity.html

Ambiguities - more examples

- Coordination:
  - ((laptop and monitor) with the Apple logo)
  - (laptop and (monitor with the Apple logo))
- Noun compound
  - ((Natural Language) Processing)
  - (Natural (Language Processing))

CFG Formalism

- **G=(Σ,N,S,R)**
  - Σ is finite set of terminal, e.g. a, b
  - N is finite set of nonterminal, e.g. A, B (V = Σ U N)
  - S is start symbol
  - R is production rule A —> α where α is V*
  - For PCFG, probability is attached to each R
- **Chomsky Normal Form (CNF) — only these rules are allowed**
  - unary terminal rule A —> w
  - binary nonterminal rule A —> B C

Why is it called context-free?

- A rule like NP —> DT NN applies regardess of the neighboring context of NP
- i.e. left-hand-side of each rule is a single non-terminal symbol

Today's Outline

1. What does it mean to be grammatical?
2. Modeling grammaticality with Context-Free Grammars