Algorithms for NLP

CS 11711, Fall 2019

Parsing

Yulia Tsvetkov
Ambiguity

- I saw a girl with a telescope
Syntactic Parsing

- **INPUT:**
  - The move followed a round of similar increases by other lenders, reflecting a continuing decline in that market

- **OUTPUT:**
Canadian Utilities had 1988 revenue of $1.16 billion, mainly from its natural gas and electric utility businesses in Alberta, where the company serves about 800,000 customers.
Outline

- Syntax:
  - Intro
  - Context Free Grammars (CFGs)
  - Probabilistic Context Free Grammars (PCFGs)
- Parsing with CFGs
- Parsing with PCFGs
- Evaluation of parsers
Syntax
Syntax

- The study of the patterns of formation of sentences and phrases from word
  - my dog Pron N
  - the dog Det N
  - the cat Det N

- and Conj

- the large cat Det Adj N
- the black cat Det Adj N

- ate a sausage V Det N
Parsing

- The process of predicting syntactic representations
- Syntactic Representations
  - Different types of syntactic representations are possible, for example:

```
  S
 /\  
NP  VP
 /  /
PN  V  NP
|    |
My  ate  a
|    |
dog
|    |
sausage
```

Constituent (a.k.a. phrase-structure) tree
Constituent trees

- Internal nodes correspond to phrases
  - S – a sentence
  - NP (Noun Phrase): My dog, a sandwich, lakes, ...
  - VP (Verb Phrase): ate a sausage, barked, ...
  - PP (Prepositional phrases): with a friend, in a car, ...

- Nodes immediately above words are PoS tags (aka preterminals)
  - PN – pronoun
  - D – determiner
  - V – verb
  - N – noun
  - P – preposition
- It is often convenient to represent a tree as a bracketed sequence

(S
  (NP (PN My) (N Dog))
  (VP (V ate)
    (NP (D a) (N sausage))
  )
)
The process of predicting syntactic representations

Syntactic Representations

Different types of syntactic representations are possible, for example:

Constituent (a.k.a. phrase-structure) tree

Dependency tree
Dependency trees

- Nodes are words (along with PoS tags)
- Directed arcs encode syntactic dependencies between them
- Labels are types of relations between the words
  - poss – possessive
  - dobj – direct object
  - nsub - subject
  - det - determiner
Some semantic information can be (approximately) derived from syntactic information

- Subjects (nsubj) are (often) agents ("initiator / doers for an action")
- Direct objects (dobj) are (often) patients ("affected entities")
Some semantic information can be (approximately) derived from syntactic information
- Subjects (nsubj) are (often) agents ("initiator / doers for an action")
- Direct objects (dobj) are (often) patients ("affected entities")

But even for agents and patients consider:
- Mary is baking a cake in the oven
- A cake is baking in the oven

In general it is not trivial even for the most shallow forms of semantics
- E.g., consider prepositions: in can encode direction, position, temporal information, …
Constituent and dependency representations

- Constituent trees can (potentially) be converted to dependency trees

- Dependency trees can (potentially) be converted to constituent trees
Constituent trees

- Internal nodes correspond to phrases
  - S – a sentence
  - NP (Noun Phrase): My dog, a sandwich, lakes, ...
  - VP (Verb Phrase): ate a sausage, barked, ...
  - PP (Prepositional phrases): with a friend, in a car, ...

- Nodes immediately above words are PoS tags (aka preterminals)
  - PN – pronoun
  - D – determiner
  - V – verb
  - N – noun
  - P – preposition
Constituency Tests

- How do we know what nodes go in the tree?

- Classic constituency tests:
  - Replacement
  - Substitution by *proform*
  - Movement
    - Clefting
    - Preposing
    - Passive
  - Modification
  - Coordination/Conjunction
  - Ellipsis/Deletion
Conflicting Tests

- Constituency isn’t always clear
  - Coordination
    - He went to and came from the store
  - Phonological reduction:
    - I will go → I’ll go
    - I want to go → I wanna go
    - a le centre → au centre
Morphology/Syntax/Semantics

- The study of the patterns of formation of sentences and phrases from word
  - Borders with semantics and morphology sometimes blurred

Afyonkarahisarlılaştırabilir bildiklerimizdenmişsinizcesiin

in Turkish means

"as if you are one of the people that we thought to be originating from Afyonkarahisar" [wikipedia]
CFGs
Context Free Grammar (CFG)

**Grammar (CFG)**

- **ROOT** → **S**
- **S** → **NP VP**
- **NP** → **DT NN**
- **NP** → **NN NNS**

**Lexicon**

- **NP** → **NP PP**
- **VP** → **VBP NP**
- **VP** → **VBP NP PP**
- **PP** → **IN NP**

- **NN** → **interest**
- **NNS** → **raises**
- **VBP** → **interest**
- **VBZ** → **raises**

- **Other grammar formalisms: LFG, HPSG, TAG, CCG…**
CFGs

\[ S \rightarrow NP \ V F \]

\[ VP \rightarrow V \]
\[ VP \rightarrow V \ NF \]
\[ VP \rightarrow VP \ PF \]
\[ NP \rightarrow NP \ PF \]
\[ NP \rightarrow D \ N \]
\[ NP \rightarrow PN \]
\[ PP \rightarrow P \ NF \]

\[ N \rightarrow girl \]
\[ N \rightarrow telescope \]
\[ N \rightarrow sandwich \]
\[ PN \rightarrow I \]
\[ V \rightarrow saw \]
\[ V \rightarrow ate \]
\[ P \rightarrow with \]
\[ P \rightarrow in \]
\[ D \rightarrow a \]
\[ D \rightarrow the \]
CFGs

\[
\begin{align*}
S & \rightarrow NP \ VF \\
VP & \rightarrow V \\
VP & \rightarrow V \ NF \\
VP & \rightarrow VP \ PF \\
NP & \rightarrow NP \ PF \\
NP & \rightarrow D \ N \\
NP & \rightarrow PN \\
PP & \rightarrow P \ NF \\
N & \rightarrow girl \\
N & \rightarrow telescope \\
N & \rightarrow sandwich \\
PN & \rightarrow I \\
V & \rightarrow saw \\
V & \rightarrow ate \\
P & \rightarrow with \\
P & \rightarrow in \\
D & \rightarrow a \\
D & \rightarrow the
\end{align*}
\]
CFGs

```
S → NP VP

VP → V
VP → VP NF

NP → NP PF
NP → D N
NP → PN

PP → P NF

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the
```
CFGs

\[
S \rightarrow NP \ VF \\
VP \rightarrow V \\
VP \rightarrow V \ NF \\
VP \rightarrow VP \ PF \\
NP \rightarrow NP \ PF \\
NP \rightarrow D \ N \\
NP \rightarrow PN \\
PP \rightarrow P \ NF \\
N \rightarrow girl \\
N \rightarrow telescope \\
N \rightarrow sandwich \\
PN \rightarrow I \\
V \rightarrow saw \\
V \rightarrow ate \\
P \rightarrow with \\
P \rightarrow in \\
D \rightarrow a \\
D \rightarrow the
\]
CFGs

\[ S \rightarrow NP \ VF \]
\[ N \rightarrow \text{girl} \]
\[ N \rightarrow \text{telescope} \]
\[ N \rightarrow \text{sandwich} \]
\[ VP \rightarrow V \]
\[ VP \rightarrow V \ NF \]
\[ VP \rightarrow VP \ PF \]
\[ PN \rightarrow I \]
\[ V \rightarrow \text{saw} \]
\[ V \rightarrow \text{ate} \]
\[ P \rightarrow with \]
\[ P \rightarrow in \]
\[ D \rightarrow a \]
\[ D \rightarrow the \]

\[ NP \rightarrow NP \ PF \]
\[ NP \rightarrow D \ N \]
\[ NP \rightarrow PN \]
\[ PP \rightarrow P \ NF \]
CFGs

\[ S \rightarrow NP \ VF \]
\[ VP \rightarrow V \]
\[ VP \rightarrow V \ NP \]
\[ VP \rightarrow VP \ PF \]
\[ NP \rightarrow NP \ PF \]
\[ NP \rightarrow D \ N \]
\[ NP \rightarrow PN \]
\[ PP \rightarrow P \ NF \]

\[ N \rightarrow girl \]
\[ N \rightarrow telescope \]
\[ N \rightarrow sandwich \]
\[ PN \rightarrow I \]
\[ V \rightarrow saw \]
\[ V \rightarrow ate \]
\[ P \rightarrow with \]
\[ P \rightarrow in \]
\[ D \rightarrow a \]
\[ D \rightarrow the \]
CFGs

$S \rightarrow NP \ VF$

$N \rightarrow girl$

$VP \rightarrow V$

$N \rightarrow telescope$

$VP \rightarrow V \ NF$

$N \rightarrow sandwich$

$VP \rightarrow VP \ PF$

$PN \rightarrow I$

$NP \rightarrow NP \ PF$

$V \rightarrow saw$

$NP \rightarrow D \ N$

$V \rightarrow ate$

$NP \rightarrow PN$

$P \rightarrow with$

$PP \rightarrow P \ NF$

$P \rightarrow in$

$D \rightarrow a$

$D \rightarrow the$
CFGs

\[ S \rightarrow NP \ VP \]
\[ NP \rightarrow D \ N \]
\[ V \rightarrow saw \]
\[ VP \rightarrow V \ NP \]
\[ PP \rightarrow P \ NP \]
\[ PN \rightarrow I \]

N \rightarrow girl
N \rightarrow telescope
N \rightarrow sandwich
P \rightarrow with
P \rightarrow in
D \rightarrow a
D \rightarrow the

\[ V \rightarrow saw \]
\[ V \rightarrow ate \]
\[ NP \rightarrow D \ N \]
\[ NP \rightarrow PN \]
\[ VP \rightarrow VP \ PP \]
\[ PP \rightarrow P \ NP \]
(S (NP-SBJ The move)
  (VP followed
   (NP (NP a round)
     (PP of
      (NP (NP similar increases)
        (PP by
          (NP other lenders))
        (PP against
          (NP Arizona real estate loans))))))

(S-ADV (NP-SBJ *))
  (VP reflecting
   (NP (NP a continuing decline)
     (PP-LOC in
      (NP that market))))

.)
A context-free grammar is a 4-tuple $<N, T, S, R>$

- $N$ : the set of non-terminals
  - Phrasal categories: S, NP, VP, ADJP, etc.
  - Parts-of-speech (pre-terminals): NN, JJ, DT, VB
- $T$ : the set of terminals (the words)
- $S$ : the start symbol
  - Often written as ROOT or TOP
  - Not usually the sentence non-terminal $S$
- $R$ : the set of rules
  - Of the form $X \rightarrow Y_1 Y_2 \ldots Y_k$, with $X, Y_i \in N$
  - Examples: $S \rightarrow NP \ VP$, $VP \rightarrow VP CC VP$
  - Also called rewrites, productions, or local trees
An example grammar

\( N = \{S, VP, NP, PP, N, V, PN, P\} \)

\( T = \{girl, telescope, sandwich, I, saw, ate, with, in, a, the\} \)

\( S = \{S\} \)

\( R \)

\( S \rightarrow NP \ VP \) (NP A girl) (VP ate a sandwich)

\( VP \rightarrow V \)

\( VP \rightarrow V \ NF \) (V ate) (NP a sandwich)

\( VP \rightarrow VP \ PP \) (VP saw a girl) (PP with a telescope)

\( NP \rightarrow NP \ PP \) (NP a girl) (PP with a sandwich)

\( NP \rightarrow D \ N \) (D a) (N sandwich)

\( NP \rightarrow PN \)

\( PP \rightarrow P \ NF \) (P with) (NP with a sandwich)

Preterminal rules

\( N \rightarrow girl \)

\( N \rightarrow telescope \)

\( N \rightarrow sandwich \)

\( PN \rightarrow I \)

\( V \rightarrow saw \)

\( V \rightarrow ate \)

\( P \rightarrow with \)

\( P \rightarrow in \)

\( D \rightarrow a \)

\( D \rightarrow the \)
What can be a sub-tree is only affected by what the phrase type is (VP) but not the context.
What can be a sub-tree is only affected by what the phrase type is (VP) but not the context.

Not grammatical
Ambiguities
Coordination ambiguity

- Here, the coarse VP and NP categories cannot enforce subject-verb agreement in number resulting in the coordination ambiguity

"Bark" can refer both to a noun or a verb

This tree would be ruled out if the context would be somehow captured (subject-verb agreement)
Why parsing is hard? Ambiguity

- Prepositional phrase attachment ambiguity
Put the block in the box on the table in the kitchen

- 3 prepositional phrases, 5 interpretations:
  - Put the block ((in the box on the table) in the kitchen)
  - Put the block (in the box (on the table in the kitchen))
  - Put ((the block in the box) on the table) in the kitchen.
  - Put (the block (in the box on the table)) in the kitchen.
  - Put (the block in the box) (on the table in the kitchen)
**Put the block in the box on the table in the kitchen**

- 3 prepositional phrases, 5 interpretations:
  - Put the block ((in the box on the table) in the kitchen)
  - Put the block (in the box (on the table in the kitchen))
  - ...

- A general case:
  - (((())))  (())()  ()(()()  (()())  ((())  (()())

\[
\text{Cat}_n = \binom{2n}{n} - \binom{2n}{n-1} \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}
\]

1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, ...
Canadian Utilities had 1988 revenue of $1.16 billion, mainly from its natural gas and electric utility businesses in Alberta, where the company serves about 800,000 customers.
Syntactic Ambiguities I

- **Prepositional phrases:**
  *They cooked the beans in the pot on the stove with handles.*

- **Particle vs. preposition:**
  *The puppy tore up the staircase.*

- **Complement structures**
  *The tourists objected to the guide that they couldn’t hear.*
  *She knows you like the back of her hand.*

- **Gerund vs. participial adjective**
  *Visiting relatives can be boring.*
  *Changing schedules frequently confused passengers.*
Syntactic Ambiguities II

- **Modifier scope within NPs**
  
  - *impractical design requirements*
  - *plastic cup holder*

- **Multiple gap constructions**
  
  - *The chicken is ready to eat.*
  - *The contractors are rich enough to sue.*

- **Coordination scope:**
  
  - *Small rats and mice can squeeze into holes or cracks in the wall.*
Dark Ambiguities

- **Dark ambiguities**: most analyses are shockingly bad (meaning, they don’t have an interpretation you can get your mind around)

This analysis corresponds to the correct parse of

“This is panic buying !”

- Unknown words and new usages
- **Solution**: We need mechanisms to focus attention on the best ones, probabilistic techniques do this
Put the block in the box on the table in the kitchen

- We want to score all the derivations to encode how plausible they are
PCFGs
Probabilistic Context-Free Grammars

- A context-free grammar is a tuple \(<N, T, S, R>\)
  - \(N\) : the set of non-terminals
    - Phrasal categories: S, NP, VP, ADJP, etc.
    - Parts-of-speech (pre-terminals): NN, JJ, DT, VB
  - \(T\) : the set of terminals (the words)
  - \(S\) : the start symbol
    - Often written as ROOT or TOP
    - Not usually the sentence non-terminal S
  - \(R\) : the set of rules
    - Of the form \(X \rightarrow Y_1 Y_2 \ldots Y_k\), with \(X, Y_i \in N\)
    - Examples: \(S \rightarrow NP VP\), \(VP \rightarrow VP CC VP\)
    - Also called rewrites, productions, or local trees

- A PCFG adds:
  - A top-down production probability per rule \(P(Y_1 Y_2 \ldots Y_k | X)\)
PCFGs

Associate probabilities with the rules:

\[ p(X \rightarrow \alpha) \]
\[ \forall X \rightarrow \alpha \in R : \quad 0 \leq p(X \rightarrow \alpha) \leq 1 \]
\[ \forall X \in N : \quad \sum_{\alpha : X \rightarrow \alpha \in R} p(X \rightarrow \alpha) = 1 \]

Now we can score a tree as a product of probabilities corresponding to the used rules:

<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td>1.0</td>
<td>(NP A girl) (VP ate a sandwich)</td>
</tr>
<tr>
<td>VP → V</td>
<td>0.2</td>
<td>(VP ate) (NP a sandwich)</td>
</tr>
<tr>
<td>VP → VP NF</td>
<td>0.4</td>
<td>(VP saw a girl) (PP with …)</td>
</tr>
<tr>
<td>NP → NP PF</td>
<td>0.3</td>
<td>(NP a girl) (PP with ….)</td>
</tr>
<tr>
<td>NP → D N</td>
<td>0.5</td>
<td>(D a) (N sandwich)</td>
</tr>
</tbody>
</table>
| NP → PN       | 0.2         | \n
\n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| N → girl      | 0.2         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| N → telescope | 0.7         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| N → sandwich  | 0.1         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| PN → I        | 1.0         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| V → saw       | 0.5         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| V → ate       | 0.5         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| P → with      | 0.6         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| P → in        | 0.4         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| P → in        | 0.4         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| D → a         | 0.3         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
</table>
| D → the       | 0.7         | \n
<table>
<thead>
<tr>
<th>Rule</th>
<th>Probability</th>
<th>Parse</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP → P NF</td>
<td>1.0</td>
<td>(P with) (NP with a sandwich)</td>
</tr>
</tbody>
</table>
## PCFGs

<table>
<thead>
<tr>
<th>Production</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \rightarrow NP\ VF$</td>
<td>1.0</td>
</tr>
<tr>
<td>$VP \rightarrow V$</td>
<td>0.2</td>
</tr>
<tr>
<td>$VP \rightarrow V\ NF$</td>
<td>0.4</td>
</tr>
<tr>
<td>$VP \rightarrow VP\ PF$</td>
<td>0.4</td>
</tr>
<tr>
<td>$NP \rightarrow NP\ PF$</td>
<td>0.3</td>
</tr>
<tr>
<td>$NP \rightarrow D\ N$</td>
<td>0.5</td>
</tr>
<tr>
<td>$NP \rightarrow PN$</td>
<td>0.2</td>
</tr>
<tr>
<td>$PP \rightarrow P\ NF$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

$N \rightarrow girl\ 0.2$

$N \rightarrow telescope\ 0.7$

$N \rightarrow sandwich\ 0.1$

$PN \rightarrow I\ 1.0$

$V \rightarrow saw\ 0.5$

$V \rightarrow ate\ 0.5$

$P \rightarrow with\ 0.6$

$P \rightarrow in\ 0.4$

$D \rightarrow a\ 0.3$

$D \rightarrow the\ 0.7$

$p(T) =$
PCFGs

\[ p(T) = 1.0 \times \]

\[ S \rightarrow NP \ VP \ 1.0 \]

\[ VP \rightarrow V \ 0.2 \]
\[ VP \rightarrow V \ NF \ 0.4 \]
\[ VP \rightarrow VP \ PF \ 0.4 \]

\[ NP \rightarrow NP \ PF \ 0.3 \]
\[ NP \rightarrow D \ N \ 0.5 \]
\[ NP \rightarrow PN \ 0.2 \]

\[ PP \rightarrow P \ NF \ 1.0 \]

\[ N \rightarrow girl \ 0.2 \]
\[ N \rightarrow telescope \ 0.7 \]
\[ N \rightarrow sandwich \ 0.1 \]
\[ PN \rightarrow I \ 1.0 \]
\[ V \rightarrow saw \ 0.5 \]
\[ V \rightarrow ate \ 0.5 \]
\[ P \rightarrow with \ 0.6 \]
\[ P \rightarrow in \ 0.4 \]
\[ D \rightarrow a \ 0.3 \]
\[ D \rightarrow the \ 0.7 \]
PCFGs

\[ S \rightarrow NP \ VP \ 1.0 \]
\[ VP \rightarrow V \ 0.2 \]
\[ VP \rightarrow V \ NF \ 0.4 \]
\[ VP \rightarrow VP \ PF \ 0.4 \]
\[ NP \rightarrow NP \ PF \ 0.3 \]
\[ NP \rightarrow D \ N \ 0.5 \]
\[ NP \rightarrow PN \ 0.2 \]
\[ PP \rightarrow P \ NF \ 1.0 \]

\[ N \rightarrow girl \ 0.2 \]
\[ N \rightarrow telescope \ 0.7 \]
\[ N \rightarrow sandwich \ 0.1 \]
\[ PN \rightarrow I \ 1.0 \]
\[ V \rightarrow saw \ 0.5 \]
\[ V \rightarrow ate \ 0.5 \]
\[ P \rightarrow with \ 0.6 \]
\[ P \rightarrow in \ 0.4 \]
\[ D \rightarrow a \ 0.3 \]
\[ D \rightarrow the \ 0.7 \]

\[ p(T) = 1.0 \times 0.2 \times \]
PCFGs

\[ S \rightarrow NP \ VP \ 1.0 \]

\[ VP \rightarrow V \ 0.2 \]
\[ NP \rightarrow NP \ PF \ 0.3 \]
\[ VP \rightarrow VP \ PF \ 0.4 \]
\[ NP \rightarrow D \ N \ 0.5 \]
\[ PN \rightarrow I \ 1.0 \]
\[ PP \rightarrow P \ NF \ 1.0 \]
\[ \]

\[ N \rightarrow girl \ 0.2 \]
\[ N \rightarrow telescope \ 0.7 \]
\[ N \rightarrow sandwich \ 0.1 \]
\[ V \rightarrow saw \ 0.5 \]
\[ V \rightarrow ate \ 0.5 \]
\[ P \rightarrow with \ 0.6 \]
\[ P \rightarrow in \ 0.4 \]
\[ D \rightarrow a \ 0.3 \]
\[ D \rightarrow the \ 0.7 \]

\[ p(T) = 1.0 \times 0.2 \times 1.0 \times \]
PCFGs

\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times \]

\[ S \rightarrow NP \; VP \; 1.0 \]
\[ V \rightarrow saw \; 0.5 \]
\[ V \rightarrow ate \; 0.5 \]
\[ P \rightarrow with \; 0.6 \]
\[ P \rightarrow in \; 0.4 \]
\[ D \rightarrow a \; 0.3 \]
\[ D \rightarrow the \; 0.7 \]
PCFGs

\[ S \rightarrow NP \ VP \ 1.0 \]
\[ NP \rightarrow WP \ PF \ 0.3 \]
\[ VP \rightarrow V \ NF \ 0.4 \]
\[ VP \rightarrow VP \ PF \ 0.4 \]
\[ PN \rightarrow I \ 1.0 \]
\[ V \rightarrow saw \ 0.5 \]
\[ V \rightarrow ate \ 0.5 \]
\[ N \rightarrow girl \ 0.2 \]
\[ N \rightarrow telescope \ 0.7 \]
\[ N \rightarrow sandwich \ 0.1 \]
\[ P \rightarrow with \ 0.6 \]
\[ P \rightarrow in \ 0.4 \]
\[ D \rightarrow a \ 0.3 \]
\[ D \rightarrow the \ 0.7 \]

\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times \]
PCFGs

\[
P(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times
\]

\[
S \rightarrow NP \ VP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
PCFGs

\[ S \rightarrow NP \ VP \ 1.0 \]
\[ VP \rightarrow V \ 0.2 \]
\[ VP \rightarrow V \ NP \ 0.4 \]
\[ VP \rightarrow VP \ PP \ 0.4 \]
\[ NP \rightarrow NP \ PP \ 0.3 \]
\[ NP \rightarrow D \ N \ 0.5 \]
\[ NP \rightarrow PN \ 0.2 \]
\[ PP \rightarrow P \ NP \ 1.0 \]
\[ N \rightarrow girl \ 0.2 \]
\[ N \rightarrow telescope \ 0.7 \]
\[ N \rightarrow sandwich \ 0.1 \]
\[ PN \rightarrow I \ 1.0 \]
\[ V \rightarrow saw \ 0.5 \]
\[ V \rightarrow ate \ 0.5 \]
\[ P \rightarrow with \ 0.6 \]
\[ P \rightarrow in \ 0.4 \]
\[ D \rightarrow a \ 0.3 \]
\[ D \rightarrow the \ 0.7 \]

\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times \\
0.5 \times 0.3 \times 0.2 \times 1.0 \times 0.6 \times 0.5 \times 0.3 \times 0.7 = 2.26 \times 10^{-5} \]
PCFG Estimation
ML estimation

- A treebank: a collection sentences annotated with constituent trees

- An estimated probability of a rule (maximum likelihood estimates)

\[ p(X \to \alpha) = \frac{C(X \to \alpha)}{C(X)} \]

- Smoothing is helpful
  - Especially important for preterminal rules
We defined a distribution over production rules for each nonterminal.

Our goal was to define a distribution over parse trees.

Unfortunately, not all PCFGs give rise to a proper distribution over trees, i.e. the sum over probabilities of all trees the grammar can generate may be less than 1: \( \sum_T P(T) < 1 \).

**Good news:** any PCFG estimated with the maximum likelihood procedure are always proper (Chi and Geman, 98).
CKY Parsing
Parsing

- Parsing is search through the space of all possible parses
  - e.g., we may want either any parse, all parses or the highest scoring parse (if PCFG):
    $$\arg \max_{T \in G(x)} P(T)$$

- Bottom-up:
  - One starts from words and attempt to construct the full tree

- Top-down
  - Start from the start symbol and attempt to expand to get the sentence
CKY algorithm (aka CYK)

- Cocke-Kasami-Younger algorithm
  - Independently discovered in late 60s / early 70s

- An efficient bottom up parsing algorithm for (P)CFGs
  - can be used both for the recognition and parsing problems
  - Very important in NLP (and beyond)

- We will start with the non-probabilistic version
Constraints on the grammar

- The basic CKY algorithm supports only rules in the Chomsky Normal Form (CNF):

\[
C \rightarrow x
\]

\[
C \rightarrow C_1 C_2
\]

Unary preterminal rules (generation of words given PoS tags)

\[
N \rightarrow \text{telescope} \quad D \rightarrow \text{the}
\]

Binary inner rules

\[
S \rightarrow NP VP \quad NP \rightarrow D \quad N
\]
Constraints on the grammar

- The basic CKY algorithm supports only rules in the Chomsky Normal Form (CNF):
  
  \[ C \rightarrow x \]
  \[ C \rightarrow C_1 C_2 \]

- Any CFG can be converted to an equivalent CNF
  - Equivalent means that they define the same language
  - However (syntactic) trees will look differently
  - It is possible to address it by defining such transformations that allows for easy reverse transformation
Transformation to CNF form

- What one need to do to convert to CNF form
  - Get rid of rules that mix terminals and non-terminals
  - Get rid of unary rules: $C \rightarrow C_1$
  - Get rid of N-ary rules: $C \rightarrow C_1 C_2 \ldots C_n \ (n > 2)$

Crucial to process them, as required for efficient parsing
Transformation to CNF form: binarization

- Consider

\[ NP \rightarrow DT \ NNP \ VBG \ NN \]

\[
\text{NP} \\
\text{DT} \quad \text{NNP} \quad \text{VBG} \quad \text{NN}
\]

- How do we get a set of binary rules which are equivalent?
Transformation to CNF form: binarization

- Consider

\[ NP \rightarrow DT\ NNP\ VBG\ NN \]

\[
\begin{array}{c}
\text{NP} \\
\text{DT} \\
\text{the} \\
\text{NNP} \\
\text{Dutch} \\
\text{VBG} \\
\text{publishing} \\
\text{NN} \\
\text{group}
\end{array}
\]

- How do we get a set of binary rules which are equivalent?

\[
\begin{align*}
NP & \rightarrow DT\ X \\
X & \rightarrow NNP\ Y \\
Y & \rightarrow VBG\ NN
\end{align*}
\]
Transformation to CNF form: binarization

- **Consider**

  \[ NP \rightarrow DT \ NNP \ VBG \ NN \]

  ![Diagram](image)

  - **How do we get a set of binary rules which are equivalent?**
    
    \[ NP \rightarrow DT \ X \]
    
    \[ X \rightarrow NNP \ Y \]
    
    \[ Y \rightarrow VBG \ NN \]

- **A more systematic way to refer to new non-terminals**

  \[ NP \rightarrow DT \ @NP|DT \]
  
  \[ @NP|DT \rightarrow NNP \ @NP|DT\_NNP \]
  
  \[ @NP|DT\_NNP \rightarrow VBG \ NN \]
Instead of binarizing tuples we can binarize trees on preprocessing:

- Also known as lossless Markovization in the context of PCFGs
- Can be easily reversed on postprocessing
CKY: Parsing task

- We are given
  - a grammar $<N, T, S, R>$
  - a sequence of words $w = (w_1, w_2, \ldots, w_n)$

- Our goal is to produce a parse tree for $w$
CKY: Parsing task

- We are given
  - a grammar $\langle N, T, S, R \rangle$
  - a sequence of words $w = (w_1, w_2, \ldots, w_n)$
- Our goal is to produce a parse tree for $w$
- We need an easy way to refer to substrings of $w$

$\text{span } (i, j)$ refers to words between fenceposts $i$ and $j$
Parsing one word

\[ C \rightarrow w_i \]
Parsing one word

\[ C \rightarrow w_i \]
 Parsing one word

$C \rightarrow w_i$

covers all words between $i-1$ and $i$
Parsing longer spans

$C \rightarrow C_1 \ C_2$

Check through all $C_1$, $C_2$, $mid$

covers all words btw $min$ and $mid$
covers all words btw $mid$ and $max$
Parsing longer spans

$C \rightarrow C_1 \ C_2$

Check through all C1, C2, mid

covers all words btw min and mid

covers all words btw mid and max
Parsing longer spans

covers all words
between \textit{min} and \textit{max}
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**Terminal rules**

**Chart (aka parsing triangle)**

- $S \rightarrow NP \ VP$
- $VP \rightarrow M \ V$
- $VP \rightarrow V$
- $NP \rightarrow N$
- $NP \rightarrow N \ NP$
- $N \rightarrow can$
- $N \rightarrow lead$
- $N \rightarrow poison$
- $M \rightarrow can$
- $M \rightarrow must$
- $V \rightarrow poison$
- $V \rightarrow lead$
Preterminal rules

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Inner rules

$S \rightarrow NP\ VP$

$VP \rightarrow M\ V$

$NP \rightarrow N$

$NP \rightarrow N\ NP$

$N \rightarrow can$

$N \rightarrow lead$

$N \rightarrow poison$

$M \rightarrow can$

$M \rightarrow must$

$V \rightarrow poison$

$V \rightarrow lead$
Preterminal rules

Inner rules

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

\[ S \rightarrow NP \ VP \]

\[ VP \rightarrow M \ V \]

\[ NP \rightarrow N \]

\[ NP \rightarrow N \ NP \]

\[ N \rightarrow can \]

\[ N \rightarrow lead \]

\[ N \rightarrow poison \]

\[ M \rightarrow can \]

\[ M \rightarrow must \]

\[ V \rightarrow poison \]

\[ V \rightarrow lead \]
Preterminal rules

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Inner rules

\[
S \rightarrow NP \ VP \\
VP \rightarrow M \ V \\
NP \rightarrow N \\
NP \rightarrow N \ NP \\
N \rightarrow \text{can} \\
N \rightarrow \text{lead} \\
N \rightarrow \text{poison} \\
M \rightarrow \text{can} \\
M \rightarrow \text{must} \\
V \rightarrow \text{poison} \\
V \rightarrow \text{lead}
\]
\[
S \rightarrow NP \ VP
\]
\[
NP \rightarrow N
\]
\[
NP \rightarrow N \ NP
\]
\[
VP \rightarrow M \ V
\]
\[
VP \rightarrow V
\]
\[
N \rightarrow can
\]
\[
N \rightarrow lead
\]
\[
N \rightarrow poison
\]
\[
M \rightarrow can
\]
\[
M \rightarrow must
\]
\[
V \rightarrow poison
\]
\[
V \rightarrow lead
\]
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ S \rightarrow NP \ VP \]

\[ VP \rightarrow M \ V \]
\[ VP \rightarrow V \]

\[ NP \rightarrow N \]
\[ NP \rightarrow N \ NP \]

\[ N \rightarrow \text{can} \]
\[ N \rightarrow \text{lead} \]
\[ N \rightarrow \text{poison} \]

\[ M \rightarrow \text{can} \]
\[ M \rightarrow \text{must} \]

\[ V \rightarrow \text{poison} \]
\[ V \rightarrow \text{lead} \]
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Preterminal rules

\[
S \rightarrow NP \ VP
\]

Inner rules

\[
VP \rightarrow M \ V
\]

\[
VP \rightarrow V
\]

\[
NP \rightarrow N
\]

\[
NP \rightarrow N \ NP
\]

\[
N \rightarrow can
\]

\[
N \rightarrow lead
\]

\[
N \rightarrow poison
\]

\[
M \rightarrow can
\]

\[
M \rightarrow must
\]

\[
V \rightarrow poison
\]

\[
V \rightarrow lead
\]
```
S → NP VP

VP → M V
VP → V

NP → N
NP → N NP

N → can
N → lead
N → poison

M → can
M → must

V → poison
V → lead
```

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

```
max = 1
max = 2
max = 3

min = 0

min = 1

min = 2
```
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\[ \begin{array}{c|c|c|c}
\text{max} & \text{max} = 1 & \text{max} = 2 & \text{max} = 3 \\
\text{min} & \begin{array}{c}
\text{min} = 0 \\
\text{min} = 1 \\
\text{min} = 2
\end{array} \\
\end{array} \]

Preterminal rules:

\[
S \rightarrow NP \ VP
\]

\[
VP \rightarrow M \ V
\]

\[
VP \rightarrow V
\]

\[
NP \rightarrow N
\]

\[
NP \rightarrow N \ NP
\]

Inner rules:

\[
N \rightarrow \text{can}
\]

\[
N \rightarrow \text{lead}
\]

\[
N \rightarrow \text{poison}
\]

\[
M \rightarrow \text{can}
\]

\[
M \rightarrow \text{must}
\]

\[
V \rightarrow \text{poison}
\]

\[
V \rightarrow \text{lead}
\]
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccc}
\text{max} = 1 & \text{max} = 2 & \text{max} = 3 \\
\text{min} = 0 & \text{min} = 1 & \text{min} = 2 \\
1 \quad N, V & 2 \quad N, M & 3 \quad N, V \\
\end{array}
\]

- Pretermial rules:
  - \( S \rightarrow NP \ VP \)
  - \( VP \rightarrow M \ V \)
  - \( VP \rightarrow V \)
  - \( NP \rightarrow N \)
  - \( NP \rightarrow N \ NP \)

- Inner rules:
  - \( N \rightarrow can \)
  - \( N \rightarrow lead \)
  - \( N \rightarrow poison \)
  - \( M \rightarrow can \)
  - \( M \rightarrow must \)
  - \( V \rightarrow poison \)
  - \( V \rightarrow lead \)
Preterminal rules

\[ S \rightarrow NP \ VP \]

\[ VP \rightarrow M \ V \]
\[ VP \rightarrow V \]
\[ NP \rightarrow N \]
\[ NP \rightarrow N \ NP \]

Inner rules

\[ N \rightarrow can \]
\[ N \rightarrow lead \]
\[ N \rightarrow poison \]
\[ M \rightarrow can \]
\[ M \rightarrow must \]
\[ V \rightarrow poison \]
\[ V \rightarrow lead \]
\[
S \rightarrow NP \ VP \\
VP \rightarrow M\ V \\
VP \rightarrow V \\
NP \rightarrow N \\
NP \rightarrow N\ NP \\
N \rightarrow can \\
N \rightarrow lead \\
N \rightarrow poison \\
M \rightarrow can \\
M \rightarrow must \\
V \rightarrow poison \\
V \rightarrow lead
\]
Preterminal rules

Inner rules

\[
S \to NP \ VP
\]

\[
VP \to M \ V
\]

\[
VP \to V
\]

\[
NP \to N
\]

\[
NP \to N \ NP
\]

\[
N \to can
\]

\[
N \to lead
\]

\[
N \to poison
\]

\[
M \to can
\]

\[
M \to must
\]

\[
V \to poison
\]

\[
V \to lead
\]
Preterminal rules

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

max = 1

max = 2

max = 3

min = 0

1. \( N, V \)
   \( NP, VP \)

2. \( N, M \)
   \( NP \)

3. \( N, V \)
   \( NP, VP \)

4. \( NP \)

5. ?

min = 1

min = 2

Inner rules

\[ S \rightarrow NP \ VP \]

\[ VP \rightarrow M \ V \]

\[ VP \rightarrow V \]

\[ NP \rightarrow N \]

\[ NP \rightarrow N \ NP \]

\[ N \rightarrow can \]

\[ N \rightarrow lead \]

\[ N \rightarrow poison \]

\[ M \rightarrow can \]

\[ M \rightarrow must \]

\[ V \rightarrow poison \]

\[ V \rightarrow lead \]
Preterminal rules

Inner rules

\[
S \rightarrow NP \ VP \\
VP \rightarrow M \ V \\
\]

\[
NP \rightarrow N \\
NP \rightarrow N \ NP \\
\]

\[
N \rightarrow can \\
N \rightarrow lead \\
N \rightarrow poison \\
M \rightarrow can \\
M \rightarrow must \\
V \rightarrow poison \\
V \rightarrow lead \\
\]
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

max = 1 max = 2 max = 3

min = 0

1: $N, V$
   $NP, VP$

2: $N, M$
   $NP$

3: $N, V$
   $NP, VP$

4: $NP$

5: $S, VP, NP$

6: ?

min = 1

min = 2

$S \rightarrow NP \ VP$

$VP \rightarrow M \ V$

$VP \rightarrow V$

$NP \rightarrow N$

$NP \rightarrow N \ NP$

$N \rightarrow can$

$N \rightarrow lead$

$N \rightarrow poison$

$M \rightarrow can$

$M \rightarrow must$

$V \rightarrow poison$

$V \rightarrow lead$
<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

max = 1  max = 2  max = 3

\[ S \rightarrow NP \ VP \]
\[ VP \rightarrow M \ V \]
\[ VP \rightarrow V \]
\[ NP \rightarrow N \]
\[ NP \rightarrow N \ NP \]
\[ N \rightarrow can \]
\[ N \rightarrow lead \]
\[ N \rightarrow poison \]
\[ M \rightarrow can \]
\[ M \rightarrow must \]
\[ V \rightarrow poison \]
\[ V \rightarrow lead \]
Preterminal rules

<table>
<thead>
<tr>
<th>lead</th>
<th>can</th>
<th>poison</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Inner rules

max = 1  max = 2  max = 3

min = 0

mid = 1

min = 1

min = 2

$S \rightarrow NP \ VP$

$VP \rightarrow M \ V$

$VP \rightarrow V$

$NP \rightarrow N$

$NP \rightarrow N \ NP$

$N \rightarrow can$

$N \rightarrow lead$

$N \rightarrow poison$

$M \rightarrow can$

$M \rightarrow must$

$V \rightarrow poison$

$V \rightarrow lead$
$S \rightarrow NP \ VP$

$VP \rightarrow M \ V$
$VP \rightarrow V$

$NP \rightarrow N$
$NP \rightarrow N \ NP$

$N \rightarrow can$
$N \rightarrow lead$
$N \rightarrow poison$

$M \rightarrow can$
$M \rightarrow must$

$V \rightarrow poison$
$V \rightarrow lead$
Apparently the sentence is ambiguous for the grammar: (as the grammar overgenerates)
No subject-verb agreement, and *poison* used as an intransitive verb.
Chart can be represented by a Boolean 3D array \( \text{chart}[\text{min}][\text{max}][\text{C}] \):

- Relevant entries have \( 0 < \text{min} < \text{max} \leq n \)

\[
\text{chart}[\text{min}][\text{max}][\text{C}] = \begin{cases} 
\text{true} & \text{if the signature (min, max, C) is already added to the chart;} \\
\text{false} & \text{otherwise.}
\end{cases}
\]

Here we assume that labels (C) are integer indices.
Implementation: binary rules

for each max from 2 to n

    for each min from max - 2 down to 0

        for each syntactic category C

            for each binary rule C -> C₁ C₂

                for each mid from min + 1 to max - 1

                    if chart[min][mid][C₁] and chart[mid][max][C₂] then

                        chart[min][max][C] = true
for each $w_i$ from left to right

for each preterminal rule $C \rightarrow w_i$

$\text{chart}[i - 1][i][C] = \text{true}$
Algorithm analysis

Time complexity?

for each max from 2 to n

   for each min from max - 2 down to 0

      for each syntactic category C

         for each binary rule C -> C₁ C₂

            for each mid from min + 1 to max - 1
Algorithm analysis

Time complexity?

\[
\text{for each max from 2 to n} \\
\text{for each min from max - 2 down to 0} \\
\text{for each syntactic category C} \\
\text{for each binary rule } C \rightarrow C_1 C_2 \\
\text{for each mid from min + 1 to max - 1}
\]

\[O(n^3|R|)\] where \(|R|\) is the number of rules in the grammar
Practical time complexity

\[ \sim n^{3.6} \]
Probabilistic CKY
PCFGs

\[
p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times 0.5 \times 0.3 \times 0.2 \times 1.0 \times 0.6 \times 0.5 \times 0.3 \times 0.7
\]
\[
= 2.26 \times 10^{-5}
\]
CKY with PCFGs

- Chart is represented by a 3d array of floats
  \[
  \text{chart}[\text{min}][\text{max}][\text{label}]
  \]
  - It stores probabilities for the most probable subtree with a given signature

- \[
  \text{chart}[0][n][S] \text{ will store the probability of the most probable full parse tree}
  \]
Intuition

For every $C$ choose $C_1$, $C_2$ and mid such that

$$P(T_1) \times P(T_2) \times P(C \rightarrow C_1C_2)$$

is maximal, where $T_1$ and $T_2$ are left and right subtrees.
for each $w_i$ from left to right

for each preterminal rule $C \rightarrow w_i$

$\text{chart}[i-1][i][C] = p(C \rightarrow w_i)$
for each max from 2 to n

    for each min from max - 2 down to 0

        for each syntactic category C

            double best = undefined

            for each binary rule C -> C₁ C₂

                for each mid from min + 1 to max - 1

                    double t₁ = chart[min][mid][C₁]

                    double t₂ = chart[mid][max][C₂]

                    double candidate = t₁ * t₂ * p(C -> C₁ C₂)

                    if candidate > best then
                        best = candidate
                    chart[min][max][C] = best