Part B: Syntactic Parsing

Structure of Part B (3 lectures):
- Overview of natural language syntax; basic formalisms (CFG / PCFG)
- Dynamic programming algorithms for syntactic parsing (CKY)
- Shortcomings of the basic PCFG model, enhancements

Project: Implement a Parser
- Implement a statistical parser of English: estimate a statistical model from a dataset and be able to process new sentences
- Most of the work (including the “boring” part – input / output, preprocessing) is already done in the code base
- You will modify it to produce a fully functional parser of English
  - But as a result you will have an end-to-end parser of your own!

[Largely based on the code kindly provided by Dan Klein and Slav Petrov]
Today

- **Syntactic representations**
  - Formalisms
  - Parsing
  - Ambiguity (and other challenges)

- **CFGs**
  - Definition and examples
  - Weaknesses

- **PCFGs**
  - Definition and examples
  - Estimating PCFG from a dataset (treebank)
Syntactic parsing

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

- **Parsing**
  - The process of predicting syntactic representations

- **Syntactic Representations**
  - Different types of syntactic representations are possible, for example:

  **Constuent (a.k.a. phrase-structure) tree**

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

  **Dependency tree**

  ```
  root
  / \  
 poss nsubj 
 / \ / \ 
 My dog ate a sausage
  ```

Afyonkarahisarlaştırabilirdiklerimizdenmişsinizcesineee in Turkish mean "as if you are one of the people that we thought to be originating from Afyonkarahisar" [wikipedia]
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
Bracketing notation

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) )
  )
)

We will use this format in the assignment
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
- nsubj - subject
- det - determiner
Constituent and dependency representations

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

  Roughly: every word along with all its dependents corresponds to a phrase = to an inner node in the constituent tree

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

  Recovering labels is harder

One potential rule to extract nsubj dependency

Roughly: every word along with all its dependents corresponds to a phrase = to an inner node in the constituent tree
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*  
  Syntactic subject corresponds to an agent (Who is baking?)
- *A cake is baking in the oven*  
  Syntactic subject corresponds to a patient (What is being baked?)

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, …
Brief history of parsing

- Before mid-90s syntactic rule-based parsers producing rich linguistic information
  - Provide low coverage (though hard to evaluate)
  - Predict much more information than the formalisms we have just discussed
- Realization that basic PCFGs do not result in accurate predictive models
  - We will discuss this in detail
- Mid-90s first accurate statistical parsers (e.g., [Magerman 1994, Collins [1996]
  - No handcrafted grammars, estimated from large datasets (Penn Treebank WSJ)
- Now: better models, more efficient algorithms, more languages, …
Ambiguity

- Ambiguous sentence:
  - I saw a girl with a telescope
- What kind of ambiguity it has?
- What implications it has for parsing?
Why parsing is hard? Ambiguity

- Prepositional phrase attachment ambiguity

What kind of clues would be useful?

PP-attachment ambiguity is just one example type of ambiguity

How serious is this problem in practice?
Why parsing is hard? Ambiguity

- **Example with 3 preposition 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)

- **A general case:**

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

  \[
  1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, \ldots
  \]

This is only one type of ambiguity but you have many types in almost all sentences.
Why parsing is hard?  Ambiguity

- A typical tree from a standard dataset (Penn treebank WSJ)

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.
Key problems

- **Recognition problem**: is the sentence grammatical?
- **Parsing problem**: what is a (most plausible) derivation (tree) corresponding the sentence?

Parsing problem encompasses the recognition problem

A more interesting question from practical point of view
Today

- **Syntactic representations**
  - Formalisms
  - Parsing
  - Ambiguity (and other challenges)

- **CFGs**
  - Definition and examples
  - Weaknesses

- **PCFGs**
  - Definition and examples
  - Estimating PCFG from a dataset (treebank)
Context-Free Grammar

- Context-free grammar is a tuple of 4 elements \( G = (V, \Sigma, R, S) \)

- \( V \) - the set of non-terminals

- \( \Sigma \) - the set of terminals

- \( R \) - the set of rules of the form \( X \rightarrow Y_1, Y_2, \ldots, Y_n \), where \( n \geq 0 \), \( X \in V, Y_i \in V \cup \Sigma \)

- \( S \) is a dedicated start symbol

In our case: phrase categories (VP, NP, ..) and PoS tags (N, V, .. – aka preterminals)

Words

\[
\begin{align*}
S & \rightarrow NP \ VP \\
NP & \rightarrow D \ N \\
NP & \rightarrow PN \\
NP & \rightarrow NP \ PP \\
PP & \rightarrow P \ NP \\
N & \rightarrow girl \\
N & \rightarrow telescope \\
V & \rightarrow saw \\
V & \rightarrow eat \\
\ldots
\end{align*}
\]
An example grammar

\[ V = \{S, VP, NP, PP, N, V, PN, P\} \]

\[ \Sigma = \{girl, telescope, sandwich, I, saw, ate, with, in, a, the\} \]

\[ S = \{S\} \]

\[ R : \]

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

\[ VP \rightarrow V \]

\[ VP \rightarrow V \ NP \] (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)

\[ NP \rightarrow PN \]

\[ PP \rightarrow P \ NP \] (P with) (NP with a sandwich)

Preterminal rules (correspond to the HMM emission model aka task model)

\[ N \rightarrow girl \]

\[ N \rightarrow telescope \]

\[ N \rightarrow sandwich \]

\[ PN \rightarrow I \]

\[ V \rightarrow saw \]

\[ V \rightarrow ate \]

\[ P \rightarrow with \]

\[ P \rightarrow in \]

\[ D \rightarrow a \]

\[ D \rightarrow the \]
CFGs

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

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

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

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

\[
\begin{align*}
S & \rightarrow NP \ VP \\
VP & \rightarrow V \\
VP & \rightarrow V \ NP \\
VP & \rightarrow VP \ PP \\
NP & \rightarrow NP \ PP \\
NP & \rightarrow D \ N \\
NP & \rightarrow PN \\
PP & \rightarrow P \ NP
\end{align*}
\]

\[
\begin{align*}
N & \rightarrow girl \\
N & \rightarrow telescope \\
N & \rightarrow sandwich \\
PN & \rightarrow I \\
V & \rightarrow saw \\
V & \rightarrow ate \\
P & \rightarrow with \\
P & \rightarrow in \\
D & \rightarrow a \\
D & \rightarrow the
\end{align*}
\]
CFGs

\[ S \rightarrow NP \ VP \]
\[ NP \rightarrow NP \ PP \]
\[ PP \rightarrow P \ NP \]
\[ VP \rightarrow V \]
\[ VP \rightarrow VP \ PP \]
\[ PN \rightarrow I \]
\[ V \rightarrow saw \]
\[ V \rightarrow ate \]
\[ P \rightarrow with \]
\[ P \rightarrow in \]
\[ D \rightarrow a \]
\[ D \rightarrow the \]

\[ N \rightarrow girl \]
\[ N \rightarrow telescope \]
\[ N \rightarrow sandwich \]
CFGs

$S \rightarrow NP \ VP$

$NP \rightarrow NP \ PP$

$VP \rightarrow V NP$

$NP \rightarrow D N$

$NP \rightarrow PN$

$PP \rightarrow P NP$

$VP \rightarrow VP \ PP$

$V \rightarrow saw$

$V \rightarrow ate$

$P \rightarrow with$

$P \rightarrow in$

$D \rightarrow a$

$D \rightarrow the$

$N \rightarrow girl$

$N \rightarrow telescope$

$N \rightarrow sandwich$

$PN \rightarrow I$
CFGs

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

\[
\begin{align*}
S & \rightarrow NP \ VP \\
VP & \rightarrow V \ NP \\
VP & \rightarrow VP \ PP \\
NP & \rightarrow NP \ PP \\
NP & \rightarrow D \ N \\
PN & \rightarrow I \\
V & \rightarrow saw \\
P & \rightarrow with \\
D & \rightarrow a \\
D & \rightarrow the
\end{align*}
\]
CFGs

\[
S \rightarrow NP \ VP \\
VP \rightarrow V \\
VP \rightarrow V \ NP \\
VP \rightarrow VP \ PP \\
NP \rightarrow NP \ PP \\
NP \rightarrow D \ N \\
NP \rightarrow PN \\
PP \rightarrow P \ NP \\
PP \rightarrow P \ PP \\
D \rightarrow a \\
D \rightarrow the
\]

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

CFG defines both:
- a set of substrings (a language)
- structures used to represent sentences (constituent trees)

```
S → NP VP
N → girl
telescope
sandwich
VP → V NP
VP → VP PP
PN → I
V → saw
ate
with
PP → P NP
in
a
the
```

```
NP → NP PP
NP → D N
NP → PN
PP → P NP
```

```
NP | PN
I | saw
V | NP
a | girl
with | D | N
a | telescope
```
Why context-free?

What can be a sub-tree is only affected by what the phrase type is (VP) but not the context.
Why context-free?

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

Not grammatical
Why context-free?

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

Not grammatical

Matters if we want to generate language (e.g., language modeling) but is this relevant to parsing?
Introduced 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)
Introduced coordination ambiguity

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

```
S
  NP
    NNS koalas
  VP
    VBP eat
    NP
      NNS leaves
        CC and
      NP
        NNS barks
  VP
    VBZ barks
```

- This tree would be ruled out if the context would be somehow captured (subject-verb agreement).

Even more detailed PoS tags are not going to help here.

"Bark" can refer both to a noun or a verb.
Key problems

- **Recognition problem:** does the sentence belong to the language defined by CFG?
  - That is, there a derivation which yields the sentence?

- **Parsing problem:** what is a (most plausible) derivation (tree) corresponding the sentence?

- Parsing problem encompasses the recognition problem
How to deal with ambiguity?

- There are (exponentially) many derivation for a typical sentence

```
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
Today

- **Syntactic representations**
  - Formalisms
  - Parsing
  - Ambiguity (and other challenges)

- **CFGs**
  - Definition and examples
  - Weaknesses

- **PCFGs**
  - Definition and examples
  - Estimating PCFG from a dataset (treebank)
An example **probabilistic CFG**

Associate probabilities with the rules \( p(X \rightarrow \alpha) \):

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

\[
S \rightarrow NP \ VP \quad 1.0 \quad \text{(NP A girl) (VP ate a sandwich)}
\]

\[
VP \rightarrow V \quad 0.2
\]

\[
VP \rightarrow V \ NP \quad 0.4 \quad \text{(VP ate) (NP a sandwich)}
\]

\[
VP \rightarrow VP \ PP \quad 0.4 \quad \text{(VP saw a girl) (PP with …)}
\]

\[
NP \rightarrow NP \ PP \quad 0.3 \quad \text{(NP a girl) (PP with ….)}
\]

\[
NP \rightarrow D \ N \quad 0.5 \quad \text{(D a) (N sandwich)}
\]

\[
NP \rightarrow PN \quad 0.2
\]

\[
PP \rightarrow P \ NP \quad 1.0 \quad \text{(P with) (NP with a sandwich)}
\]

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

\[
N \rightarrow girl \quad 0.2
\]

\[
N \rightarrow telescope \quad 0.7
\]

\[
N \rightarrow sandwich \quad 0.1
\]

\[
PN \rightarrow I \quad 1.0
\]

\[
V \rightarrow saw \quad 0.5
\]

\[
V \rightarrow ate \quad 0.5
\]

\[
P \rightarrow with \quad 0.6
\]

\[
P \rightarrow in \quad 0.4
\]

\[
D \rightarrow a \quad 0.3
\]

\[
D \rightarrow the \quad 0.7
\]
CFGs

\[ p(T) = \]

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

\[
VP \rightarrow V \ 0.2
\]

\[
VP \rightarrow V \ NP \ 0.4
\]

\[
VP \rightarrow VP \ PP \ 0.4
\]

\[
NP \rightarrow NP \ PP \ 0.3
\]

\[
NP \rightarrow D \ N \ 0.5
\]

\[
NP \rightarrow PN \ 0.2
\]

\[
PP \rightarrow P \ NP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
N \rightarrow sandwich \ 0.1
\]

\[
PN \rightarrow I \ 1.0
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
CFGs

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

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

\[
VP \rightarrow V \ 0.2
\]

\[
VP \rightarrow V \ NP \ 0.4
\]

\[
VP \rightarrow VP \ PP \ 0.4
\]

\[
NP \rightarrow NP \ PP \ 0.3
\]

\[
NP \rightarrow D \ N \ 0.5
\]

\[
NP \rightarrow PN \ 0.2
\]

\[
PP \rightarrow P \ NP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
N \rightarrow sandwich \ 0.1
\]

\[
PN \rightarrow I \ 1.0
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
\[ p(T) = 1.0 \times 0.2 \times \]

CFGs

\[
\begin{align*}
S & \rightarrow NP \ VP \ 1.0 \\
NP & \rightarrow PP \ 0.3 \\
NP & \rightarrow PN \ 0.2 \\
VP & \rightarrow V \ NP \ 0.4 \\
VP & \rightarrow VP \ PP \ 0.4 \\
V & \rightarrow saw \ 0.5 \\
P & \rightarrow with \ 0.6 \\
D & \rightarrow a \ 0.3 \\
D & \rightarrow the \ 0.7 \\
N & \rightarrow girl \ 0.2 \\
N & \rightarrow telescope \ 0.7 \\
N & \rightarrow sandwich \ 0.1 \\
PN & \rightarrow I \ 1.0 \\
P & \rightarrow in \ 0.4
\end{align*}
\]
$p(T) = 1.0 \times 0.2 \times 1.0 \times$

CFGs

$S \rightarrow NP \ VP \ 1.0$
$NP \rightarrow NP \ PP \ 0.3$
$NP \rightarrow D \ N \ 0.5$
$NP \rightarrow PN \ 0.2$
$VP \rightarrow VP \ PP \ 0.4$
$VP \rightarrow V \ NP \ 0.4$
$VP \rightarrow V \ 0.2$

$N \rightarrow girl \ 0.2$
$N \rightarrow telescope \ 0.7$
$N \rightarrow sandwich \ 0.1$

$PN \rightarrow I \ 1.0$
$V \rightarrow saw \ 0.5$
$V \rightarrow ate \ 0.5$
$P \rightarrow with \ 0.6$
$P \rightarrow in \ 0.4$
$D \rightarrow a \ 0.3$
$D \rightarrow the \ 0.7$
p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times

CFGs

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

\[
VP \rightarrow V \ 0.2
\]

\[
VP \rightarrow V \ NP \ 0.4
\]

\[
VP \rightarrow VP \ PP \ 0.4
\]

\[
NP \rightarrow NP \ PP \ 0.3
\]

\[
NP \rightarrow D \ N \ 0.5
\]

\[
NP \rightarrow PN \ 0.2
\]

\[
PP \rightarrow P \ NP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
N \rightarrow sandwich \ 0.1
\]

\[
PN \rightarrow I \ 1.0
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times \]

CFGs

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

CFGs

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

\[
VP \rightarrow V \ 0.2
\]

\[
VP \rightarrow V \ NP \ 0.4
\]

\[
VP \rightarrow VP \ PP \ 0.4
\]

\[
NP \rightarrow NP \ PP \ 0.3
\]

\[
NP \rightarrow D \ N \ 0.5
\]

\[
NP \rightarrow PN \ 0.2
\]

\[
PP \rightarrow P \ NP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
N \rightarrow sandwich \ 0.1
\]

\[
PN \rightarrow I \ 1.0
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
$p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times 0.5 \times$
\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times 0.5 \times 0.3 \times \]

CFGs

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

\[
VP \rightarrow V \ 0.2
\]

\[
VP \rightarrow V \ NP \ 0.4
\]

\[
VP \rightarrow VP \ PP \ 0.4
\]

\[
NP \rightarrow NP \ PP \ 0.3
\]

\[
NP \rightarrow D \ N \ 0.5
\]

\[
NP \rightarrow PN \ 0.2
\]

\[
PP \rightarrow P \ NP \ 1.0
\]

\[
N \rightarrow girl \ 0.2
\]

\[
N \rightarrow telescope \ 0.7
\]

\[
N \rightarrow sandwich \ 0.1
\]

\[
PN \rightarrow I \ 1.0
\]

\[
V \rightarrow saw \ 0.5
\]

\[
V \rightarrow ate \ 0.5
\]

\[
P \rightarrow with \ 0.6
\]

\[
P \rightarrow in \ 0.4
\]

\[
D \rightarrow a \ 0.3
\]

\[
D \rightarrow the \ 0.7
\]
\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times \]
\[ 0.5 \times 0.3 \times 0.2 \]
\[ p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times 0.5 \times 0.3 \times 0.2 \times 1.0 \times 0.6 \times 0.5 \times 0.3 \times 0.7 \]
\[ = 2.26 \times 10^{-5} \]
Distribution over trees

- 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]
Distribution over trees

- Let us denote by $G(x)$ the set of derivations for the sentence $x$
- The probability distribution defines the scoring $P(T)$ over the trees $T \in G(x)$
- Finding the best parse for the sentence according to PCFG:

$$\arg \max_{T \in G(x)} P(T)$$

We will look at this at the next class
ML estimation

- A treebank: a collection sentences annotated with constituent trees

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

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

- Smoothing is helpful
  - Especially important for preterminal rules, i.e. generation of words ( = task mode in PoS tagging)
  - The same smoothing techniques as studied before can be used (e.g., add 1 smoothing)
ML estimation: an example

- **A toy treebank:**

  ![Treebank Example](image)

- **Without smoothing:**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Count</th>
<th>Prob. estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \rightarrow B\ C$</td>
<td>$n_1$</td>
<td>$n_1/(n_1+n_2+n_3)$</td>
</tr>
<tr>
<td>$S \rightarrow C$</td>
<td>$n_2$</td>
<td>$n_2/(n_1+n_2+n_3)$</td>
</tr>
<tr>
<td>$S \rightarrow B$</td>
<td>$n_3$</td>
<td>$n_3/(n_1+n_2+n_3)$</td>
</tr>
<tr>
<td>$B \rightarrow a\ a$</td>
<td>$n_1$</td>
<td>$n_1/(n_1+n_3)$</td>
</tr>
<tr>
<td>$B \rightarrow a$</td>
<td>$n_3$</td>
<td>$n_3/(n_1+n_3)$</td>
</tr>
<tr>
<td>$C \rightarrow a\ a$</td>
<td>$n_1$</td>
<td>$n_1/(n_1+n_2)$</td>
</tr>
<tr>
<td>$C \rightarrow a\ a\ a$</td>
<td>$n_2$</td>
<td>$n_2/(n_1+n_2)$</td>
</tr>
</tbody>
</table>
Penn Treebank: peculiarities

- Wall street journal: around 40,000 annotated sentences, 1,000,000 words
- Fine-grain part of speech tags (45), e.g., for verbs
  - VBD: Verb, past tense
  - VBG: Verb, gerund or present participle
  - VBP: Verb, present (non-3rd person singular)
  - VBZ: Verb, present (3rd person singular)
  - MD: Modal
- Flat NPs (no attempt to disambiguate NP attachment)
  
  ![Diagram of a flat NPs structure]

  - DT: a
  - JJ: hot
  - NN: dog
  - NN: food
  - NN: cart
Today we considered syntactic formalisms and defined (P)CFGs.

you are ready to start working on the assignment step 1.

Next time

We will learn how to parse (efficiently) with (P)CFGs.