Context Free Grammars

Many slides from Michael Collins

Overview

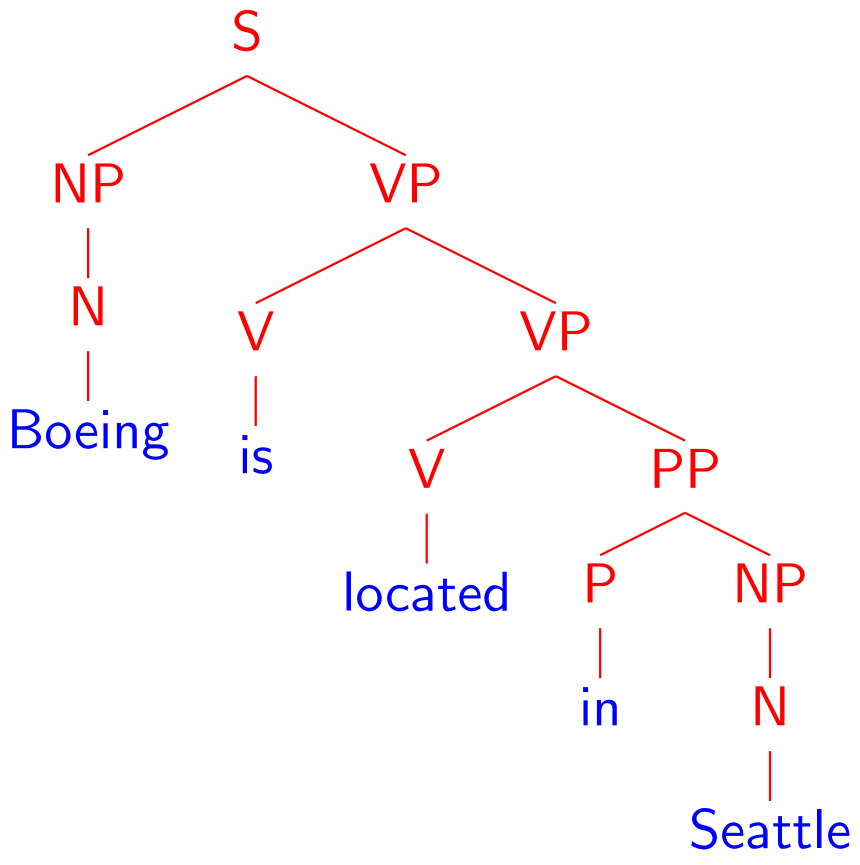
- An introduction to the parsing problem
- Context free grammars
- A brief(!) sketch of the syntax of English
- Examples of ambiguous structures

Parsing (Syntactic Structure)

INPUT:

Boeing is located in Seattle.

OUTPUT:



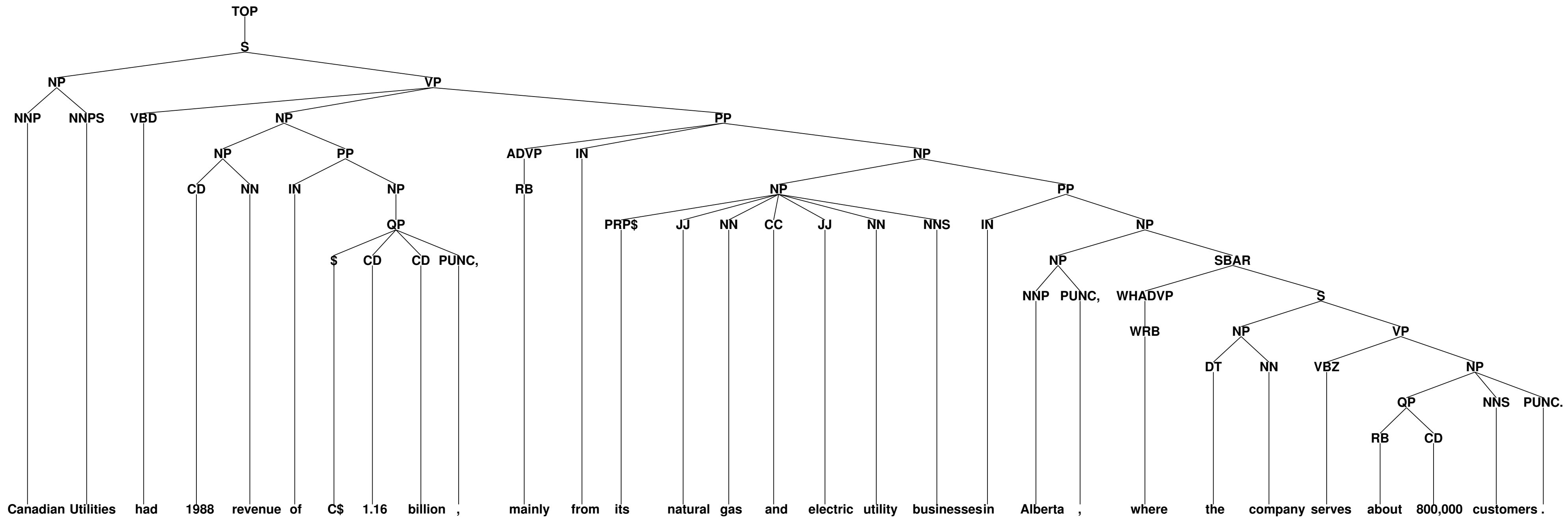
Syntactic Formalisms

- Work in formal syntax goes back to Chomsky's PhD thesis in the 1950s
- Examples of current formalisms: minimalism, lexical functional grammar (LFG), head-driven phrase-structure grammar (HPSG), tree adjoining grammars (TAG), categorial grammars

Data for Parsing Experiments

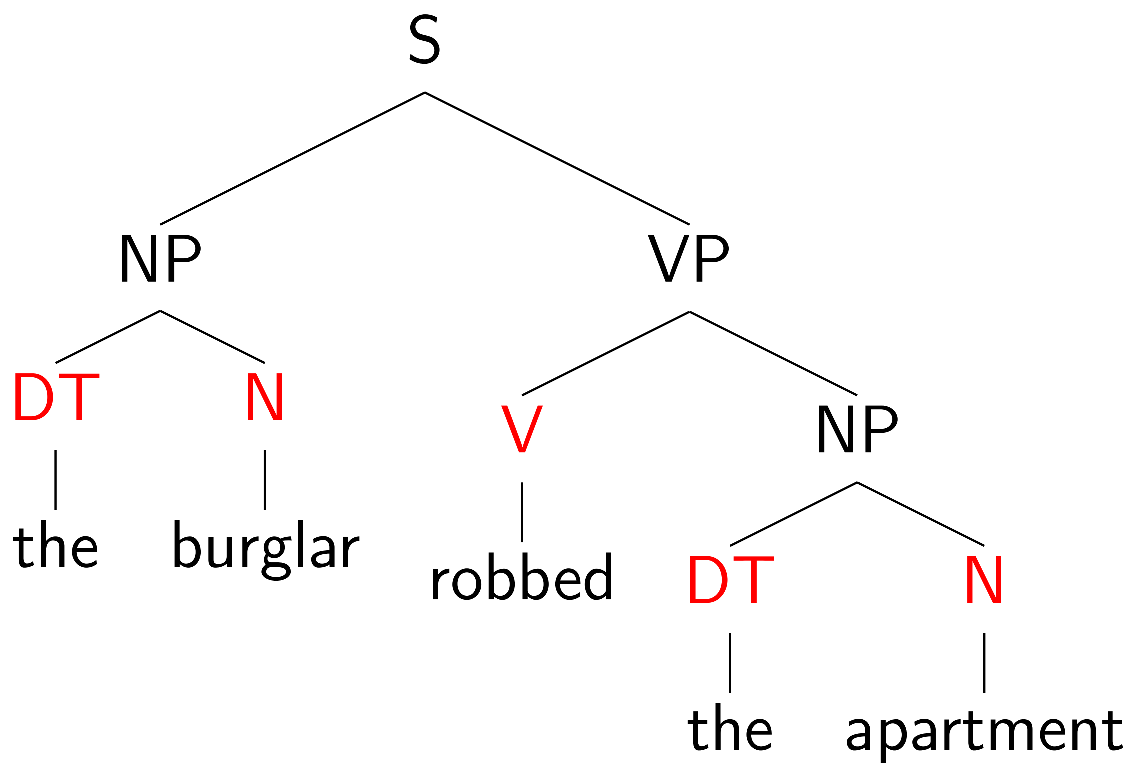
- Penn WSJ Treebank = 50,000 sentences with associated trees
- Usual set-up: 40,000 training sentences, 2400 test sentences

An example tree:



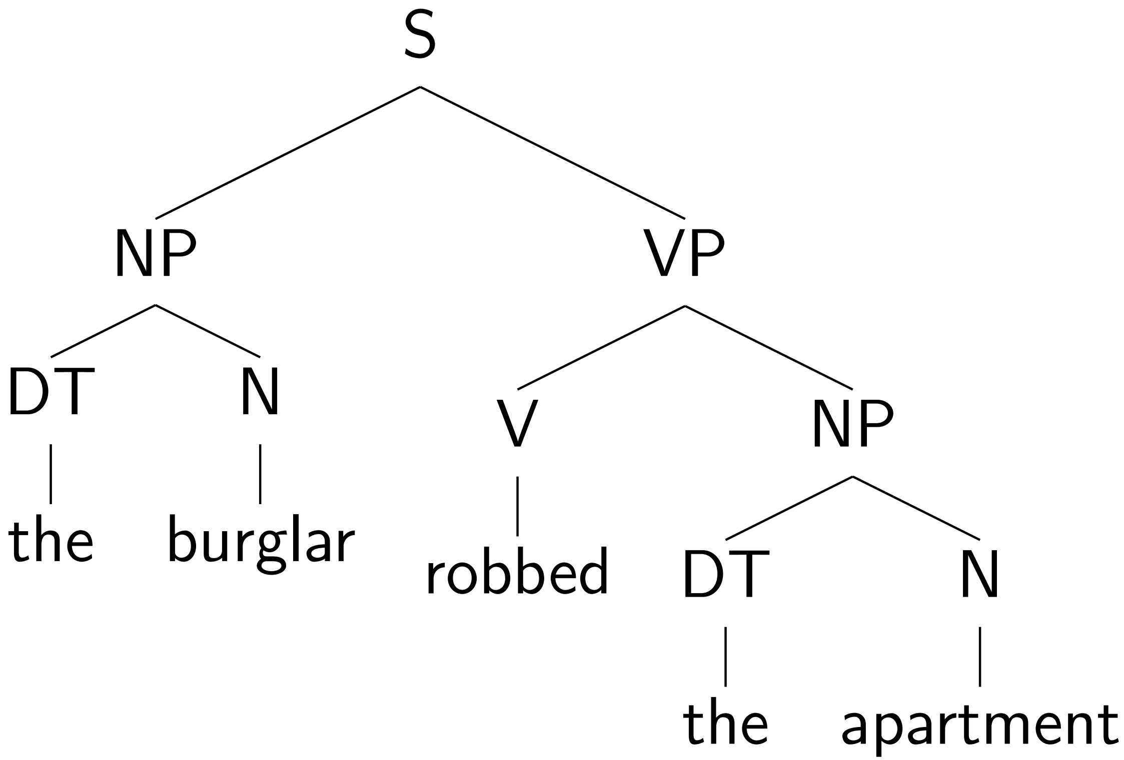
The Information Conveyed by Parse Trees

(1) Part of speech for each word
(N = noun, V = verb, DT = determiner)



The Information Conveyed by Parse Trees (continued)

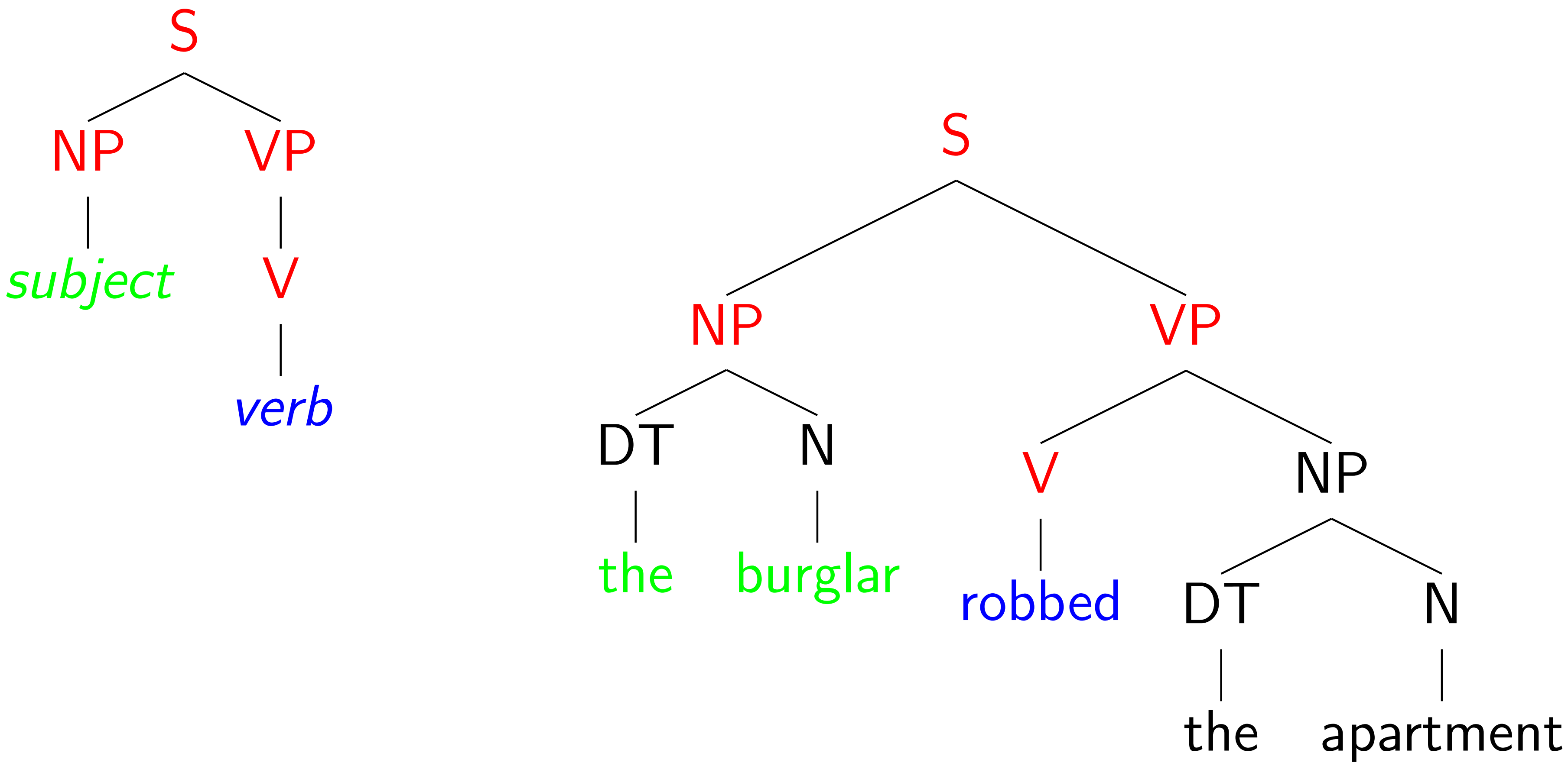
(2) Phrases



Noun Phrases (NP): "the burglar", "the apartment"
Verb Phrases (VP): "robbed the apartment"
Sentences (S): "the burglar robbed the apartment"

The Information Conveyed by Parse Trees (continued)

(3) Useful Relationships



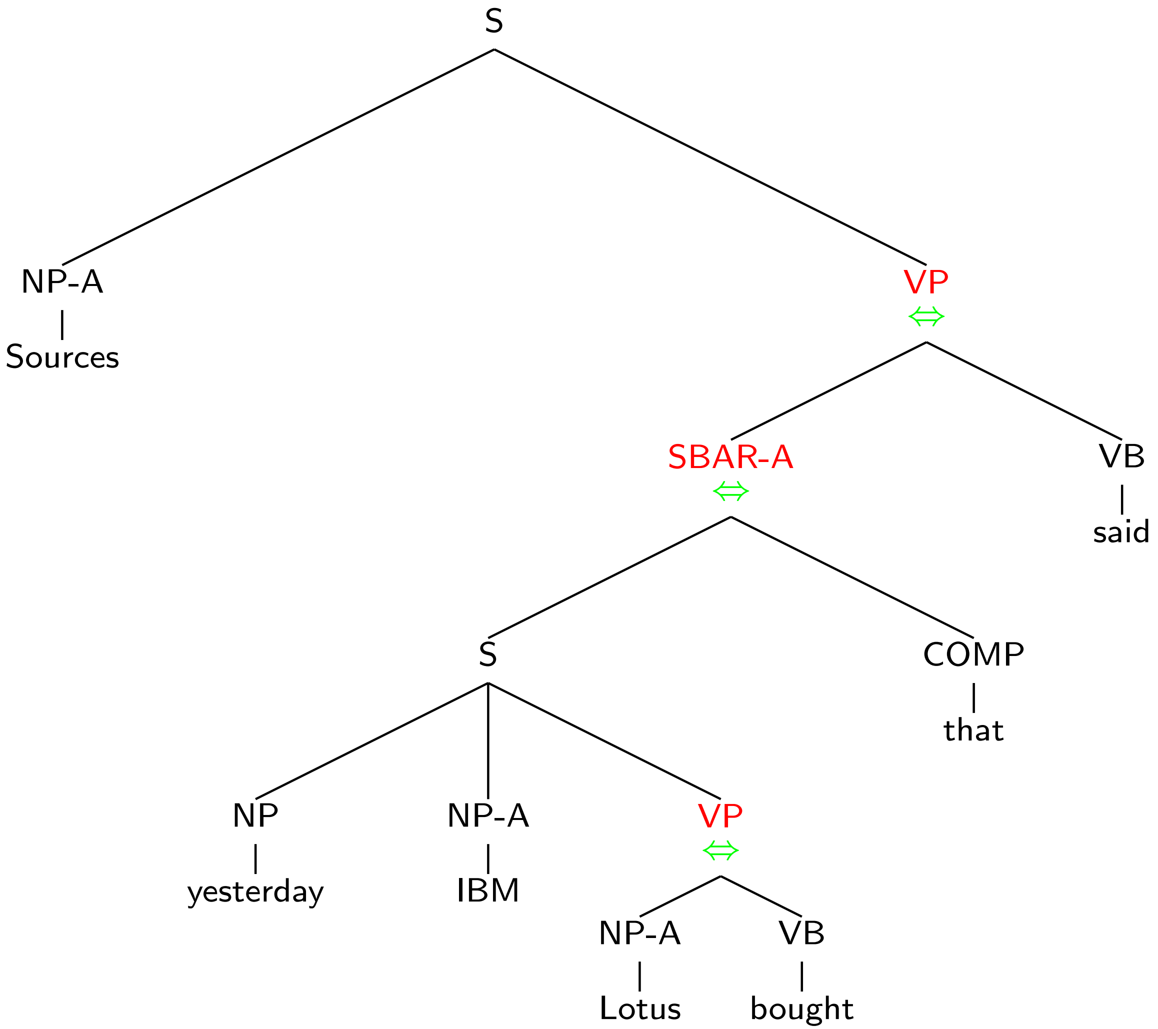
$\Rightarrow$ "the burglar" is the subject of "robbed"

An Example Application: Machine Translation

- English word order is *subject – verb – object*
- Japanese word order is *subject – object – verb*

English: IBM bought Lotus
Japanese: *IBM Lotus bought*

English: Sources said that IBM bought Lotus yesterday
Japanese: *Sources yesterday IBM Lotus bought that said*


S
NP-A
Sources
VP
⇔
SBAR-A
⇔
VB
said
S
COMP
that
NP
yesterday
NP-A
IBM
VP
⇔
NP-A
Lotus
VB
bought

Overview

- An introduction to the parsing problem
- Context free grammars
- A brief(!) sketch of the syntax of English
- Examples of ambiguous structures

Context-Free Grammars

Hopcroft and Ullman, 1979

A context free grammar $G = (N, \Sigma, R, S)$ where:

- N is a set of non-terminal symbols
- Σ is a set of terminal symbols
- R is a set of rules of the form $X \rightarrow Y_1 Y_2 \ldots Y_n$ for $n \geq 0$, $X \in N$, $Y_i \in (N \cup \Sigma)$
- $S \in N$ is a distinguished start symbol

A Context-Free Grammar for English

N = {S, NP, VP, PP, DT, Vi, Vt, NN, IN}
S = S
Σ = {sleeps, saw, man, woman, telescope, the, with, in}

$R =$

S	$\rightarrow$	NP	VP
VP	$\rightarrow$	Vi	
VP	$\rightarrow$	Vt	NP
VP	$\rightarrow$	VP	PP
NP	$\rightarrow$	DT	NN
NP	$\rightarrow$	NP	PP
PP	$\rightarrow$	IN	NP

Vi	$\rightarrow$	sleeps
Vt	$\rightarrow$	saw
NN	$\rightarrow$	man
NN	$\rightarrow$	woman
NN	$\rightarrow$	telescope
DT	$\rightarrow$	the
IN	$\rightarrow$	with
IN	$\rightarrow$	in

Note: S=sentence, VP=verb phrase, NP=noun phrase, PP=prepositional phrase, DT=determiner, Vi=intransitive verb, Vt=transitive verb, NN=noun, IN=preposition

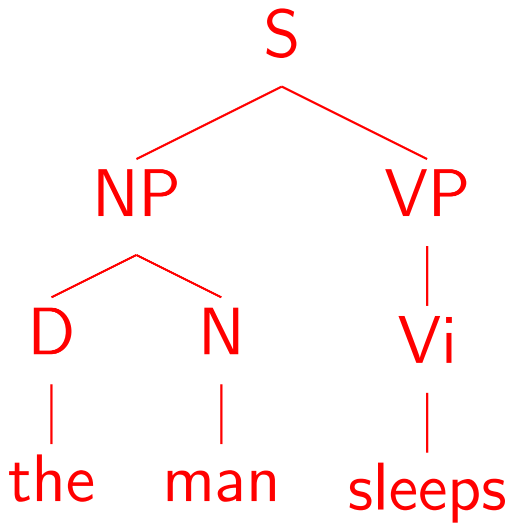
Left-Most Derivations

A left-most derivation is a sequence of strings $s_1 \ldots s_n$, where

- $s_1 = S$, the start symbol
- $s_n \in \Sigma^*$, i.e. s_n is made up of terminal symbols only
- Each s_i for $i = 2 \ldots n$ is derived from s_{i-1} by picking the left-most non-terminal X in s_{i-1} and replacing it by some β where $X \rightarrow \beta$ is a rule in R

For example: [S], [NP VP], [D N VP], [the N VP], [the man VP], [the man Vi], [the man sleeps]

Representation of a derivation as a tree:



An Example

DERIVATION	RULES USED
S	

An Example

DERIVATION	RULES USED
S	
	S $\rightarrow$ NP VP
NP VP	

An Example

DERIVATION	RULES USED
S	
	S $\rightarrow$ NP VP
NP VP	
	NP $\rightarrow$ DT N
DT N VP	

An Example

DERIVATION	RULES USED
S	S $\rightarrow$ NP VP
NP VP	NP $\rightarrow$ DT N
DT N VP	DT $\rightarrow$ the
the N VP	

An Example

DERIVATION	RULES USED
S	
	S $\rightarrow$ NP VP
NP VP	
	NP $\rightarrow$ DT N
DT N VP	
	DT $\rightarrow$ the
the N VP	
	N $\rightarrow$ dog
the dog VP	

An Example

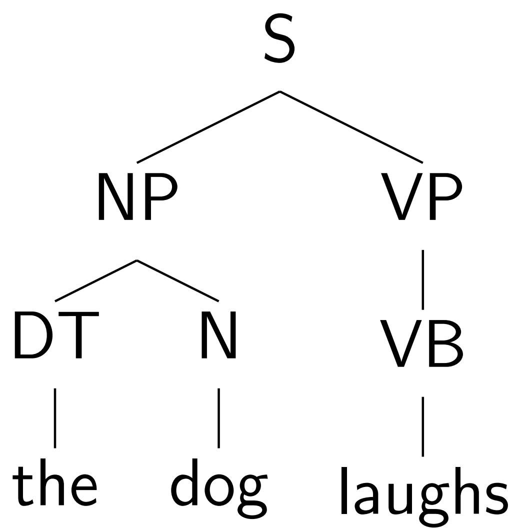
DERIVATION	RULES USED
S	S $\rightarrow$ NP VP
NP VP	NP $\rightarrow$ DT N
DT N VP	DT $\rightarrow$ the
the N VP	N $\rightarrow$ dog
the dog VP	VP $\rightarrow$ VB
the dog VB	

An Example

DERIVATION	RULES USED
S	
NP VP	S $\rightarrow$ NP VP
DT N VP	NP $\rightarrow$ DT N
the N VP	DT $\rightarrow$ the
the dog VP	N $\rightarrow$ dog
the dog VB	VP $\rightarrow$ VB
the dog laughs	VB $\rightarrow$ laughs

An Example

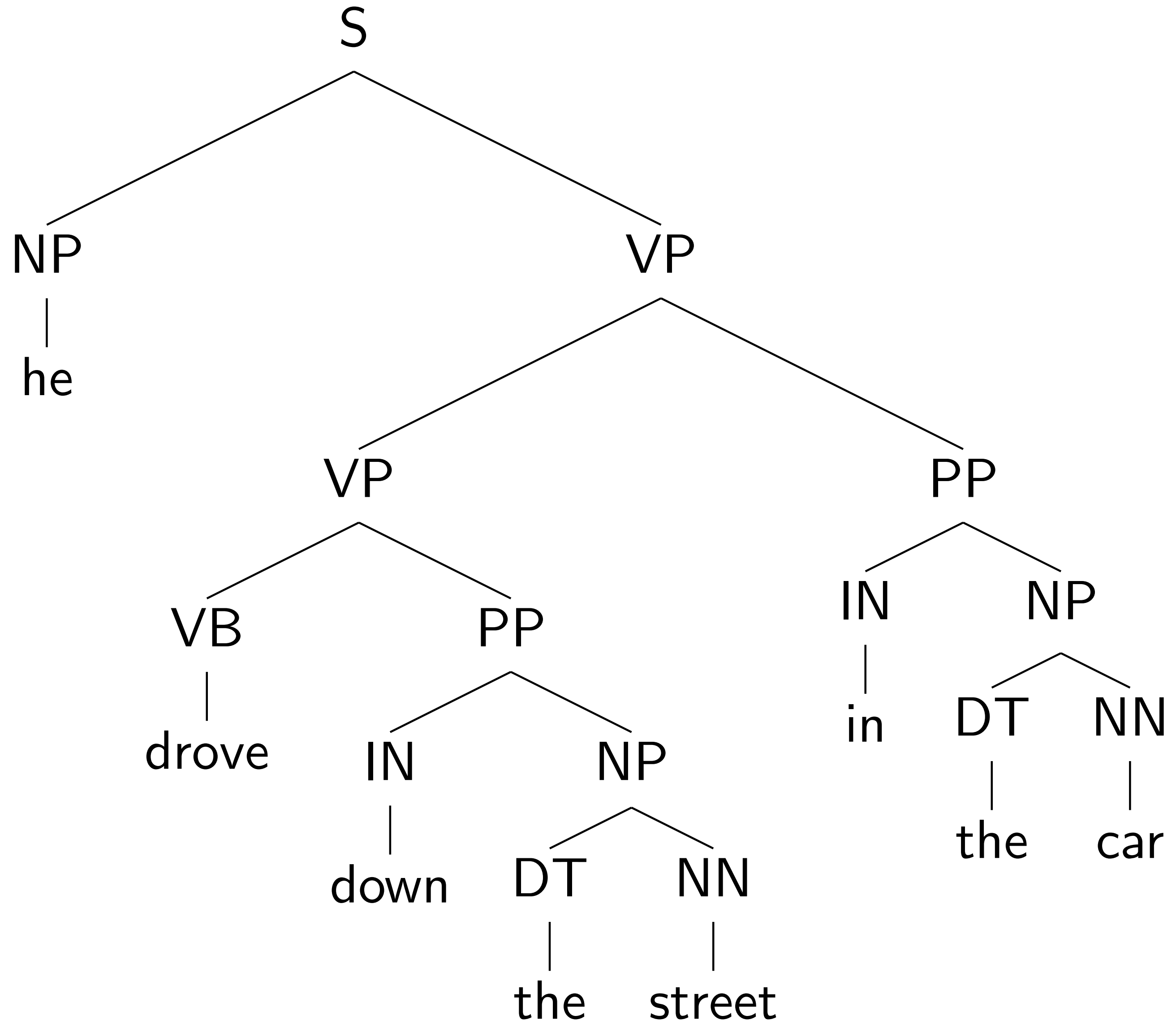
DERIVATION	RULES USED
S	
NP VP	$S \rightarrow NP\ VP$
DT N VP	$NP \rightarrow DT\ N$
the N VP	$DT \rightarrow$ the
the dog VP	$N \rightarrow$ dog
the dog VB	$VP \rightarrow VB$
the dog laughs	$VB \rightarrow$ laughs



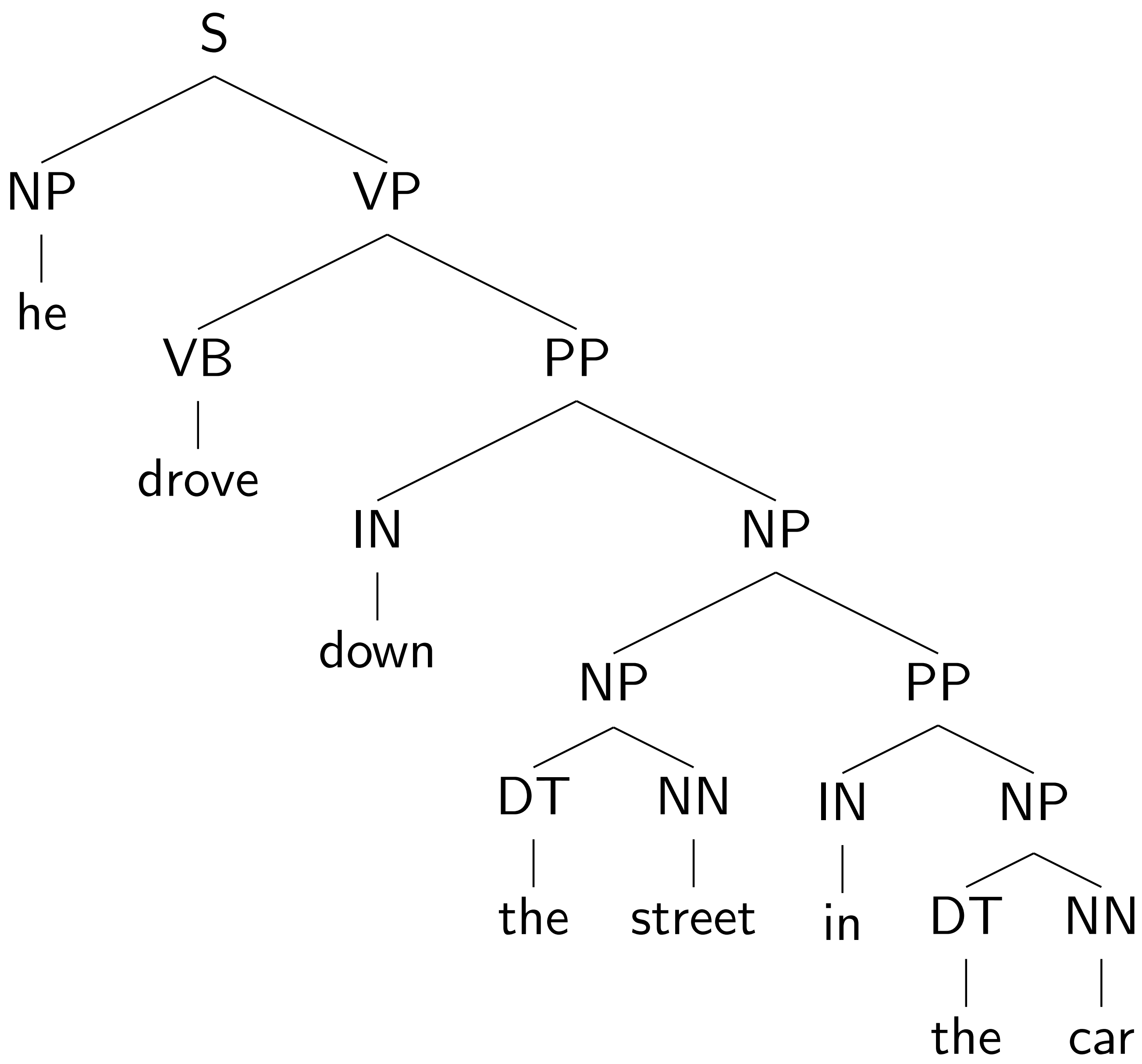
Properties of CFGs

- A CFG defines a set of possible derivations
- A string $s \in \Sigma^*$ is in the *language* defined by the CFG if there is at least one derivation that yields s
- Each string in the language generated by the CFG may have more than one derivation ("ambiguity")

An Example of Ambiguity



An Example of Ambiguity (continued)



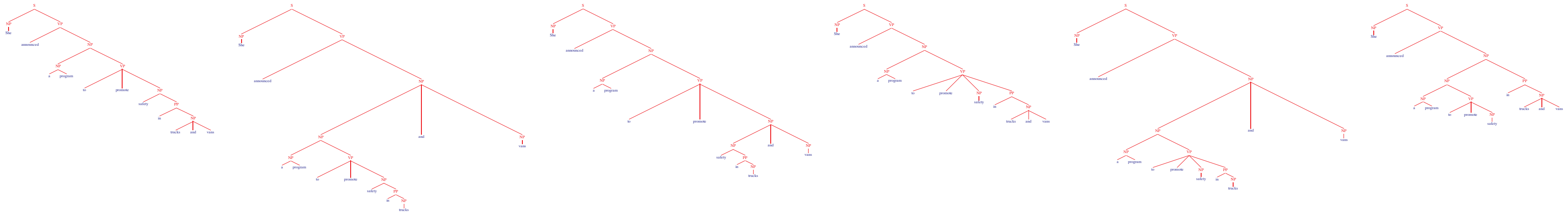
The Problem with Parsing: Ambiguity

INPUT:

She announced a program to promote safety in trucks and vans

⇓

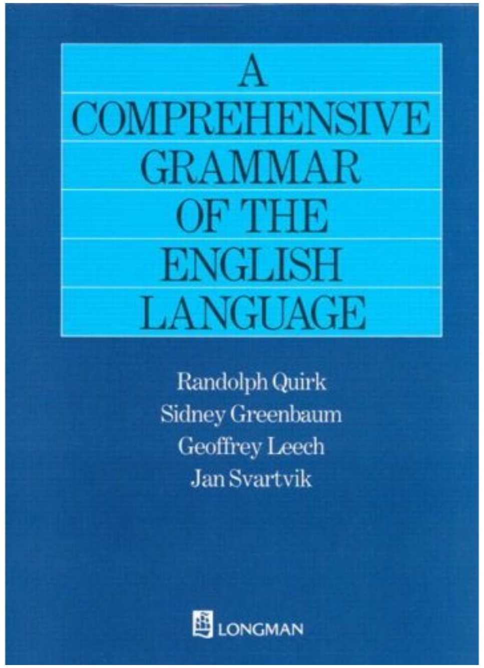
POSSIBLE OUTPUTS:



And there are more...

Overview

- An introduction to the parsing problem
- Context free grammars
- A brief(!) sketch of the syntax of English
- Examples of ambiguous structures



Product Details (from Amazon)

Hardcover: 1779 pages

Publisher: Longman; 2nd Revised edition

Language: English

ISBN-10: 0582517346

ISBN-13: 978-0582517349

Product Dimensions: 8.4 x 2.4 x 10 inches

Shipping Weight: 4.6 pounds

A Brief Overview of English Syntax

Parts of Speech (tags from the Brown corpus):

- Nouns
  NN = singular noun e.g., man, dog, park
  NNS = plural noun e.g., telescopes, houses, buildings
  NNP = proper noun e.g., Smith, Gates, IBM
- Determiners
  DT = determiner e.g., the, a, some, every
- Adjectives
  JJ = adjective e.g., red, green, large, idealistic

A Fragment of a Noun Phrase Grammar

$\bar{N}$	$\Rightarrow$	NN	
$\bar{N}$	$\Rightarrow$	NN	$\bar{N}$
$\bar{N}$	$\Rightarrow$	JJ	$\bar{N}$
$\bar{N}$	$\Rightarrow$	$\bar{N}$	$\bar{N}$
NP	$\Rightarrow$	DT	$\bar{N}$

NN	$\Rightarrow$	box
NN	$\Rightarrow$	car
NN	$\Rightarrow$	mechanic
NN	$\Rightarrow$	pigeon
DT	$\Rightarrow$	the
DT	$\Rightarrow$	a

JJ	$\Rightarrow$	fast
JJ	$\Rightarrow$	metal
JJ	$\Rightarrow$	idealistic
JJ	$\Rightarrow$	clay

Prepositions, and Prepositional Phrases

- Prepositions
  IN = preposition e.g., of, in, out, beside, as

An Extended Grammar

$$
\begin{array}{|lcll|}
\bar{N} & \Rightarrow & NN & \\
\bar{N} & \Rightarrow & NN & \bar{N} \\
\bar{N} & \Rightarrow & JJ & \bar{N} \\
\bar{N} & \Rightarrow & \bar{N} & \bar{N} \\
NP & \Rightarrow & DT & \bar{N} \\
 & & & \\
PP & \Rightarrow & IN & NP \\
\bar{N} & \Rightarrow & \bar{N} & PP
\end{array}
\quad
\begin{array}{|lcl|}
NN & \Rightarrow & \text{box} \\
NN & \Rightarrow & \text{car} \\
NN & \Rightarrow & \text{mechanic} \\
NN & \Rightarrow & \text{pigeon} \\
 & & \\
DT & \Rightarrow & \text{the} \\
DT & \Rightarrow & \text{a}
\end{array}
\quad
\begin{array}{|lcl|}
JJ & \Rightarrow & \text{fast} \\
JJ & \Rightarrow & \text{metal} \\
JJ & \Rightarrow & \text{idealistic} \\
JJ & \Rightarrow & \text{clay} \\
 & & \\
IN & \Rightarrow & \text{in} \\
IN & \Rightarrow & \text{under} \\
IN & \Rightarrow & \text{of} \\
IN & \Rightarrow & \text{on} \\
IN & \Rightarrow & \text{with} \\
IN & \Rightarrow & \text{as}
\end{array}
$$

Generates:

in a box, under the box, the fast car mechanic under the pigeon in the box, . . .

Verbs, Verb Phrases, and Sentences

- Basic Verb Types

Vi = Intransitive verb		e.g., sleeps, walks, laughs
Vt = Transitive verb		e.g., sees, saw, likes
Vd = Ditransitive verb		e.g., gave

- Basic VP Rules

$$\begin{array}{llll} \text{VP} & \rightarrow & \text{Vi} & \\ \text{VP} & \rightarrow & \text{Vt} & \text{NP} \\ \text{VP} & \rightarrow & \text{Vd} & \text{NP} \quad \text{NP} \end{array}$$

- Basic S Rule

$$\text{S} \rightarrow \text{NP} \quad \text{VP}$$

Examples of VP:

sleeps, walks, likes the mechanic, gave the mechanic the fast car

Examples of S:

the man sleeps, the dog walks, the dog gave the mechanic the fast car

PPs Modifying Verb Phrases

A new rule: VP $\rightarrow$ VP PP

New examples of VP:
sleeps in the car, walks like the mechanic, gave the mechanic the fast car on Tuesday, . . .

Complementizers, and SBARs

- Complementizers
  COMP = complementizer e.g., that
- SBAR Subordinate clause
  SBAR $\rightarrow$ COMP S

Examples:

that the man sleeps, that the mechanic saw the dog ...

More Verbs

- New Verb Types

V[5]	e.g., said, reported
V[6]	e.g., told, informed
V[7]	e.g., bet

- New VP Rules

$$\begin{array}{lllll} VP & \rightarrow & V[5] & SBAR & \\ VP & \rightarrow & V[6] & NP & SBAR \\ VP & \rightarrow & V[7] & NP & NP \quad SBAR \end{array}$$

Examples of New VPs:

said that the man sleeps

told the dog that the mechanic likes the pigeon

bet the pigeon $50 that the mechanic owns a fast car

Coordination

- A New Part-of-Speech:
  CC = Coordinator e.g., and, or, but
- New Rules

NP	$\rightarrow$	NP	CC	NP
$\bar{\text{N}}$	$\rightarrow$	$\bar{\text{N}}$	CC	$\bar{\text{N}}$
VP	$\rightarrow$	VP	CC	VP
S	$\rightarrow$	S	CC	S
SBAR	$\rightarrow$	SBAR	CC	SBAR

We've Only Scratched the Surface...

- Agreement

  The dogs laugh *vs.* The dog laughs

- Wh-movement Long-distance dependency

  The dog that the cat liked ___

- Active vs. passive

  The dog saw the cat *vs.*
  The cat was seen by the dog

- If you're interested in reading more:

  Syntactic Theory: A Formal Introduction, 2nd Edition. Ivan A. Sag, Thomas Wasow, and Emily M. Bender.

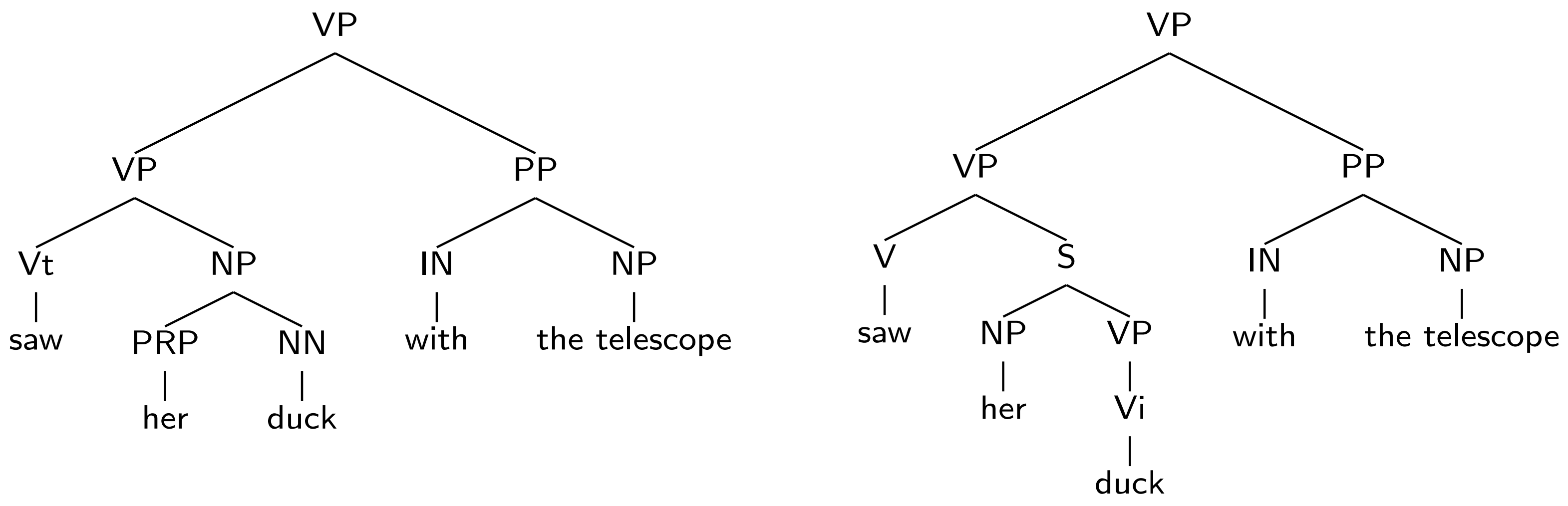
Overview

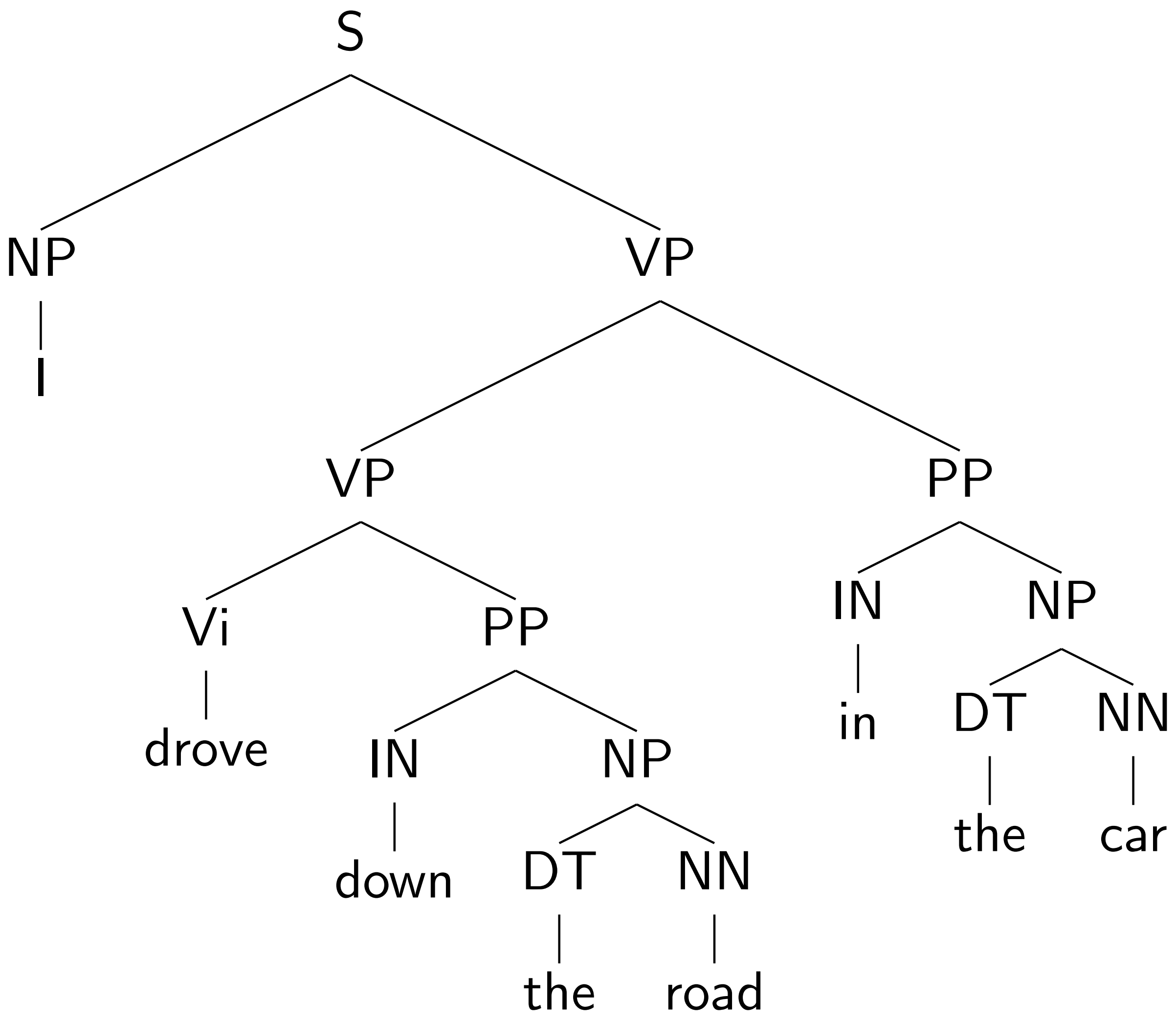
- An introduction to the parsing problem
- Context free grammars
- A brief(!) sketch of the syntax of English
- Examples of ambiguous structures

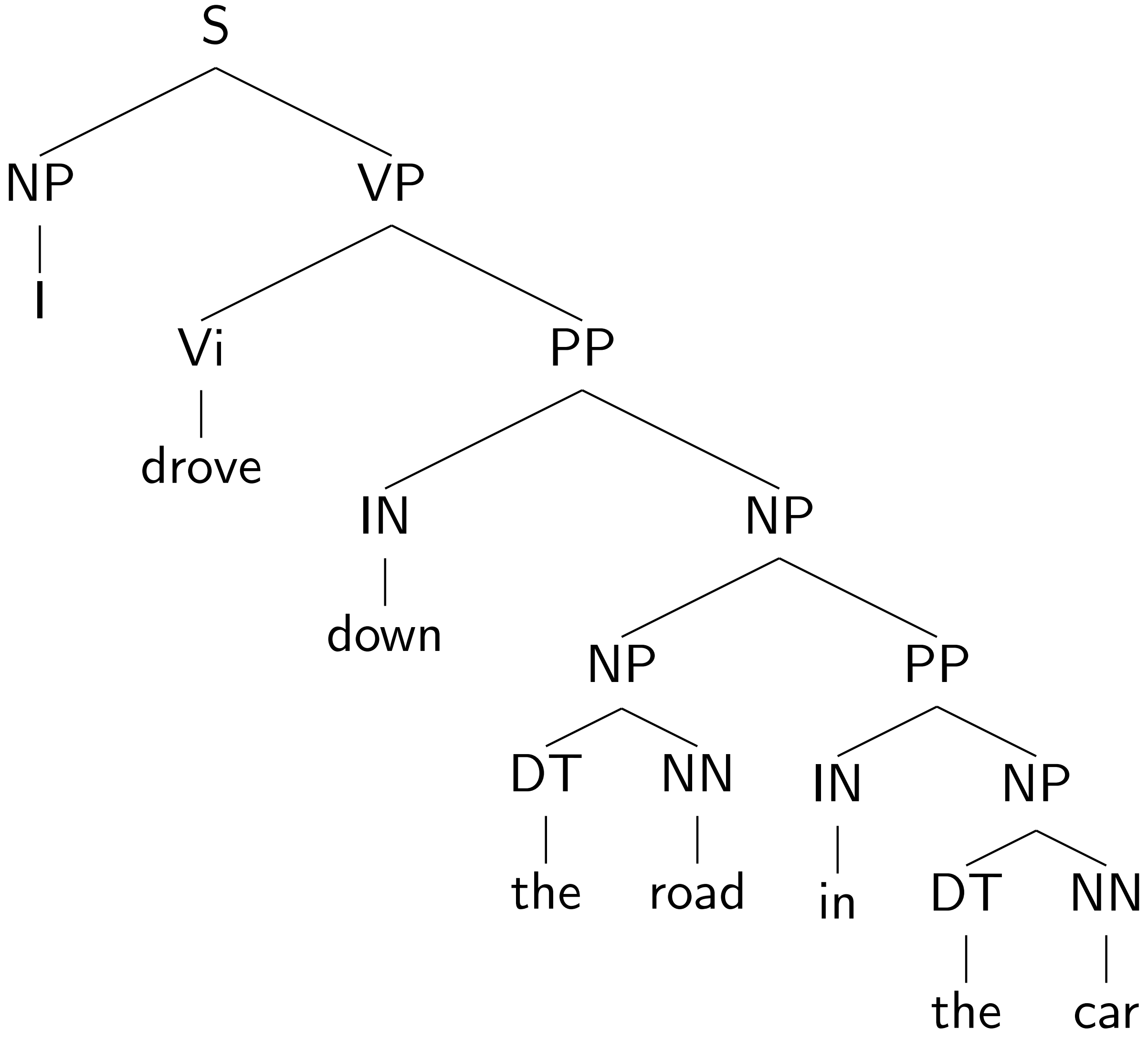
Sources of Ambiguity

- Part-of-Speech ambiguity

$$NN \rightarrow \text{duck}$$
$$Vi \rightarrow \text{duck}$$




S
NP
I
VP
VP
Vi
drove
PP
IN
down
NP
DT
the
NN
road
PP
IN
in
NP
DT
the
NN
car


S
NP
VP
I
Vi
PP
drove
IN
NP
down
NP
PP
DT
NN
IN
NP
the
road
in
DT
NN
the
car

Two analyses for: John was believed to have been shot by Bill

With the same set of grammar rules

Sources of Ambiguity: Noun Premodifiers

- Noun premodifiers: