

A decorative vertical bar on the left side of the slide, composed of two parallel lines in shades of orange and brown.

# Formal Grammars of English

JURAFSKY AND MARTIN CHAPTER 10

*Based on slides from Sameer Singh, Dan Jurafsky, Noah Smith, Slav Petrov, and everyone else they copied from.*

# Limitations of Sequence Tags

John Smith shot Bill in his pajamas.

NNP NNP VBD NNP IN PRP\$ NNS .  
John Smith shot Bill in his pajamas.

Person Pers  
John Smith shot Bill in his pajamas.

What happened?  
Who shot who?  
Who was wearing the pajamas?

# How do words group together in English?

A **noun phrase** is a sequence of words surrounding at least one noun.

What evidence do we have that these words group together?

## Noun Phrases

**Harry the Horse**

**the Broadway coppers**

**they**

**a high-class spot such as Mindy's**

**the reason he comes into the Hot Box**

**three parties from Brooklyn**

# Constituents

- Constituent behave as a unit that can be rearranged:

John talked [to the children] [about drugs].

John talked [about drugs] [to the children].

John talked drugs to the children about

- Or substituted/expanded:

John talked [to the children taking the drugs] [about alcohol].

Harry the Horse

a high-class spot such as Mindy's

the Broadway coppers

the reason he comes into the Hot Box

they

three parties from Brooklyn

X

arrive(s)

attract(s)

love(s)

sit(s)

“Noun phrases appear before verbs in English.”

# Constituents and Grammars

## Grammar

- Tells you how the constituents can be arranged
- Implicit knowledge for us (we often can't tell *why* something is wrong)
- Generate all, and only, the possible sentences of the language
- Different from meaning:

Colorless green ideas sleep furiously.

- The words are in the right order,
- And that ideas are green and colorless,
- And that ideas sleep,
- And that sleeping is done furiously,
- As opposed to: “sleep green furiously ideas colorless”

# Uses of Parsing

[ send [the text message from James] [to Sharon] ]

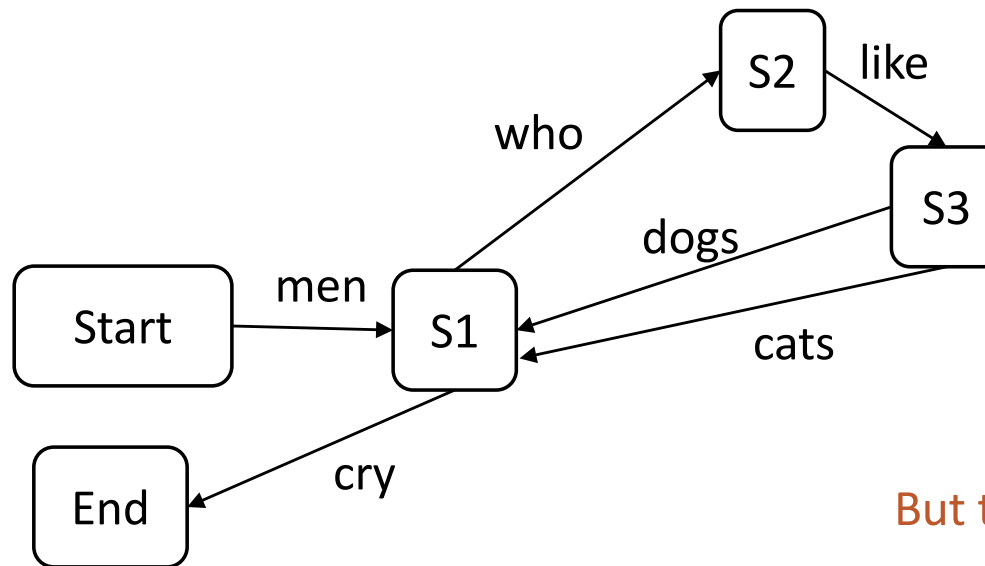
[ translate [the message] [from Hindi] [to English] ]

- Grammar checkers
- Dialog systems
- High precision question answering
- Named entity recognition
- Sentence compression
- Extracting opinions about products
- Improved interaction in computer games
- Helping linguists find data
- Machine translation
- Relation extraction systems

# Basic Grammar: Regular Expr.

- You can capture individual words:
  - (man|dog|cat)
- Simple sentences:
  - (man|dog|cat)(ate|loves|consumed)(.|food|lunch)
- Infinite length? Yes!
  - men (who like (cats|dogs))\* cry.

Finite State  
Machine



But too weak for English.

# Context-Free Grammars

Grammar, G

T

Terminal Symbols  
words

{man, dog, food...}

N "Constituents"

Non-terminal Symbols

S, NP, VP, ...

Rules

$A \rightarrow BCD$

$A \rightarrow w$

$A \rightarrow BAB$

$A \rightarrow a \quad B \rightarrow b$

"a"

"bab" "baba" — x

$\rightarrow b^* a b^* \leftarrow$

Grammar applies rules recursively..

If we can construct the input sentence, it is in the grammar, otherwise not.



# Context Free Grammars

$$G \equiv \langle T, N, R, S \rangle$$

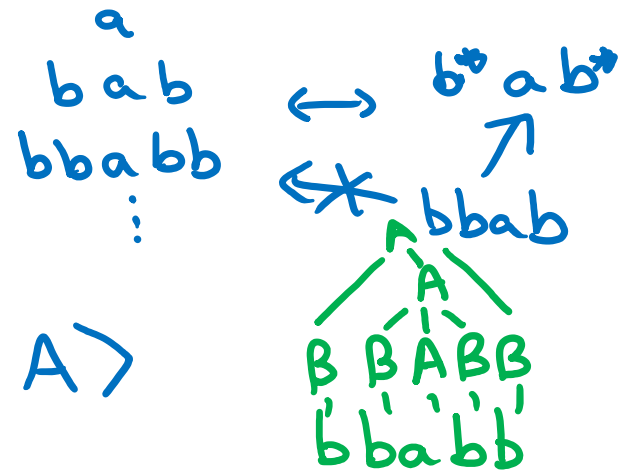
words      constituents      how to compose them      sentence

$$A \rightarrow ABA$$

$$A \rightarrow a$$

$$B \rightarrow b$$

$$G \equiv \langle \{a, b\}, \{A, B\}, R, A \rangle$$



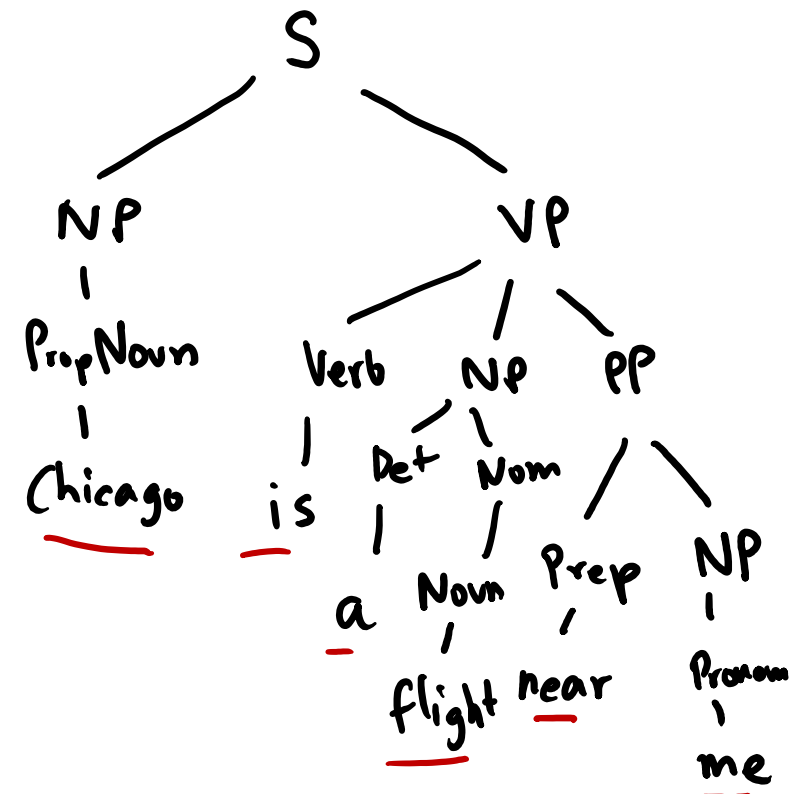
"Lexicon"

# Example CFG

$A \rightarrow w_1 \mid w_2$   
 $A \rightarrow w_1 \quad A \rightarrow w_2$

Noun  $\rightarrow$  flights | breeze | trip | morning  
 Verb  $\rightarrow$  is | prefer | like | need | want | fly  
 Adjective  $\rightarrow$  cheapest | non-stop | first | latest  
                   | other | direct  
 Pronoun  $\rightarrow$  me | I | you | it  
 Proper-Noun  $\rightarrow$  Alaska | Baltimore | Los Angeles  
                   | Chicago | United | American  
 Determiner  $\rightarrow$  the | a | an | this | these | that  
 Preposition  $\rightarrow$  from | to | on | near  
 Conjunction  $\rightarrow$  and | or | but

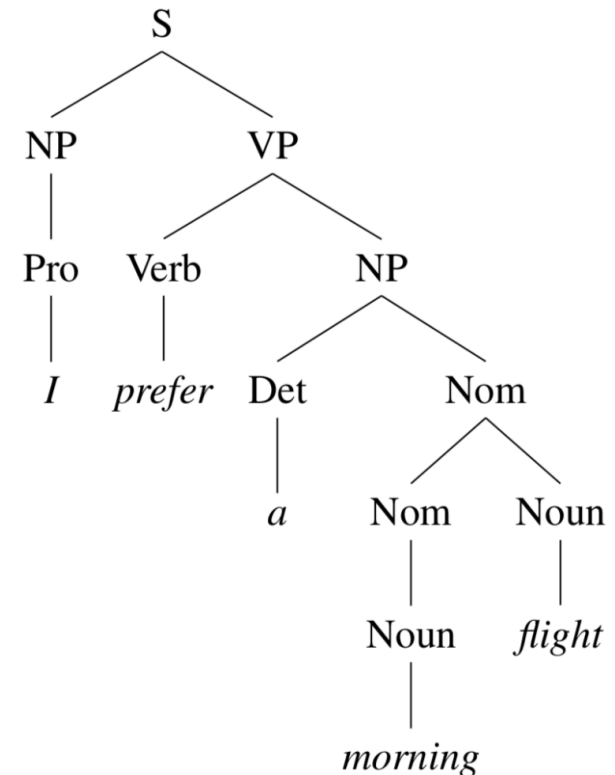
Grammar Rules	Examples
$S \rightarrow NP VP$	I + want a morning flight
$NP \rightarrow$ Pronoun	I
Proper-Noun	Los Angeles
Det Nominal	a + flight
Nominal $\rightarrow$ Nominal Noun	morning + flight
Noun	flights
$VP \rightarrow$ Verb	do
Verb NP	want + a flight
Verb NP PP	leave + Boston + in the morning
Verb PP	leaving + on Thursday
$PP \rightarrow$ Preposition NP	from + Los Angeles



# Example Parse Tree

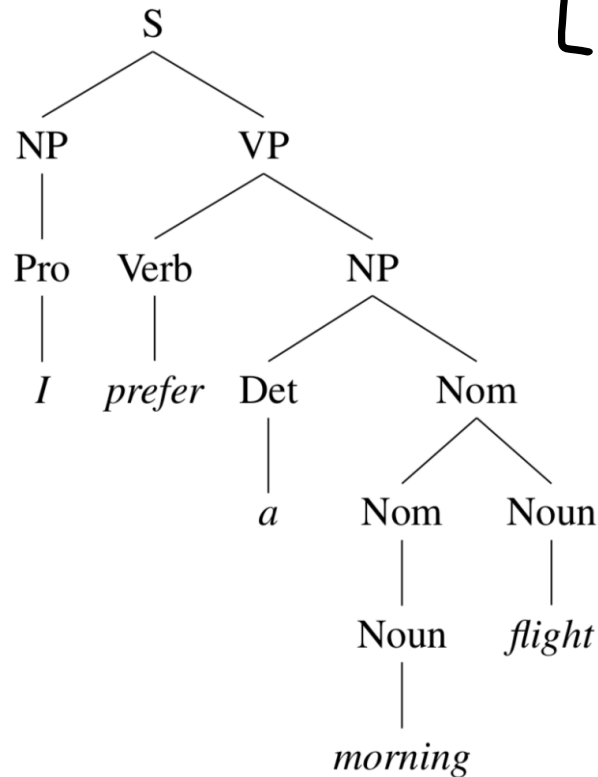
I prefer a morning flight.

Grammar Rules		Examples
$S \rightarrow NP VP$		I + want a morning flight
$NP \rightarrow$	<i>Pronoun</i>	I
	<i>Proper-Noun</i>	Los Angeles
	<i>Det Nominal</i>	a + flight
	<i>Nominal</i>	morning + flight
$Nominal \rightarrow$	<i>Nominal Noun</i>	morning + flight
	<i>Noun</i>	flights
$VP \rightarrow$	<i>Verb</i>	do
	<i>Verb NP</i>	want + a flight
	<i>Verb NP PP</i>	leave + Boston + in the morning
	<i>Verb PP</i>	leaving + on Thursday
$PP \rightarrow$	<i>Preposition NP</i>	from + Los Angeles



# Example Parse Tree: Brackets

I prefer a morning flight.



$[S [NP [pro\ I]]] [VP [verb\ prefer] [NP [det\ a] .]]]$

# More details: Noun Phrases

## Simple Noun Phrases

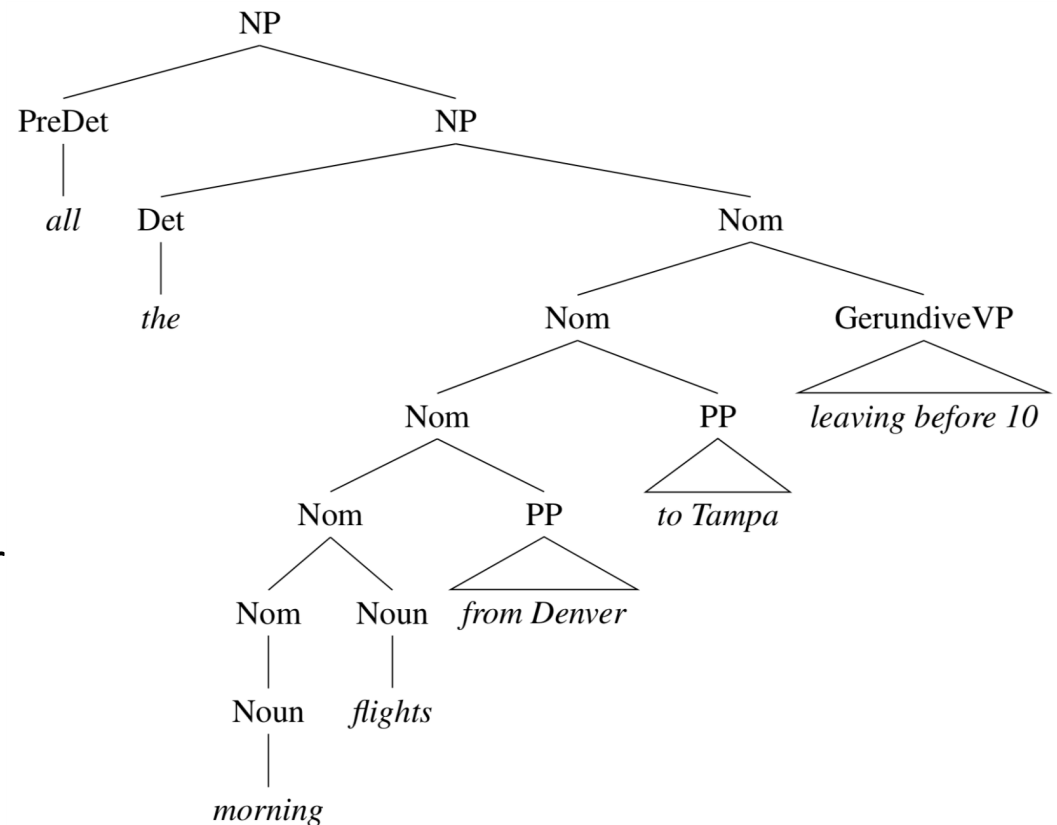
*NP → ProperNoun*

*NP → Det Nominal*

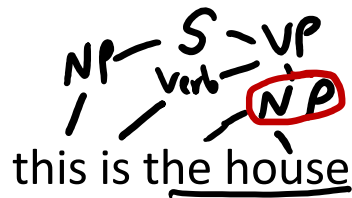
*Nominal → Noun / Noun Nominal*

## Complex Noun Phrases

“all the morning flights from Denver to Tampa leaving before 10”



# Recursive Noun Phrases



this is the house that Jack built

this is the cat that lives in the house that Jack built

this is the dog that chased the cat that lives in the house that Jack built

this is the flea that bit the dog that chased the cat that lives in the house the Jack built

this is the virus that infected the flea that bit the dog that chased the cat that lives in the house that Jack built

# More details: Verb Phrases

## Simple Verb Phrases

$VP \rightarrow Verb$

$VP \rightarrow Verb\ NP$

$VP \rightarrow Verb\ NP\ PP$

$VP \rightarrow Verb\ PP$

disappear

prefer a morning flight

leave Boston in the morning

leave in the morning

But all verbs are not the same!  
(this grammar overgenerates)

**Solution:** subcategorize!

**Sneezed:** John sneezed.

**Find:** Please find a flight to NY.

**Give:** Give me a cheaper fare.

**Help:** Can you help me with a flight?

**Prefer:** I prefer to leave earlier.

**Told:** I was told United has a  
flight.

# Types of Sentences

Declarative

$S \rightarrow NP VP$

A plane left.

Imperative

$S \rightarrow VP$

Show the plane.

Yes/no Questions

$S \rightarrow Aux NP VP$

Did the plane leave?

Aux

Wh-Questions

$S \rightarrow \underline{WhNP} Aux NP VP$

When did the plane leave?

WhNP



# Source of Grammar?

Manual



Noam Chomsky

Write symbolic grammar (CFG or often richer) and lexicon

$S \rightarrow NP VP$

$NN \rightarrow \textit{interest}$

$NP \rightarrow (DT) NN$

$NNS \rightarrow \textit{rates}$

$NP \rightarrow NN NNS$

$NNS \rightarrow \textit{raises}$

$NP \rightarrow NNP$

$VP \rightarrow \textit{interest}$

$VP \rightarrow V NP$

$VBZ \rightarrow \textit{rates}$

Used grammar/proof systems to prove parses from words

*Fed raises interest rates 0.5% in effort to control inflation*

- Minimal grammar: 36 parses
- Simple 10 rule grammar: 592 parses
- Real-size broad-coverage grammar: millions of parses

# Source of Grammar?

From data!

## The Penn Treebank

Building a treebank seems a lot slower and less useful than building a grammar

But a treebank gives us many things

- Reusability of the labor
  - Many parsers, POS taggers, etc.
  - Valuable resource for linguistics
- Broad coverage
- Frequencies and distributional information
- A way to evaluate systems

[Marcus et al. 1993, *Computational Linguistics*]

```

( (S
  (NP-SBJ (DT The) (NN move))
  (VP (VBD followed)
    (NP
      (NP (DT a) (NN round))
      (PP (IN of)
        (NP
          (NP (JJ similar) (NNS increases))
          (PP (IN by)
            (NP (JJ other) (NNS lenders)))
          (PP (IN against)
            (NP (NNP Arizona) (JJ real) (NN estate) (NNS loans)))))))
    (, ,)
    (S-ADV
      (NP-SBJ (-NONE- *))
      (VP (VBG reflecting)
        (NP
          (NP (DT a) (VBG continuing) (NN decline))
          (PP-LOC (IN in)
            (NP (DT that) (NN market)))))))
    (. .)))

```