

Grammars and Parsing: Part 1

Acknowledgements

I have used the following sources for preparing these slides:

1. Statistical Techniques for Natural Language Parsing, AI Mag article by Charniak, AI Magazine Volume 18 Number 4 (1997) (© AAI)
2. J&M Textbook
3. <http://www.classes.cs.uchicago.edu/archive/2006/winter/35100-1/slides/class3.ppt>
4. http://www.rci.rutgers.edu/~cfs/305_html/Understanding/Parsing.html
5. <http://www.cis.upenn.edu/~cis530/slides-2008/TAGs-2008.ppt>
6. www.cis.upenn.edu/~cis530/slides-2008/530-cfparsing-2008.ppt
7. Pinker's non-fiction "The Language Instinct"

Parsing

the man took the book

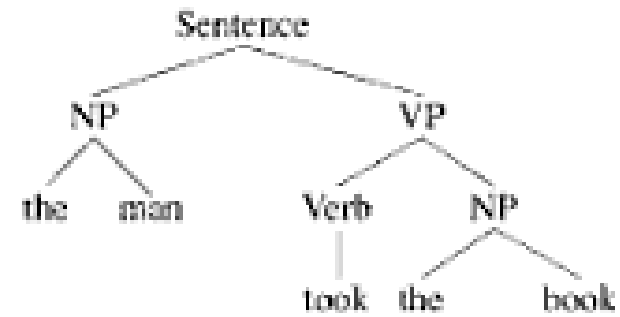
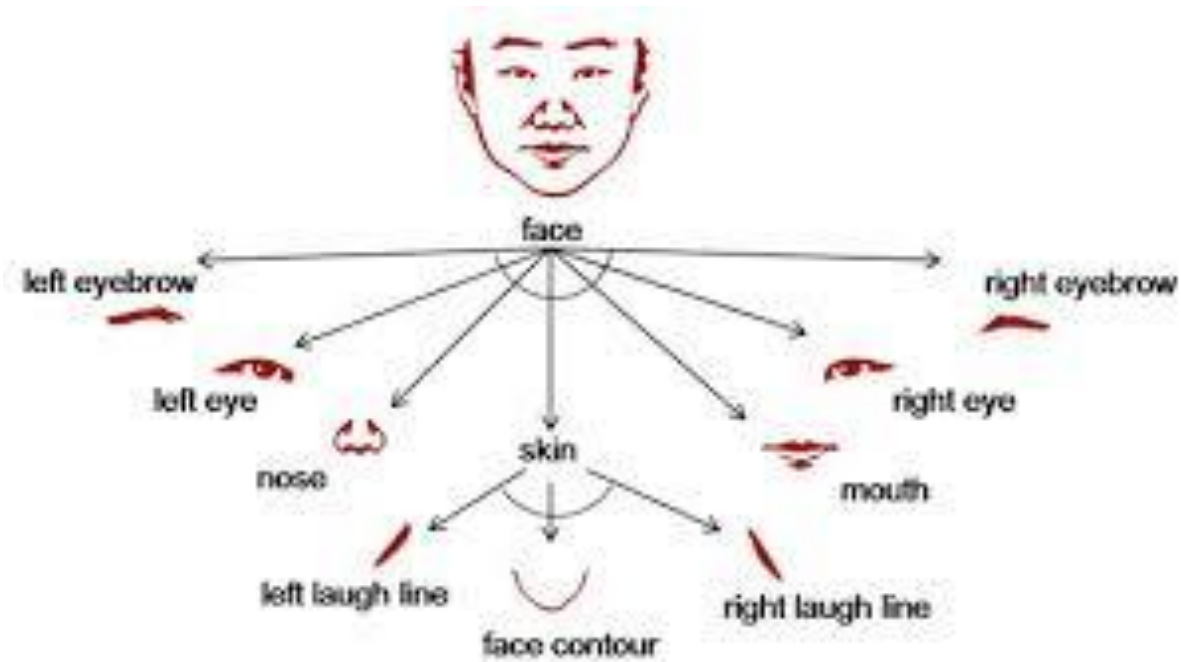


Image Parsing



Ack: http://www.stat.ucla.edu/~sczhu/Project_pages/Paper-cut/Paper-cut.htm

Original paper: [Xu Z¹](#), [Chen H](#), [Zhu SC](#), [Luo J](#), **A hierarchical compositional model for face representation and sketching** ([IEEE Trans Pattern Anal Mach Intell.](#) 2008 Jun;30(6):955-69. doi: 10.1109/TPAMI.2008.50)

Ambiguity: Parsing is useful in determining meaning



Ack: <http://specgram.com/CLIII.4/08.phlogiston.cartoon.zhe.html>

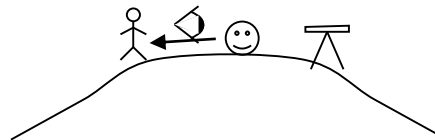
Ambiguity: Parsing is useful in determining meaning



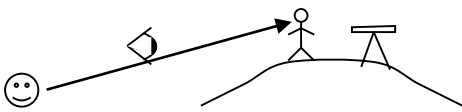
Ack: <http://specgram.com/CLIII.4/08.phlogiston.cartoon.zhe.html>

Ambiguity: Parsing is useful in determining meaning

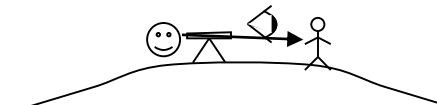
I saw the man on the hill with the telescope




“I was on the hill that has a telescope when I saw a man.”



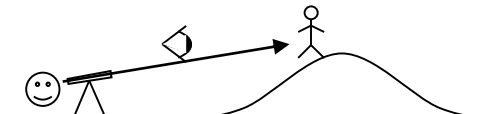
“I saw a man who was on the hill that has a telescope on it.”



“I was on the hill when I used the telescope to see a man.”



“I saw a man who was on a hill and who had a telescope.”

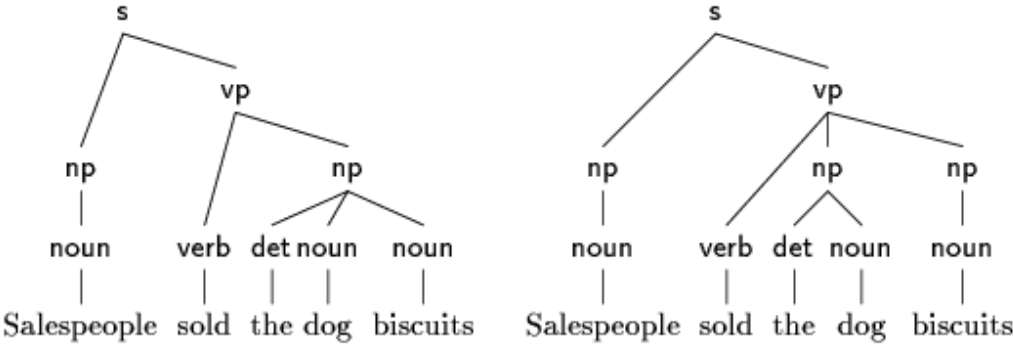


“Using a telescope, I saw a man who was on a hill.”

...

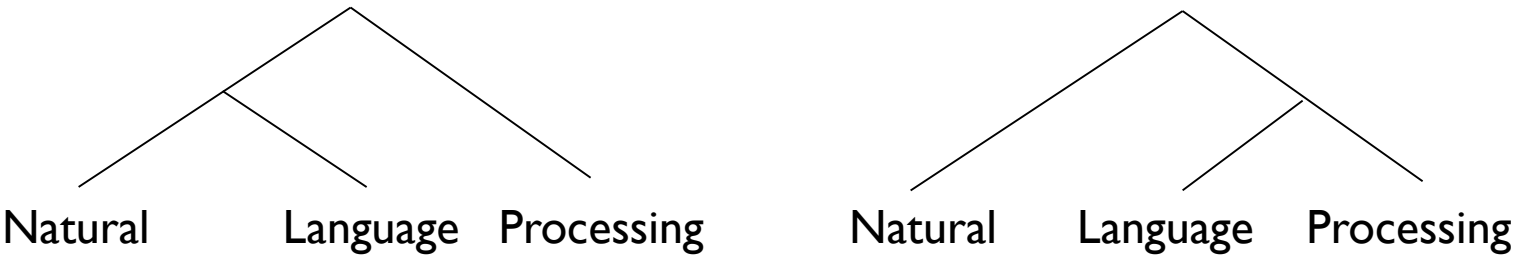
☺ Me 🔭 See 🧑 A man 📡 The telescope 🏞 The hill

Ambiguity: Parsing is useful in determining meaning



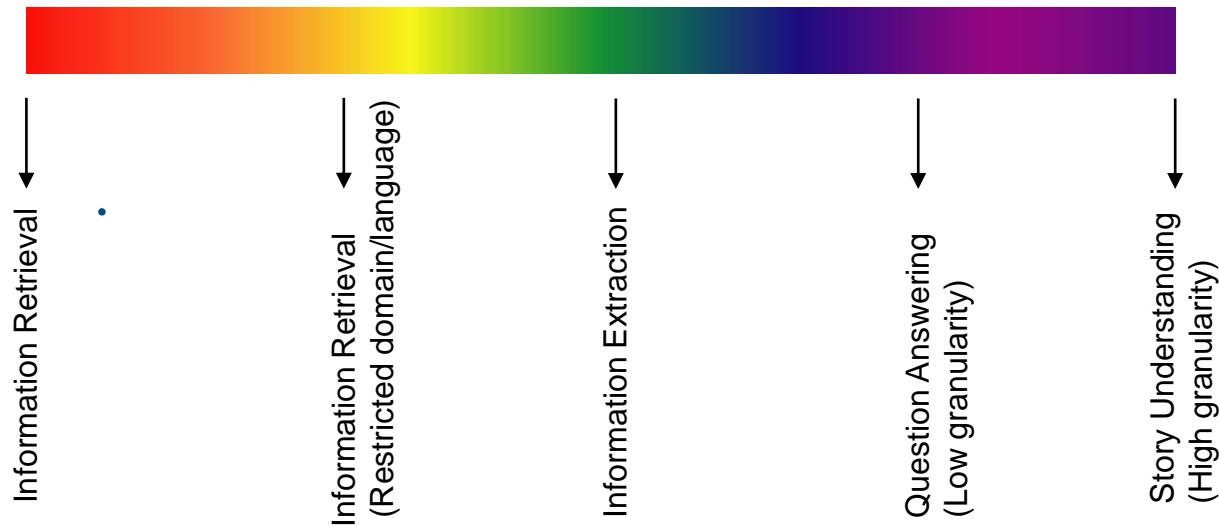
A third interpretation?

Statistical Techniques for Natural Language Parsing, AI Mag article by Charniak, AI Magazine Volume 18 Number 4 (1997) (© AAAI)



To parse or not to parse

- Not all NLP applications may need accurate parsing.
- Factors : Granularity, Volume, Domain Specificity



References

- For slides to follow: From beginning till Section 9.4 of the chapter from an earlier edition of J&M (shared)
- Suggested self study: Section 9.5, 9.6 and 9.7 (may cover briefly in next class), and some preliminaries of Chomsky hierarchy

Properties of words

- Why are Parts of Speech important ?
 - Gives significant amount of information about the word and its neighbours
 - Example : personal pronouns (I, you, me) and possessive pronouns (my, your, his, her, its)
 - Morphology
 - Word Sense Disambiguation
 - Partial Parsing in Information Extraction
 - Pronunciation of words
 - Linguistic Research
- One categorization : Closed and open classes

Open Classes

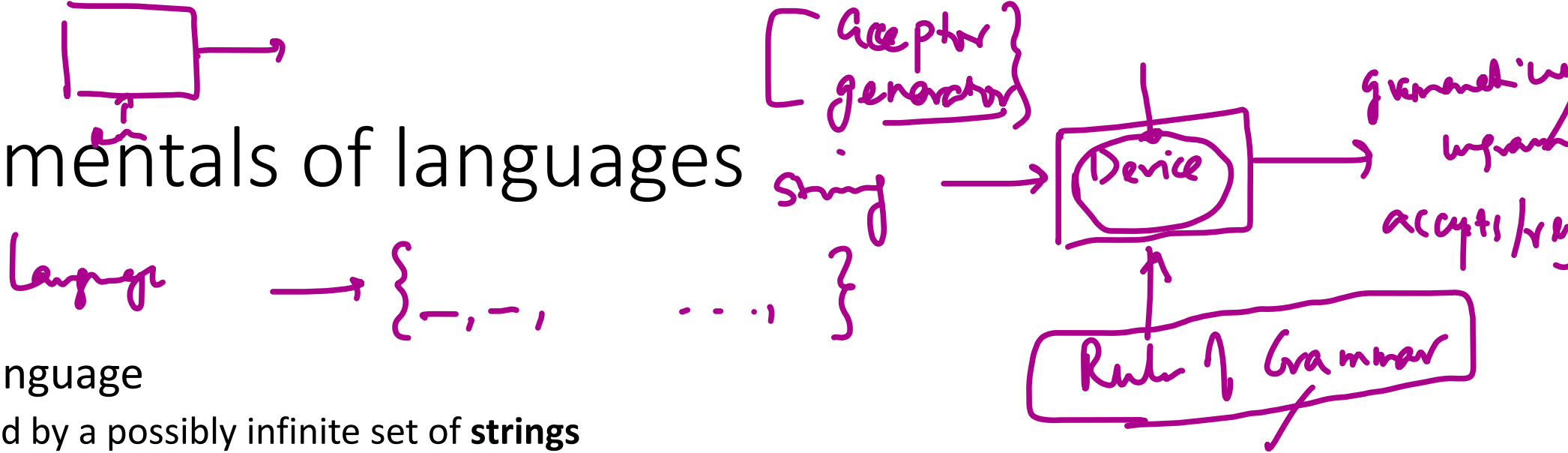
- Four major open classes: nouns, verbs, adjectives, adverbs
- Nouns
 - Proper nouns, common nouns
 - Count nouns, mass nouns
- Verbs
- Adjectives (colour, age, value)
- Adverbs
 - Directional/locative (downhill, here)
 - Degree (extremely)
 - Manner (delicately)
 - Temporal (yesterday)

Unfortunately, John walked home extremely slowly yesterday.

Closed Classes

- Prepositions : on, under, over
- Determiners : a, an, the
- Pronouns : she, who, I, others
- Conjunctions : and, but, or, if
- Auxiliary verbs : can, may, should, are
- Particles : up, down, on, off, in
- Numerals : one, two, three, first, second, third
- Additional classes:
 - Interjections (oh, ah, hey, man)
 - Negatives (no, not)
 - Politeness markers (please, thank you)
 - Greetings (hello, goodbye)
 - Existential *there*

Fundamentals of languages



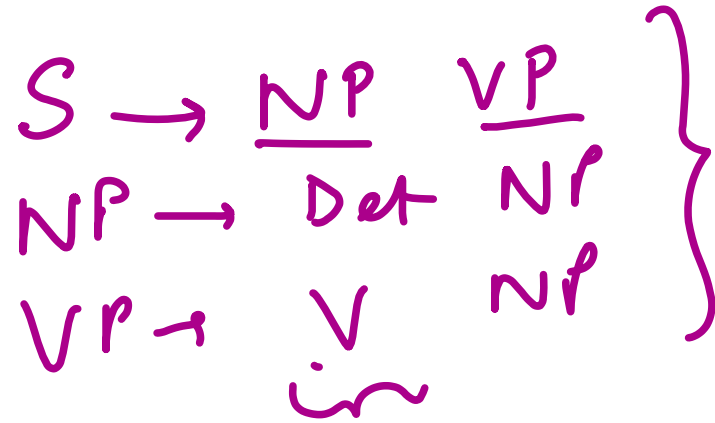
- Formal language

- Defined by a possibly infinite set of **strings**
- Each string is a concatenation of **terminal symbols**, sometimes called words
- E.g. when using FOL, with terminal symbols P, Q and \wedge , string " $P \wedge Q$ " is a member of the language, but " $\wedge P Q$ " is not.
- Formal languages like FOL and Java have strict mathematical definitions. In contrast, **natural languages** like Chinese, Danish and English have no strict definition, but are used by a community of speakers.
- We will attempt to treat natural languages as if they were formal languages

✓ Rajeev Sangat, Vineet Chaturanya.

Syntax / Semantics.

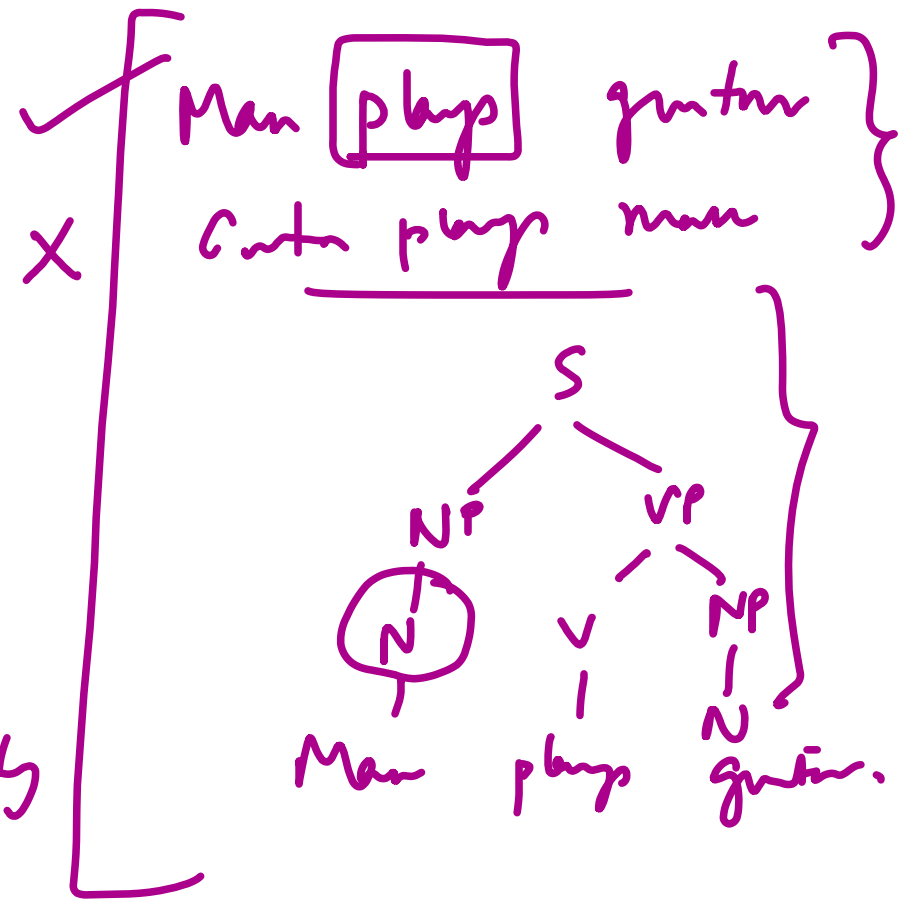
Part of speech



{ hardware / software

low - 150

↓
higher
granularity
↓



Grammar

- Syntax: from Greek syntaxis, “setting out together, arrangement”
- Refers to the way words are arranged together, and the relationship between them.
- Goal of syntax is to model the knowledge of that people unconsciously have about the grammar of their native language
- Important in NLP : Grammar checkers, Question answering systems, Information extraction, Machine translation

Grammar

- A finite set of rules that specifies the language
- Formal languages always have an official grammar
- Linguists are scientists who attempt to discover properties of language by a process of scientific inquiry and codifying their discoveries in a grammar.
 - Note similarities with describing nature using a set of physical laws.
- Till date no linguist has succeeded completely.
- Linguists are also prescriptive – they try to dictate how the language should be
 - don't split infinitives" [*The population is expected **to more than double** in the next ten years, to **boldly go***]

Three key ideas

- Constituency
 - Noun phrases like “a thoroughly entertaining movie”
- Grammatical Relations
 - Subject-object relations
- Subcategorization and Dependency Relations
 - “I *want* to fly to Detroit” : but *find* cannot be followed by an infinitive

Context-Free Grammars

- Capture constituency and ordering
 - Ordering:
 - What are the rules that govern the ordering of words and bigger units in the language?
 - Constituency:
How words group into units and how the various kinds of units behave

Constituency

- Noun phrases (NPs)
 - Three parties from Brooklyn
 - A high-class spot such as Mindy's
 - The Broadway coppers
 - They
 - Harry the Horse
 - The reason he comes into the Hot Box
- How do we know these form a constituent?
 - They can all appear before a verb:
 - Three parties from Brooklyn arrive...
 - A high-class spot such as Mindy's attracts...
 - The Broadway coppers love...
 - They sit...

How do we know these form a constituent?

- They can all appear before a verb:
 - Three parties from Brooklyn arrive...
 - A high-class spot such as Mindy's attracts...
 - The Broadway coppers love...
 - They sit ...
- But individual words can't always appear before verbs:
 - *from arrive...
 - *as attracts...
 - *the is
 - *spot is...
- Must be able to state generalizations like:
 - Noun phrases can occur before verbs

How do we know these form a constituent?

- Preposing and postposing:
 - On September 17th, I'd like to fly from Atlanta to Denver
 - I'd like to fly on September 17th from Atlanta to Denver
 - I'd like to fly from Atlanta to Denver on September 17th.
- But not:
 - *On September, I'd like to fly 17th from Atlanta to Denver
 - *On I'd like to fly September 17th from Atlanta to Denver

Phrase Structure

- Most grammar formalisms are based on the idea of **phrase structure**
 - Strings are composed of substrings called **phrases**, which come in different categories
 - One such category is the **noun phrase (NP)**, examples are “the wumpus”, “the king”, “the agent in the corner”
- Two reasons for identifying phrases this way:
 - Phrases correspond to natural semantic elements from which the meaning of an utterance can be constructed, e.g. **noun phrases refer to objects in the world**
 - Categorizing phrases helps us to describe the allowable strings of the language
 - We can say that noun phrase combines legally with a **verb phrase (VP)**, such as “is dead” to form a phrase of category **sentence (S)**. Without the notions of NP and VP, we could not say why “**The wumpus is dead**” is a sentence while “**wumpus the dead is**” is not.

Key Constituents (English)

- Sentences
- Noun phrases
- Verb phrases
- Prepositional phrases

Prepositional Phrases (PPs)

- PP → Preposition NP
 - From LA
 - To Boston
 - On Tuesday
 - With lunch
 - In the large cardboard box

CFG Rules: Examples

- $S \rightarrow NP VP$
- $NP \rightarrow Det Noun$
- $VP \rightarrow Verb$
- $Det \rightarrow a$
- $Noun \rightarrow flight$
- $Verb \rightarrow left$

CFGs

- $S \rightarrow NP VP$
 - This says that there are units called S, NP, and VP in this language
 - That an S consists of an NP followed immediately by a VP
 - Doesn't say that that's the only kind of S
 - Nor does it say that this is the only place that NPs and VPs occur

Generativity

- This is a generative model that can also be used for analysis, somewhat like Markov Models
 - Generate strings in the language
 - Reject strings not in the language
 - Impose structures (trees) on strings in the language

Derivations

- A derivation is a sequence of rules applied to a string that accounts for that string
 - Covers all the elements in the string
 - Covers only the elements in the string

Context Free Grammars: Derivations

$S \rightarrow NP S$

Who does Bill think Harry likes?

$S \rightarrow V S$

S

$S \rightarrow NP VP$

$VP \rightarrow V S$

$VP \rightarrow V$

$NP \rightarrow \text{who}$

$NP \rightarrow \text{Bill}$

$NP \rightarrow \text{Harry}$

$V \rightarrow \text{does}$

$V \rightarrow \text{think}$

$V \rightarrow \text{likes}$

Context Free Grammars: Derivations

$S \rightarrow NP S$

Who does Bill think Harry likes?

$S \rightarrow V S$

$S \rightarrow NP VP$

$VP \rightarrow V S$

$VP \rightarrow V$

$NP \rightarrow \text{who}$

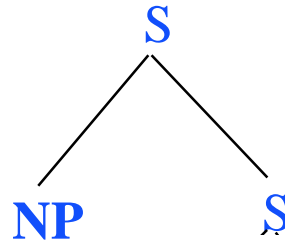
$NP \rightarrow \text{Bill}$

$NP \rightarrow \text{Harry}$

$V \rightarrow \text{does}$

$V \rightarrow \text{think}$

$V \rightarrow \text{likes}$



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

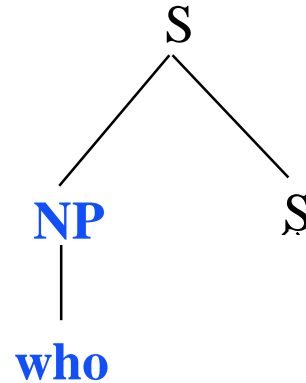
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

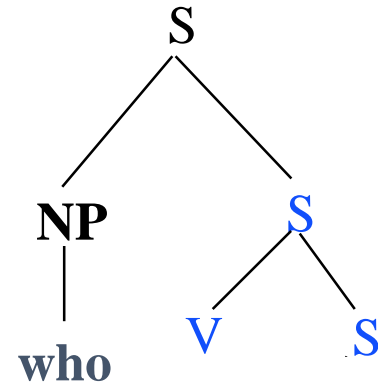
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

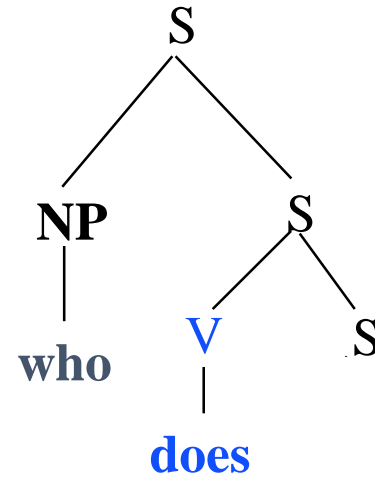
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

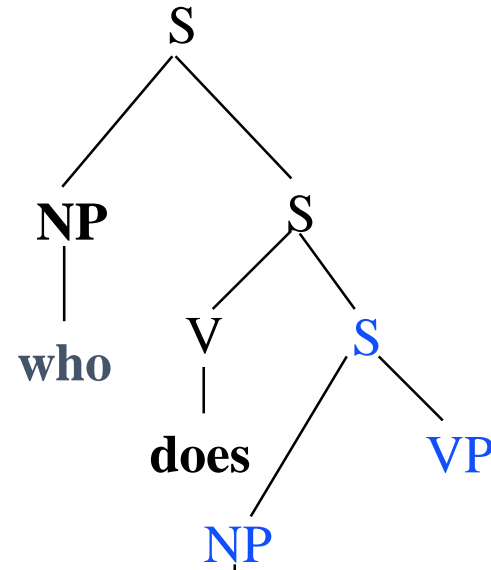
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

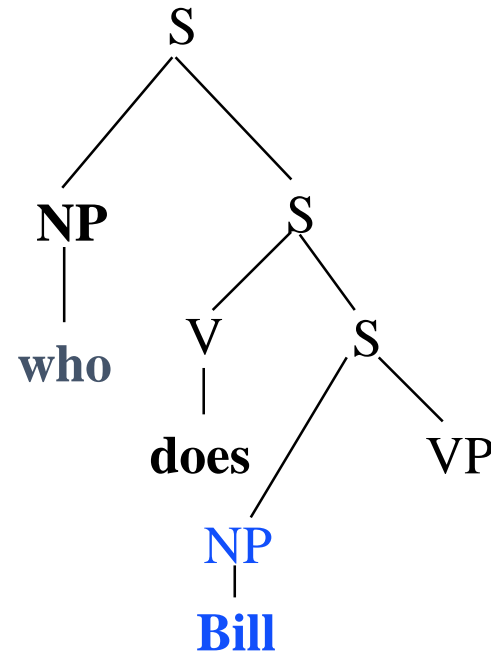
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

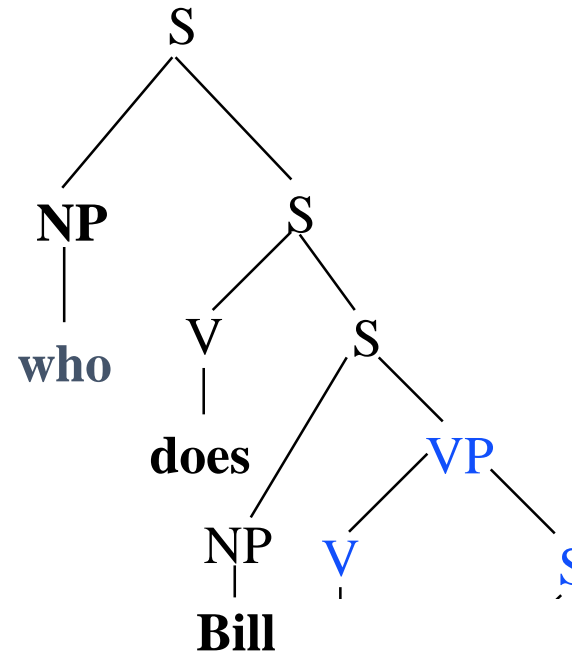
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

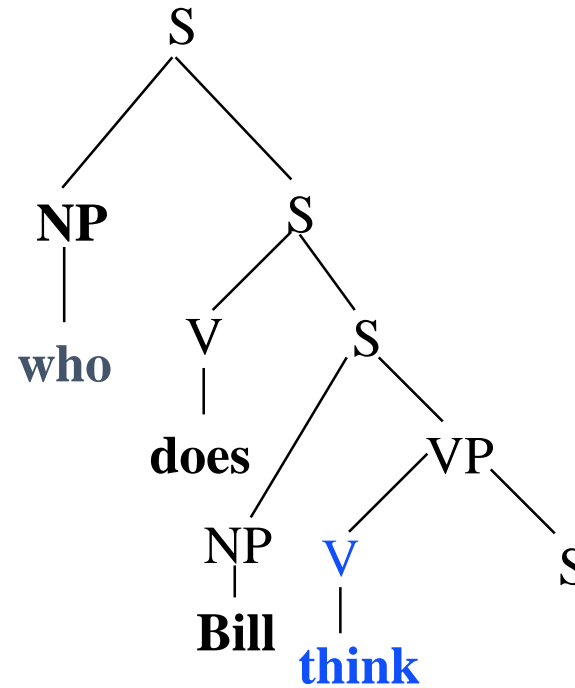
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

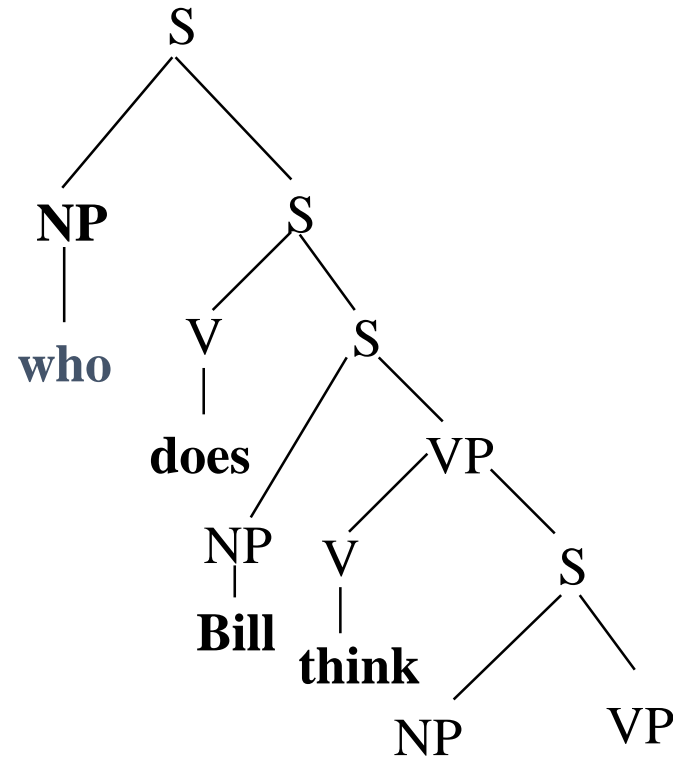
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

$S \rightarrow NP S$

$S \rightarrow V S$

$S \rightarrow NP VP$

$VP \rightarrow V S$

$VP \rightarrow V$

$NP \rightarrow \text{who}$

$NP \rightarrow \text{Bill}$

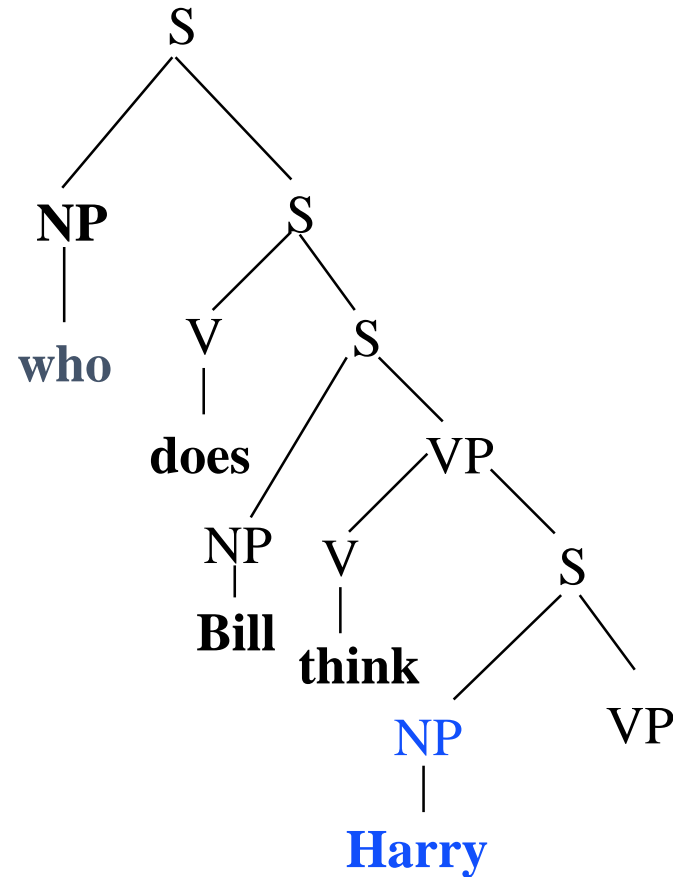
$NP \rightarrow \text{Harry}$

$V \rightarrow \text{does}$

$V \rightarrow \text{think}$

$V \rightarrow \text{likes}$

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

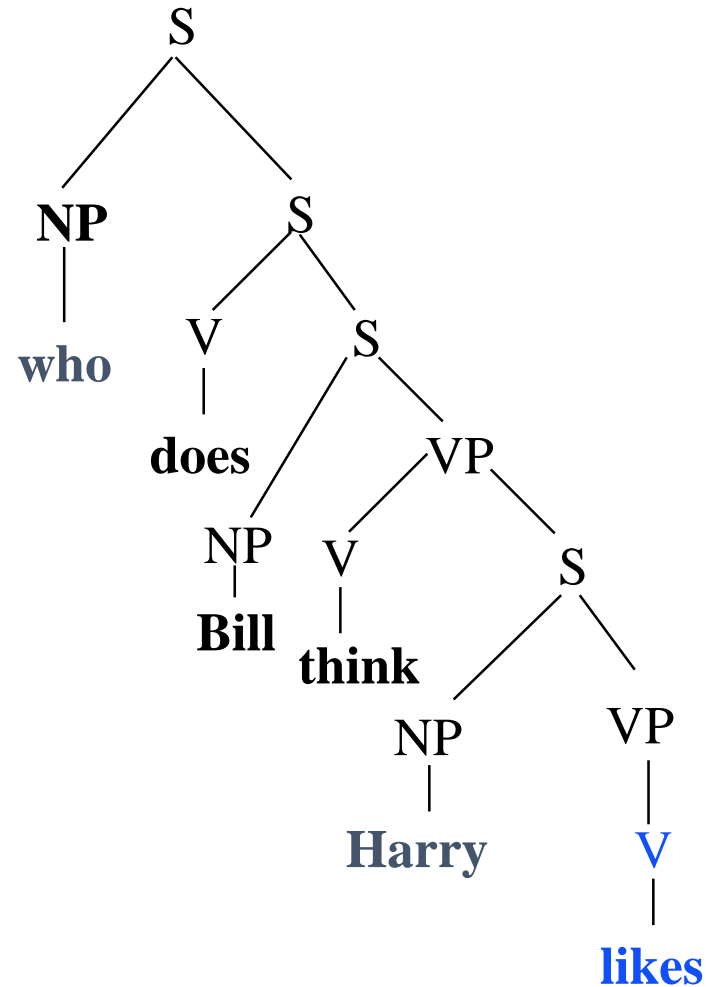
NP → Harry

V → does

V → think

V → likes

Who does Bill think Harry likes?



Context Free Grammars: Derivations

S → NP S

S → V S

S → NP VP

VP → V S

VP → V

NP → who

NP → Bill

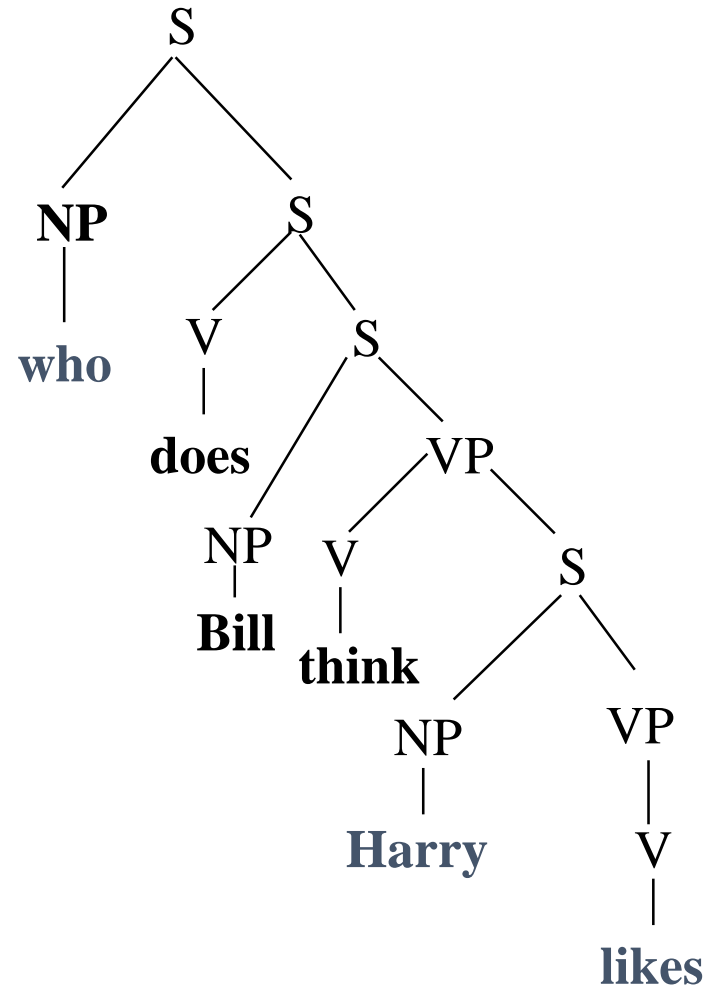
NP → Harry

V → does

V → think

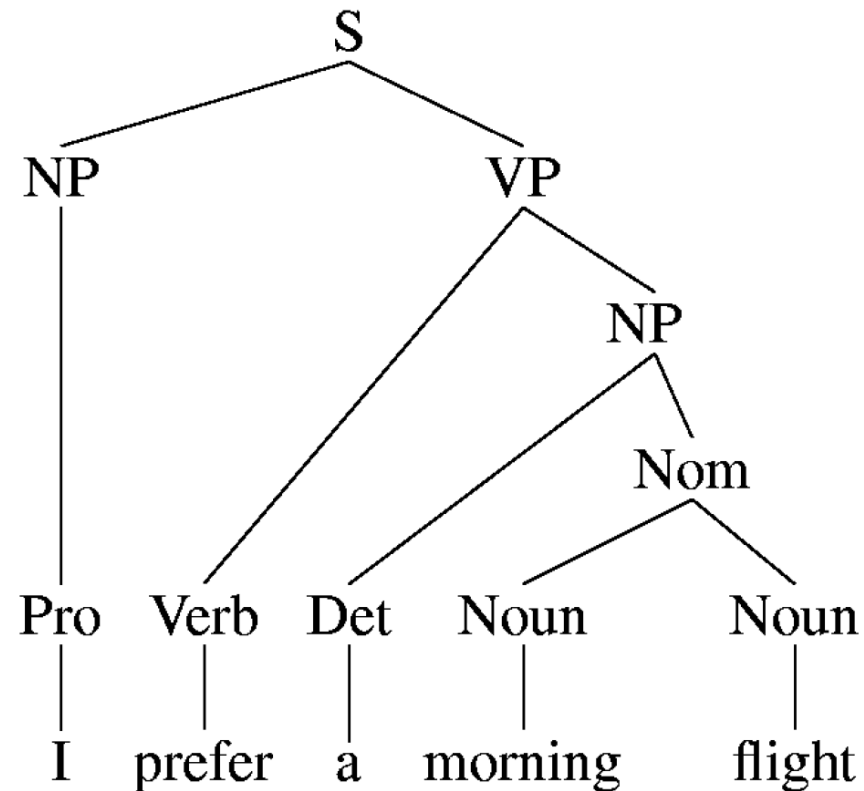
V → likes

Who does Bill think Harry likes?



Bracketed Notation

- $[_S [_{NP} [_{PRO} I] [_{VP} [_V \text{prefer} [_{NP} [_{NP} [_{Det} a] [_{Nom} [_N \text{morning}] [_N \text{flight}]]]]]]]$



Context free?

- The notion of context in CFGs has nothing to do with the ordinary meaning of the word context in language.
- All it really means is that the non-terminal on the left-hand side of a rule is out there all by itself (free of context)

$A \rightarrow BC$

This means that I can rewrite an A as a B followed by a C regardless of the context in which A is found

Chomsky Hierarchy

- A terse overview in around 10 minutes (you can use the Wikipedia article as a companion):
<https://www.youtube.com/watch?v=xlw3KFAQKj8>
- Of historical importance: the original paper by Chomsky:
<https://chomsky.info/wp-content/uploads/195609-.pdf>