ANTONIS ANASTASOPOULOS

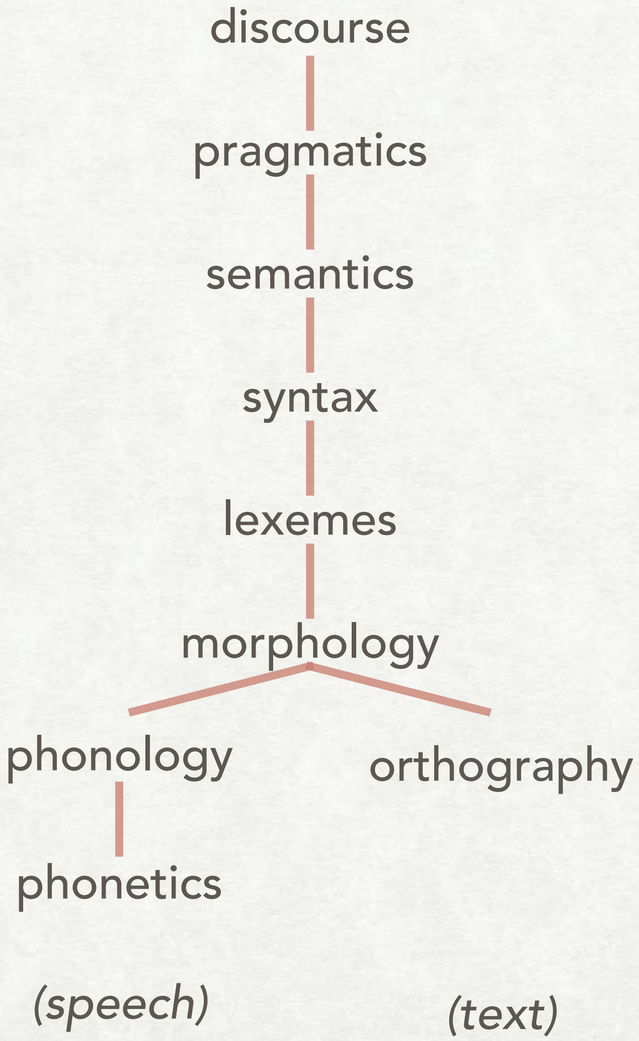
CS499 INTRODUCTION TO NLP

PARSING WITH CFGs

https://cs.gmu.edu/~antonis/course/cs499-spring21/

With adapted slides by David Mortensen and Alan Black

LEVELS OF REPRESENTATION



STRUCTURE OF THIS LECTURE

1. Context-free Grammars
2. Parsing Algorithms
3. Chomsky Normal Form
4. CKY Algorithm

CONTEXT-FREE GRAMMARS

Using grammars:

- recognition
- parsing

Parsing algorithms:

- top down
- bottom up

Chomsky Normal Form (CNF)

CKY Algorithm (Cocke-Younger-Kasami)

PARSING VS WORD MATCHING

Consider the following sentence:

The student who was taught be Dr. A won the prize.

Who won the prize?

String matching:

"Dr. A won the prize."

Parsing based

[[The student [who was taught be Dr. A]] won the prize.]

"the student won the prize"

CONTEXT-FREE GRAMMARS

Vocabulary of terminal symbols: Σ

Set of non-terminal symbols (aka variables): N

Special start symbols: $S \in N$

Production rules of the form $X \rightarrow \alpha$, where

$X \in N$

$\alpha \in (N \cup \Sigma)^*$

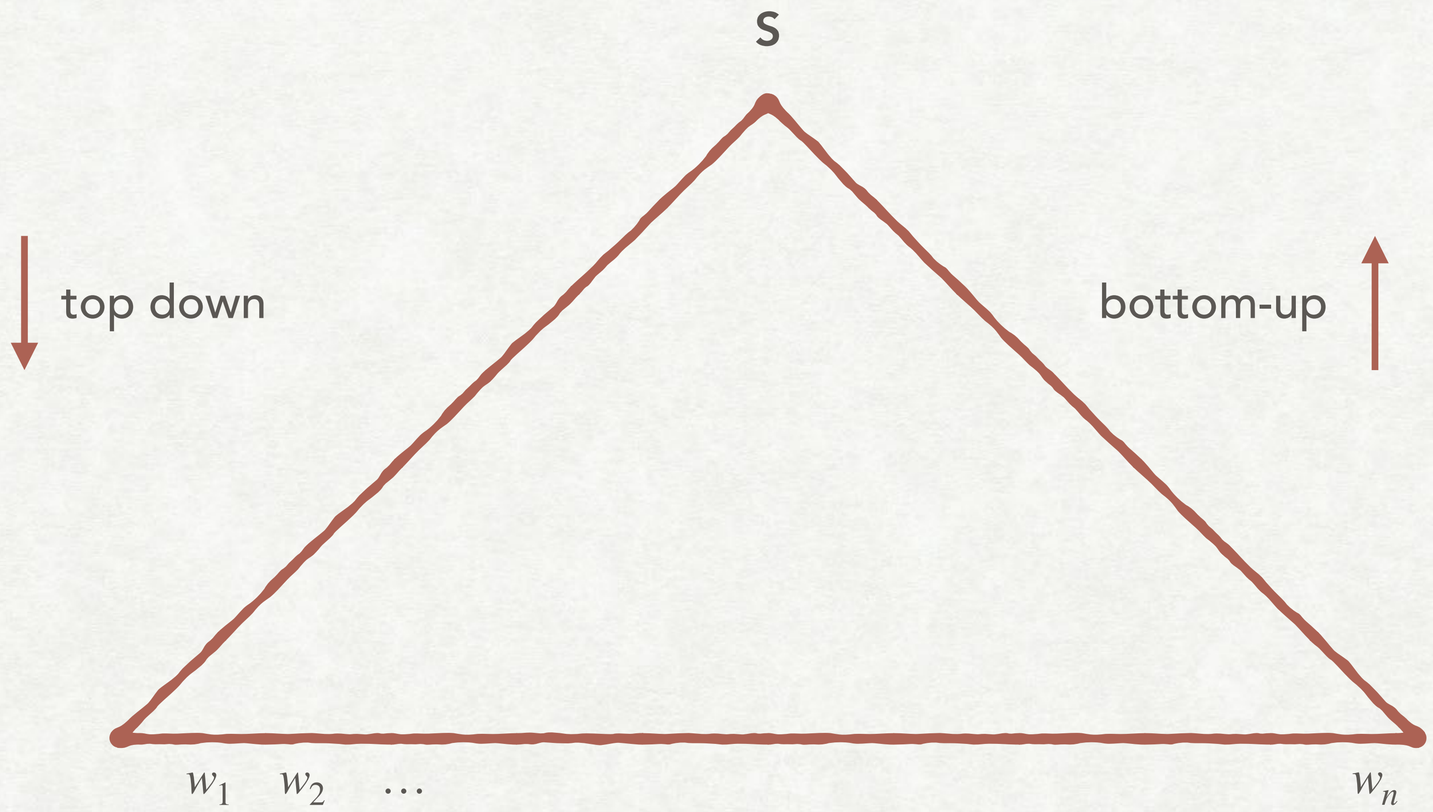
TWO RELATED PROBLEMS

Input: sentence $\mathbf{w} = (w_1, \ldots w_n)$ and CFG $\mathcal{G}$

Output (recognition): tree iff $\mathbf{w} \in$ Language($\mathcal{G}$)

Output (parsing) one or more derivations for $\mathbf{w}$, under $\mathcal{G}$.

PARSING AS SEARCH



RECOGNIZERS AS SEARCH

```
Agenda = { state0 }

while(Agenda not empty)
  s = pop a state from Agenda
  if s is a success-state return s // valid parse tree
  else if s is not a failure-state:
    generate new states from s
    push new states onto Agenda

return nil
```

EXAMPLE GRAMMAR AND LEXIVON

Grammar	Lexicon
$S \rightarrow NP\ VP$	$Det \rightarrow that \mid this \mid a$
$S \rightarrow Aux\ NP\ VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$Proper\text{-}Noun \rightarrow Houston \mid NWA$
$NP \rightarrow Det\ Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal\ Noun$	
$Nominal \rightarrow Nominal\ PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb\ NP$	
$VP \rightarrow Verb\ NP\ PP$	
$VP \rightarrow Verb\ PP$	
$VP \rightarrow VP\ PP$	
$PP \rightarrow Preposition\ NP$	

Figure 13.1 The $\mathscr{L}_1$ miniature English grammar and lexicon.

CHOMSKY NORMAL FORM

CONTEXT-FREE GRAMMARS IN CHOMSKY NORMAL FORM

Vocabulary of terminal symbols: Σ

Set of non-terminal symbols (aka variables): N

Special start symbols: $S \in N$

Production rules of the form $X \rightarrow \alpha$, where

$X \in N$

$\alpha \in N, N \cup \Sigma$

CONVERTING CFGS TO CNF

For each rule

$$X \rightarrow A\ B\ C$$

Rewrite as:

$$X \rightarrow A\ X_2$$

$$X_2 \rightarrow B\ C$$

(Introduces a new non-terminal)

$\mathcal{L}_1$ Grammar	$\mathcal{L}_1$ in CNF
$S \rightarrow NP\ VP$	$S \rightarrow NP\ VP$
$S \rightarrow Aux\ NP\ VP$	$S \rightarrow X1\ VP$
	$X1 \rightarrow Aux\ NP$
$S \rightarrow VP$	$S \rightarrow book \mid include \mid prefer$
	$S \rightarrow Verb\ NP$
	$S \rightarrow X2\ PP$
	$S \rightarrow Verb\ PP$
	$S \rightarrow VP\ PP$
$NP \rightarrow Pronoun$	$NP \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$NP \rightarrow TWA \mid Houston$
$NP \rightarrow Det\ Nominal$	$NP \rightarrow Det\ Nominal$
$Nominal \rightarrow Noun$	$Nominal \rightarrow book \mid flight \mid meal \mid money$
$Nominal \rightarrow Nominal\ Noun$	$Nominal \rightarrow Nominal\ Noun$
$Nominal \rightarrow Nominal\ PP$	$Nominal \rightarrow Nominal\ PP$
$VP \rightarrow Verb$	$VP \rightarrow book \mid include \mid prefer$
$VP \rightarrow Verb\ NP$	$VP \rightarrow Verb\ NP$
$VP \rightarrow Verb\ NP\ PP$	$VP \rightarrow X2\ PP$
	$X2 \rightarrow Verb\ NP$
$VP \rightarrow Verb\ PP$	$VP \rightarrow Verb\ PP$
$VP \rightarrow VP\ PP$	$VP \rightarrow VP\ PP$
$PP \rightarrow Preposition\ NP$	$PP \rightarrow Preposition\ NP$

Figure 13.8 $\mathcal{L}_1$ Grammar and its conversion to CNF. Note that although they aren't shown here all the original lexical entries from $\mathcal{L}_1$ carry over unchanged as well.

CKY ALGORITHM

CKY ALGORITHM

For $i = [1 \ \ldots \ n]$

$C[i-1, i] = \{V \mid V \rightarrow w_i\}$

For $l = 2 \ \ldots \ n$: // width

For $i = 0 \ \ldots \ n - l$: // left boundary

$k = i + l$ // right boundary

For $j = i + 1 \ \ldots \ k - 1$: // midpoint

$C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$

Return true if $S \in C[0,n]$

CKY: CHART

Book					
	this				
		flight			
			through		
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun				
Book					
	this				
		flight			
			through		
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb				
Book					
	this				
		flight			
			through		
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb				
Book		Det			
	this		Noun		
		flight		Prep	
			through		PNoun
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb				
Book		Det			
	this		Noun		
		flight		Prep	
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––			
Book		Det			
	this		Noun		
		flight		Prep	
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	--			
Book		Det	NP		
	this		Noun		
		flight		Prep	
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––			
Book		Det	NP		
	this		Noun	––	
		flight		Prep	
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––			
Book		Det	NP	––	
	this		Noun	––	
		flight		Prep	
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	--			
Book		Det	NP	--	
	this		Noun	--	
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––			
Book		Det	NP	––	
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––			
Book		Det	NP	––	NP
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ,\ Y \in C[i,j],\ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––	VP		
Book		Det	NP	––	NP
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––	VP, S		
Book		Det	NP	––	NP
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––	VP, S	––	
Book		Det	NP	––	NP
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i+1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY: CHART

	Noun, Verb	––	VP, S	––	S
Book		Det	NP	––	NP
	this		Noun	––	––
		flight		Prep	PP
			through		PNoun, NP
				Houston	

For $i = [1 \ \ldots \ n]$
  $C[i-1, i] = \{V \mid V \rightarrow w_i\}$
For $l = 2 \ \ldots \ n$: **// width**
  For $i = 0 \ \ldots \ n-l$: **// left boundary**
    $k = i + l$ **// right boundary**
    For $j = i + 1 \ \ldots \ k-1$: **// midpoint**
      $C[i,k] = C[i,k] \cup \{V \mid V \rightarrow YZ, \ Y \in C[i,j], \ Z \in C[j,k]\}$
Return true if $S \in C[0,n]$

CKY EQUATIONS

$$C[i-1, i, w_i] = \text{TRUE}$$

$$C[i-1, i, V] = \begin{cases} \text{TRUE} & \text{if } V \rightarrow w_i \\ \text{FALSE} & \text{otherwise} \end{cases}$$

$$C[i, j, V] = \begin{cases} \text{TRUE} & \text{if } \exists j, Y, Z \text{ such that} \\ & V \rightarrow Y\,Z \\ & \quad \text{and } C[i, k, Y] \\ & \quad \text{and } C[k, j, Z] \\ & \quad \text{and } i < k < j \\ \text{FALSE} & \text{otherwise} \end{cases}$$

$$\text{goal} = C[0, n, S]$$

CKY COMPLEXITY

Worst case: $\mathcal{O}(n^3 \cdot |\mathcal{G}|)$

Best is also worst case

(Others better in average case)

NEXT CLASS

Treebanks and probabilistic CFGs