

# Parsing, and Context-Free Grammars

Michael Collins, Columbia University

# Overview

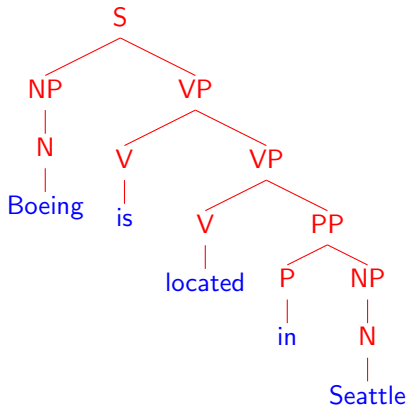
- ▶ An introduction to the parsing problem
- ▶ Context free grammars
- ▶ A brief(!) sketch of the syntax of English
- ▶ Examples of ambiguous structures

# Parsing (Syntactic Structure)

INPUT:

Boeing is located in Seattle.

OUTPUT:



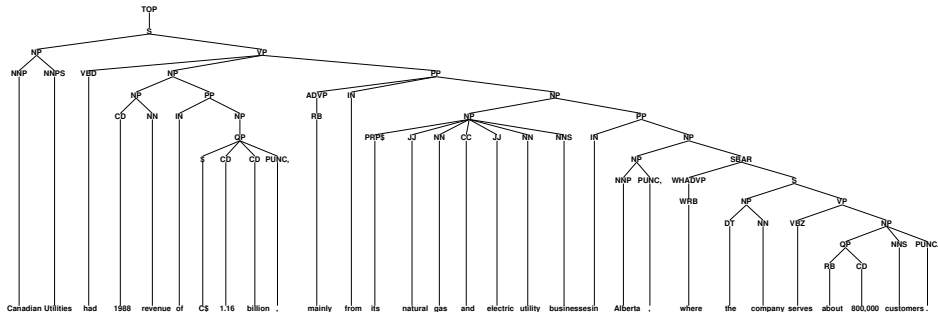
# Syntactic Formalisms

- ▶ 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

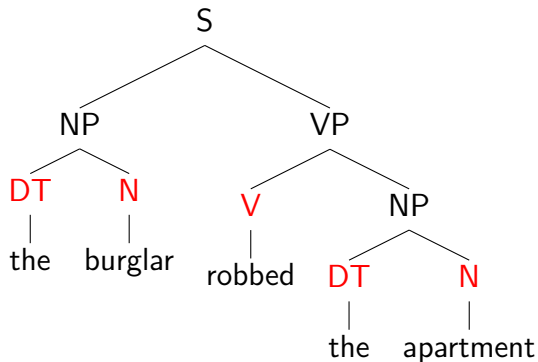
## An example tree:



# The Information Conveyed by Parse Trees

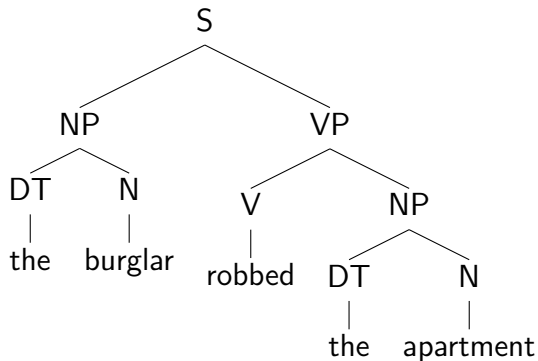
(1) Part of speech for each word

(N = noun, V = verb, DT = determiner)



# The Information Conveyed by Parse Trees (continued)

## (2) Phrases



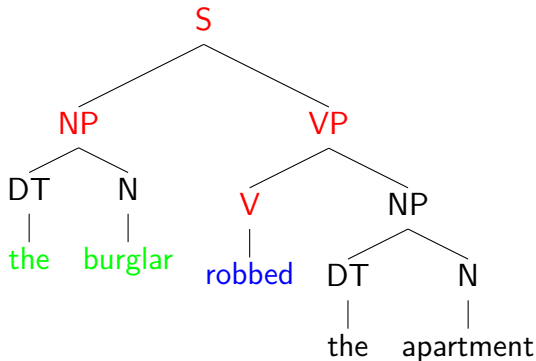
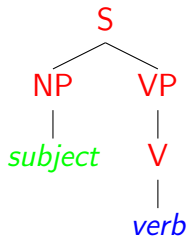
Noun Phrases (NP): “the burglar”, “the apartment”

Verb Phrases (VP): “robbed the apartment”

Sentences (S): “the burglar robbed the apartment”

# The Information Conveyed by Parse Trees (continued)

## (3) Useful Relationships



⇒ “the burglar” is the subject of “robbed”



# An Example Application: Machine Translation

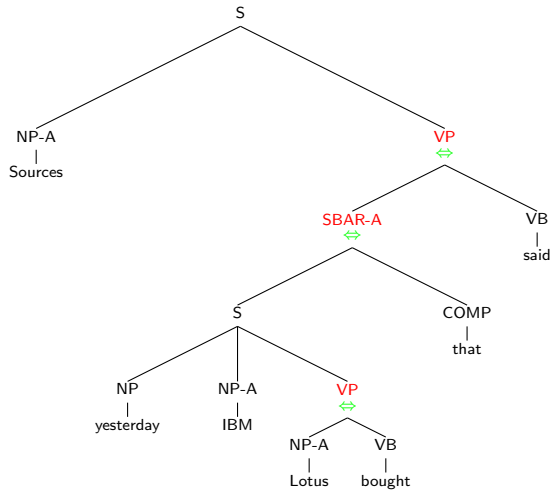
- ▶ English word order is *subject – verb – object*
- ▶ Japanese word order is *subject – object – verb*

English: IBM bought Lotus

Japanese: *IBM Lotus bought*

English: Sources said that IBM bought Lotus yesterday

Japanese: *Sources yesterday IBM Lotus bought that said*



# 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
- ▶  $\Sigma$  is a set of terminal symbols
- ▶  $R$  is a set of rules of the form  $X \rightarrow Y_1 Y_2 \dots Y_n$   
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

$N = \{S, NP, VP, PP, DT, Vi, Vt, NN, IN\}$

$S = S$

$\Sigma = \{\text{sleeps, saw, man, woman, telescope, the, with, in}\}$

$R =$

S	→	NP	VP
VP	→	Vi	
VP	→	Vt	NP
VP	→	VP	PP
NP	→	DT	NN
NP	→	NP	PP
PP	→	IN	NP

Vi	→	sleeps
Vt	→	saw
NN	→	man
NN	→	woman
NN	→	telescope
DT	→	the
IN	→	with
IN	→	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

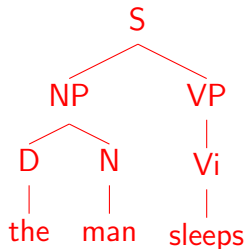
## Left-Most Derivations 最左派生

A left-most derivation is a sequence of strings  $s_1 \dots s_n$ , 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  $i = 2 \dots n$  is derived from  $s_{i-1}$  by picking the left-most non-terminal  $X$  in  $s_{i-1}$  and replacing it by some  $\beta$  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

S

RULES USED

# An Example

## DERIVATION

S

NP VP

## RULES USED

$S \rightarrow NP VP$



# An Example

## DERIVATION

S

NP VP

DT N VP

## RULES USED

$S \rightarrow NP VP$

$NP \rightarrow DT N$

# An Example

## DERIVATION

S

NP VP

DT N VP

the N VP

## RULES USED

$S \rightarrow NP VP$

$NP \rightarrow DT N$

$DT \rightarrow \text{the}$

# An Example

## DERIVATION

S

NP VP

DT N VP

the N VP

the dog VP

## RULES USED

$S \rightarrow NP VP$

$NP \rightarrow DT N$

$DT \rightarrow \text{the}$

$N \rightarrow \text{dog}$

# An Example

## DERIVATION

S

NP VP

DT N VP

the N VP

the dog VP

the dog VB

## RULES USED

$S \rightarrow NP VP$

$NP \rightarrow DT N$

$DT \rightarrow \text{the}$

$N \rightarrow \text{dog}$

$VP \rightarrow VB$

# An Example

## DERIVATION

S

NP VP

DT N VP

the N VP

the dog VP

the dog VB

the dog laughs

## RULES USED

$S \rightarrow NP VP$

$NP \rightarrow DT N$

$DT \rightarrow \text{the}$

$N \rightarrow \text{dog}$

$VP \rightarrow VB$

$VB \rightarrow \text{laughs}$

# An Example

## DERIVATION

S

NP VP

DT N VP

the N VP

the dog VP

the dog VB

the dog laughs

## RULES USED

$S \rightarrow NP VP$

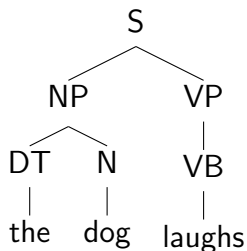
$NP \rightarrow DT N$

$DT \rightarrow \text{the}$

$N \rightarrow \text{dog}$

$VP \rightarrow VB$

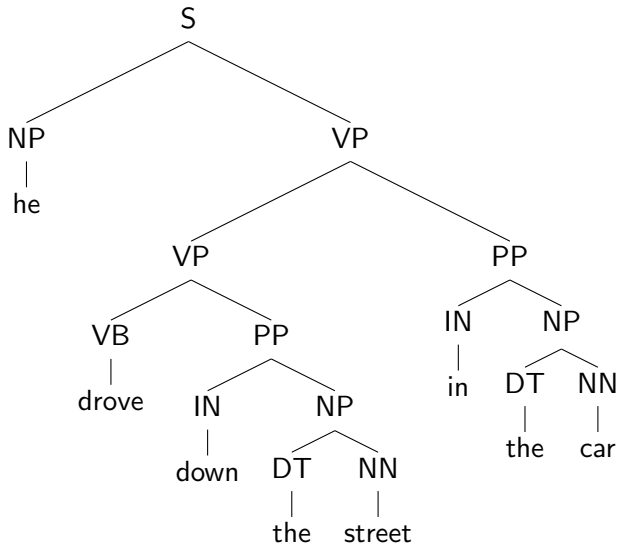
$VB \rightarrow \text{laughs}$



## Properties of CFGs

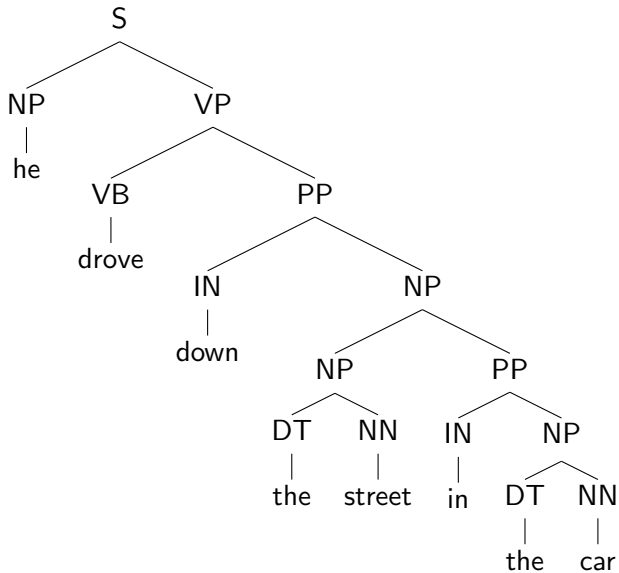
- ▶ A CFG defines a set of possible derivations
- ▶ A string  $s \in \Sigma^*$  is in the *language* defined by the CFG if there is at least one derivation that yields  $s$
- ▶ Each string in the language generated by the CFG may have more than one derivation (“ambiguity”)

## An Example of Ambiguity





## An Example of Ambiguity (continued)



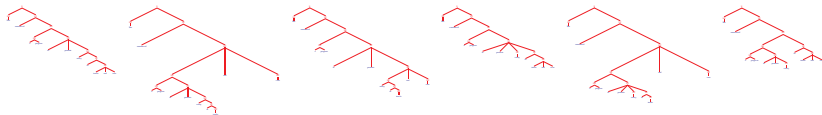
# The Problem with Parsing: Ambiguity

INPUT:

She announced a program to promote safety in trucks and vans



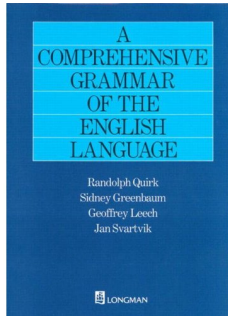
POSSIBLE OUTPUTS:



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 专有名词
  - 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

$\bar{N} \Rightarrow NN$	$NN \Rightarrow box$	$JJ \Rightarrow fast$
$\bar{N} \Rightarrow NN \quad \bar{N}$	$NN \Rightarrow car$	$JJ \Rightarrow metal$
$\bar{N} \Rightarrow JJ \quad \bar{N}$	$NN \Rightarrow mechanic$	$JJ \Rightarrow idealistic$
$\bar{N} \Rightarrow \bar{N} \quad \bar{N}$	$NN \Rightarrow pigeon$	$JJ \Rightarrow clay$
$NP \Rightarrow DT \quad \bar{N}$		$IN \Rightarrow in$
	$DT \Rightarrow the$	$IN \Rightarrow under$
$PP \Rightarrow IN \quad NP$	$DT \Rightarrow a$	$IN \Rightarrow of$
$\bar{N} \Rightarrow \bar{N} \quad PP$		$IN \Rightarrow on$
		$IN \Rightarrow with$
		$IN \Rightarrow as$

## Generates:

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



# An Extended 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	$\bar{N}$
PP	$\Rightarrow$	IN	NP
$\bar{N}$	$\Rightarrow$	$\bar{N}$	PP

# 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

VP → Vi

VP → Vt NP

VP → Vd NP NP

- ▶ Basic S Rule

S → NP 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 标句词  
COMP = complementizer e.g., that
- ▶ SBAR 非限定状态  
SBAR → 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

VP → V[5] SBAR

VP → V[6] NP SBAR

VP → V[7] NP NP SBAR

### **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

并列连词

- ▶ New Rules

NP        →    NP        CC    NP

$\bar{N}$         →     $\bar{N}$         CC     $\bar{N}$

VP        →    VP        CC    VP

S         →    S         CC    S

SBAR     →    SBAR    CC    SBAR

# We've Only Scratched the Surface...

- ▶ Agreement

The dogs laugh vs. The dog laughs

- ▶ Wh-movement

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

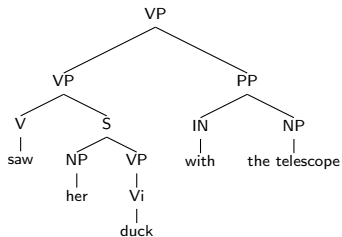
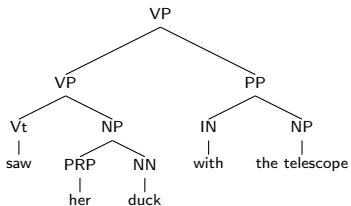


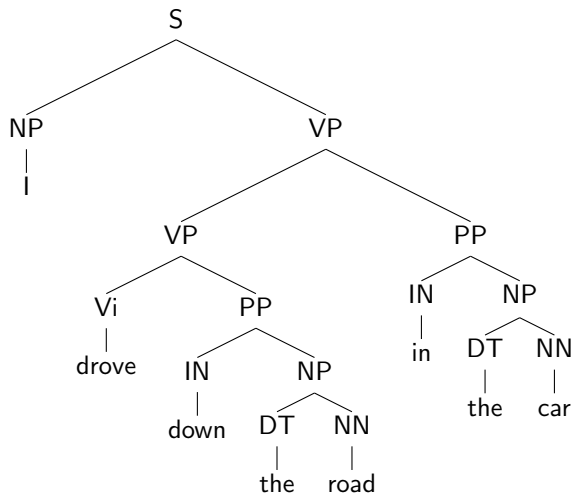
# Sources of Ambiguity

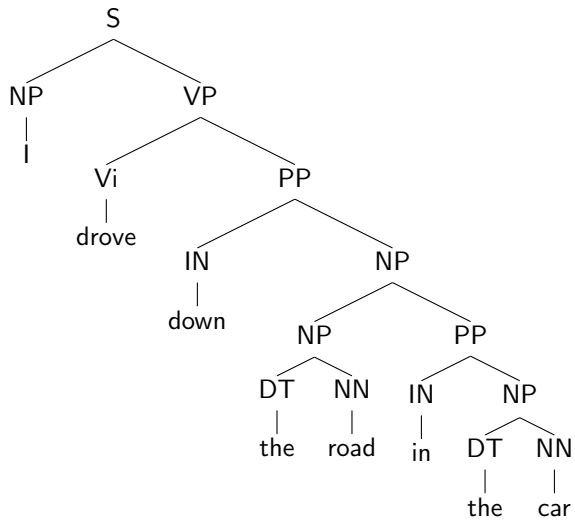
- Part-of-Speech ambiguity

NN → duck

Vi → duck







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

# Sources of Ambiguity: Noun Premodifiers

- Noun premodifiers:

