Statistical Parsing: Part 1

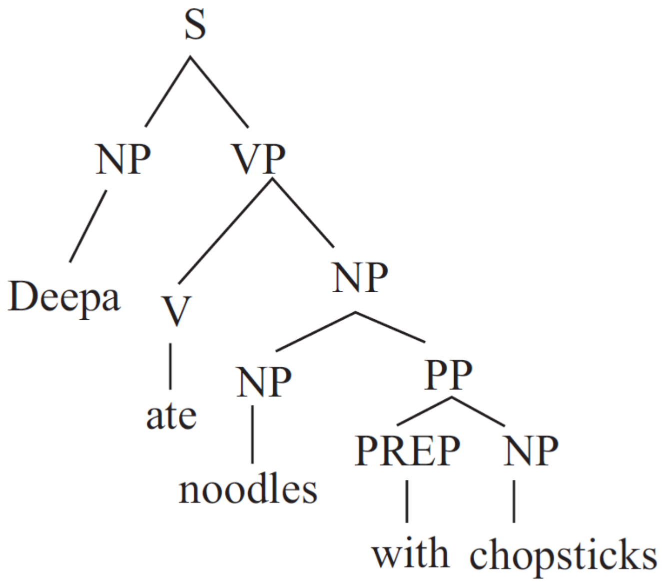
Recap of Classical Parsing

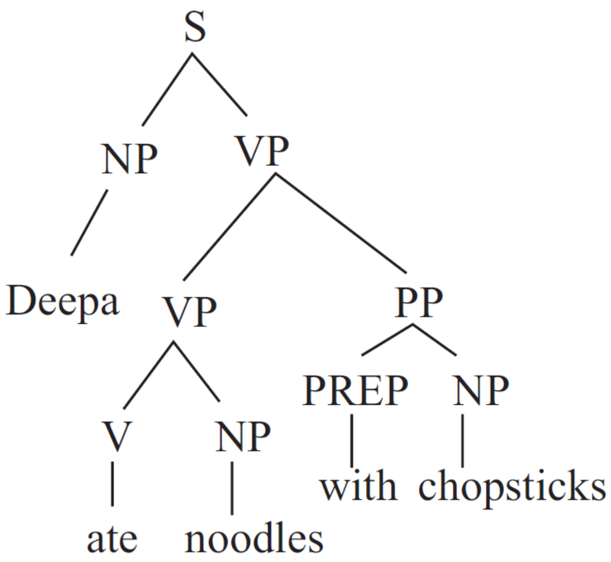
Deepa ate noodles with chopsticks

S → NP VP
VP → V NP
VP → VP PP
PP → PREP NP

NP → NP PP
NP → Deepa
NP → noodles
NP → chopsticks
NP → spoons

V → ate
PREP → with





Statistical versus Non-statistical approaches

- Non-statistical : parser + syntactic disambiguation

- Statistical : parsing all the way down

Key Idea in Statistical Parsing

- Find possible parses
- Assign probabilities to them
- Find out the most probable one

Probabilistic Parsing

$$
\begin{array}{lll}
S & \longrightarrow & NP\ VP\ (1.0) \\
VP & \longrightarrow & V\ NP\ (0.5) \\
VP & \longrightarrow & VP\ PP\ (0.5) \\
PP & \longrightarrow & PREP\ NP\ (1.0)
\end{array}
\qquad
\begin{array}{lll}
NP & \longrightarrow & NP\ PP\ (0.2) \\
NP & \longrightarrow & \text{Deepa}\ (0.2) \\
NP & \longrightarrow & \text{noodles}\ (0.3) \\
NP & \longrightarrow & \text{chopsticks}\ (0.15) \\
NP & \longrightarrow & \text{spoons}\ (0.15)
\end{array}
\qquad
\begin{array}{lll}
V & \longrightarrow & \text{ate}\ (1.0) \\
PREP & \longrightarrow & \text{with}\ (1.0)
\end{array}
$$

AFCAI Chapter, pages 691-692

Machine Learning for NLP

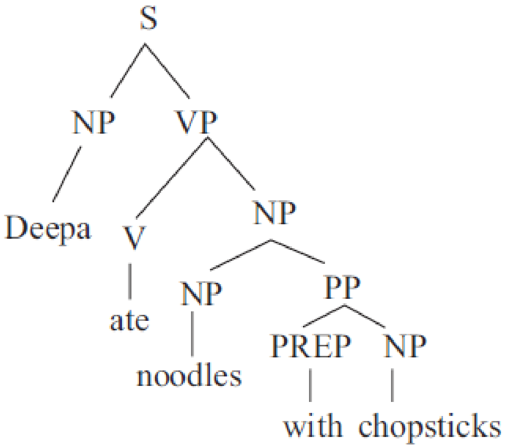
- Symbolic Backbone as a philosophy for integrating top down and bottom up knowledge

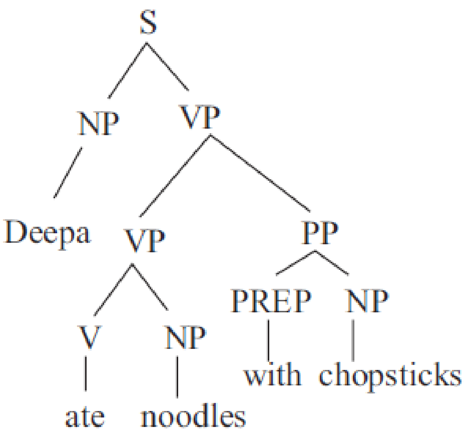
Probabilistic Parsing

$S \longrightarrow NP\ VP\ (1.0)$	$NP \longrightarrow NP\ PP\ (0.2)$	$V \longrightarrow ate\ (1.0)$
$VP \longrightarrow V\ NP\ (0.5)$	$NP \longrightarrow Deepa\ (0.2)$	$PREP \longrightarrow with\ (1.0)$
$VP \longrightarrow VP\ PP\ (0.5)$	$NP \longrightarrow noodles\ (0.3)$	
$PP \longrightarrow PREP\ NP\ (1.0)$	$NP \longrightarrow chopsticks\ (0.15)$	
	$NP \longrightarrow spoons\ (0.15)$	

$$\begin{aligned}P(\text{Parse}_1) = {} & P(S \to NP\ VP) \times P(VP \to V\ NP) \times P(NP \to NP\ PP) \times P(PP \to PREP\ NP) \\ & \times P(NP \to Deepa) \to P(V \to ate) \times P(NP \to noodles) \\ & \times P(PREP \to with) \times P(NP \to chopsticks) \\ = {} & 0.0009\end{aligned}$$

$$\begin{aligned}P(\text{Parse}_2) = {} & P(S \to NP\ VP) \times P(VP \to VP\ PP) \times P(VP \to V\ NP) \times P(PP \to PREP\ NP) \\ & \times P(NP \to Deepa) \times P(V \to ate) \times P(NP \to noodles) \\ & \times P(PREP \to with) \times P(NP \to chopsticks) \\ = {} & 0.0022\end{aligned}$$

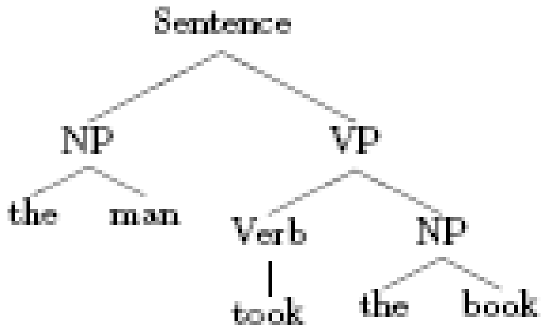




AFCAI Chapter, pages 691-692

Property 1: Rule probabilities corresponding to each non-terminal must sum to 1

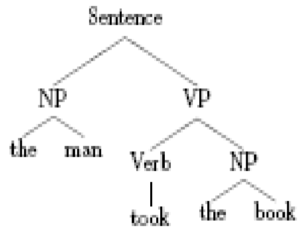
CFG rule	Rule probability
Sentence $\longrightarrow$ NP VP	$p($ Sentence $\longrightarrow$ NP VP $) = 1$
NP $\longrightarrow$ the man	$p($ NP $\longrightarrow$ the man $)$ summing to 1
NP $\longrightarrow$ the book	$p($ NP $\longrightarrow$ the book $)$ summing to 1
VP $\longrightarrow$ Verb NP	$p($ VP $\longrightarrow$ Verb NP $) = 1$
Verb $\longrightarrow$ took	$p($ Verb $\longrightarrow$ took $) = 1$



$$\boxed{p(\text{S} \longrightarrow \text{NP VP}) \cdot p(\text{NP} \longrightarrow \text{the man}) \cdot p(\text{VP} \longrightarrow \text{Verb NP}) \cdot p(\text{Verb} \longrightarrow \text{took}) \cdot p(\text{NP} \longrightarrow \text{the book})}$$

Workbook

Property 2: Probabilities of all full parse-trees must sum up to 1



$$p(\text{S} \longrightarrow \text{NP VP}) \cdot p(\text{NP} \longrightarrow \text{the man}) \cdot p(\text{VP} \longrightarrow \text{Verb NP}) \cdot p(\text{Verb} \longrightarrow \text{took}) \cdot p(\text{NP} \longrightarrow \text{the book})$$

$$p(\text{NP} \longrightarrow \text{the man}) \cdot p(\text{NP} \longrightarrow \text{the book})$$

CFG rule	**Rule probability**
Sentence $\longrightarrow$ NP VP	$p(\text{Sentence} \longrightarrow \text{NP VP}) = 1$
NP $\longrightarrow$ the man	$p(\text{NP} \longrightarrow \text{the man})$ } summing to 1
NP $\longrightarrow$ the book	$p(\text{NP} \longrightarrow \text{the book})$ }
VP $\longrightarrow$ Verb NP	$p(\text{VP} \longrightarrow \text{Verb NP}) = 1$
Verb $\longrightarrow$ took	$p(\text{Verb} \longrightarrow \text{took}) = 1$

Probabilistic Grammar Checking :

$$\begin{aligned}
p(\mathcal{X}) &= p(\text{NP} \longrightarrow \text{the man}) \cdot p(\text{NP} \longrightarrow \text{the book}) + \\
&\quad p(\text{NP} \longrightarrow \text{the book}) \cdot p(\text{NP} \longrightarrow \text{the book}) + \\
&\quad p(\text{NP} \longrightarrow \text{the man}) \cdot p(\text{NP} \longrightarrow \text{the man}) + \\
&\quad p(\text{NP} \longrightarrow \text{the book}) \cdot p(\text{NP} \longrightarrow \text{the man}) \\
&= \Big(p(\text{NP} \longrightarrow \text{the man}) + p(\text{NP} \longrightarrow \text{the book})\Big) \cdot p(\text{NP} \longrightarrow \text{the book}) + \\
&\quad \Big(p(\text{NP} \longrightarrow \text{the man}) + p(\text{NP} \longrightarrow \text{the book})\Big) \cdot p(\text{NP} \longrightarrow \text{the man}) \\
&= 1
\end{aligned}$$

Workbook

A Non-standard PCFG

$$\begin{array}{lll} S & \longrightarrow NP\ sleeps & (1.0) \\ S & \longrightarrow John\ sleeps & (0.7) \\ NP & \longrightarrow John & (0.3) \end{array}$$

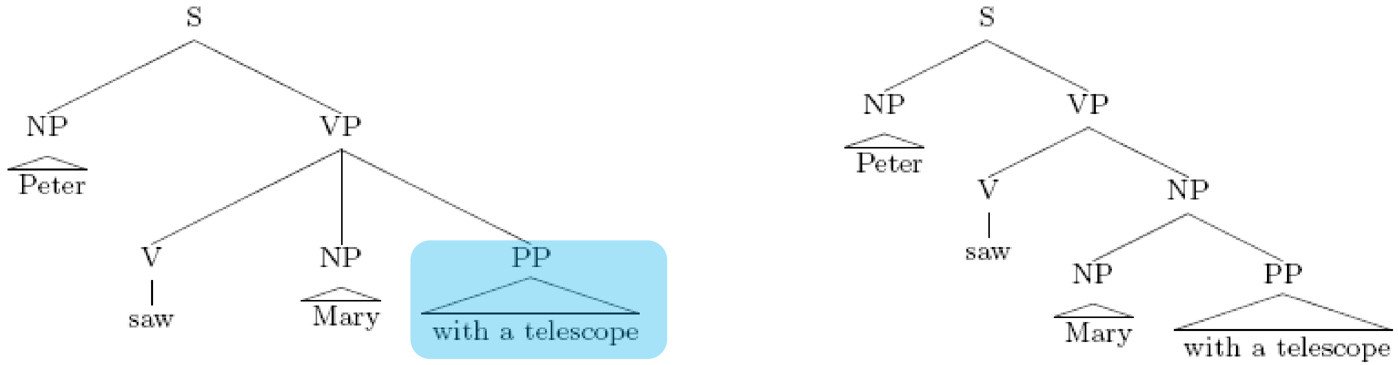
After Repair

$$\begin{array}{lll} S & \longrightarrow NP\ sleeps & (0.3) \\ S & \longrightarrow John\ sleeps & (0.7) \\ NP & \longrightarrow John & (1.0) \end{array}$$

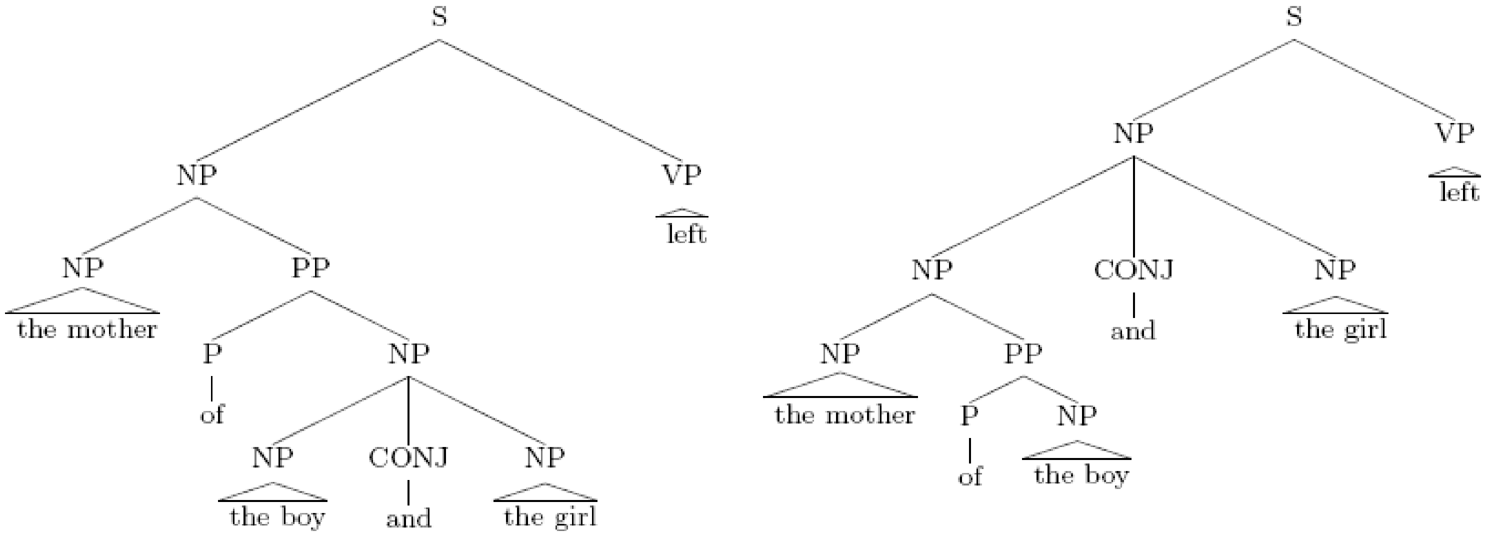
Workbook

Resolving Ambiguities using PCFG

Ambiguity caused by prepositional-phrase attachment:

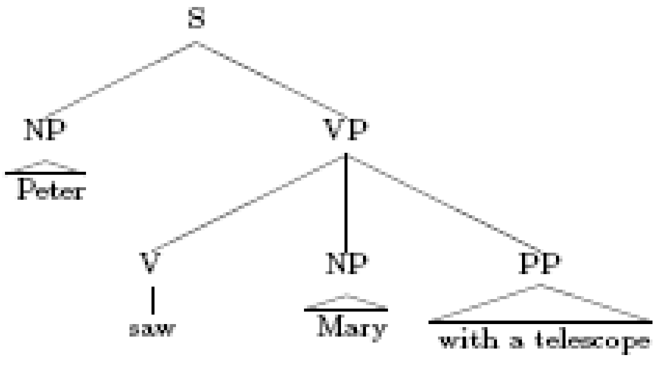


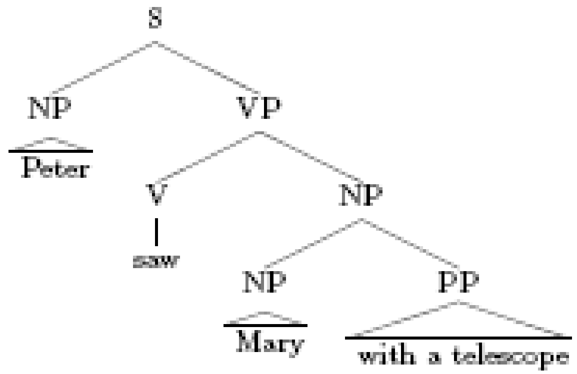
Ambiguity caused by conjunctions:



Workbook

Resolving Ambiguities





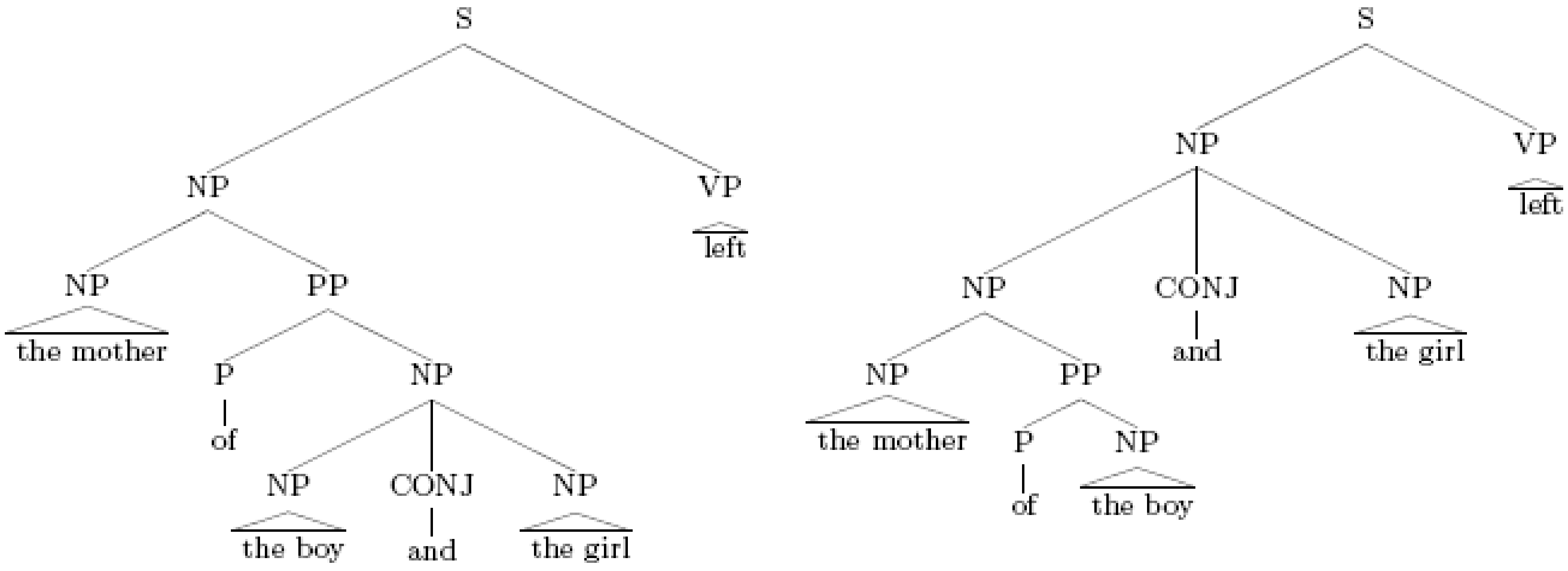
$$p(\mathrm{S} \longrightarrow \mathrm{NP\ VP}) \cdot p\left(\overset{\mathrm{NP}}{\widehat{\text{Peter}}}\right) \cdot$$
$$p(\mathrm{VP} \longrightarrow \mathrm{V\ NP\ PP}) \cdot$$
$$p(\mathrm{V} \longrightarrow \text{saw}) \cdot p\left(\overset{\mathrm{NP}}{\widehat{\text{Mary}}}\right) \cdot p\left(\overset{\mathrm{PP}}{\widehat{\text{with a telescope}}}\right)$$

$$p(\mathrm{S} \longrightarrow \mathrm{NP\ VP}) \cdot p\left(\overset{\mathrm{NP}}{\widehat{\text{Peter}}}\right) \cdot$$
$$p(\mathrm{VP} \longrightarrow \mathrm{V\ NP}) \cdot p(\mathrm{NP} \longrightarrow \mathrm{NP\ PP})$$
$$p(\mathrm{V} \longrightarrow \text{saw}) \cdot p\left(\overset{\mathrm{NP}}{\widehat{\text{Mary}}}\right) \cdot p\left(\overset{\mathrm{PP}}{\widehat{\text{with a telescope}}}\right)$$

$$p(\mathrm{VP} \longrightarrow \mathrm{V\ NP\ PP}) > p(\mathrm{VP} \longrightarrow \mathrm{V\ NP}) \cdot p(\mathrm{NP} \longrightarrow \mathrm{NP\ PP})$$

Workbook

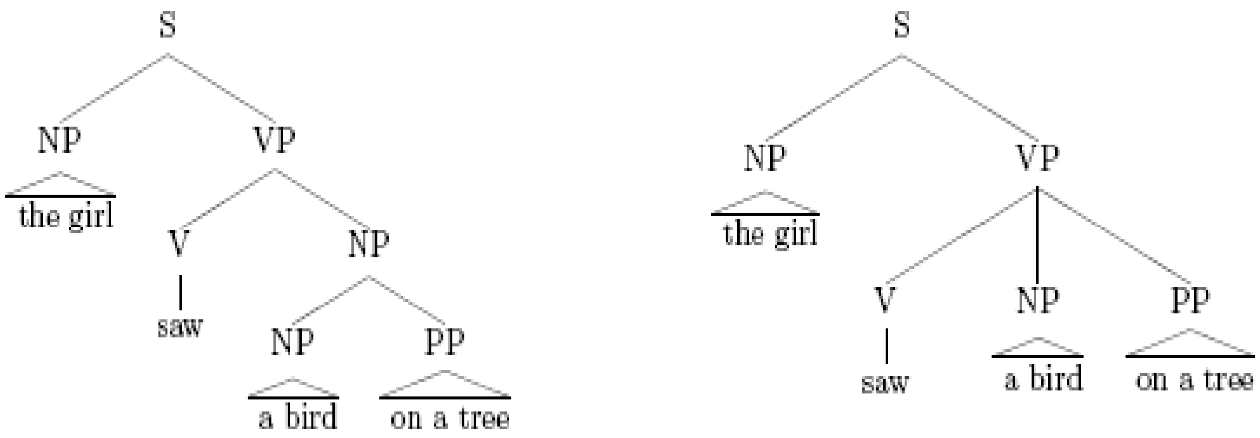
Catch 1



Different structures but the same set of CGF rules, hence probabilistic parser is not able to disambiguate

Workbook

Catch 2



$$p(\mathbf{VP} \longrightarrow \mathbf{V\ NP}) \cdot p(\mathbf{NP} \longrightarrow \mathbf{NP\ PP}) > p(\mathbf{VP} \longrightarrow \mathbf{V\ NP\ PP})$$

Contrast this with the requirement in slide 12:

$$p(\mathbf{VP} \longrightarrow \mathbf{V\ NP\ PP}) > p(\mathbf{VP} \longrightarrow \mathbf{V\ NP}) \cdot p(\mathbf{NP} \longrightarrow \mathbf{NP\ PP})$$

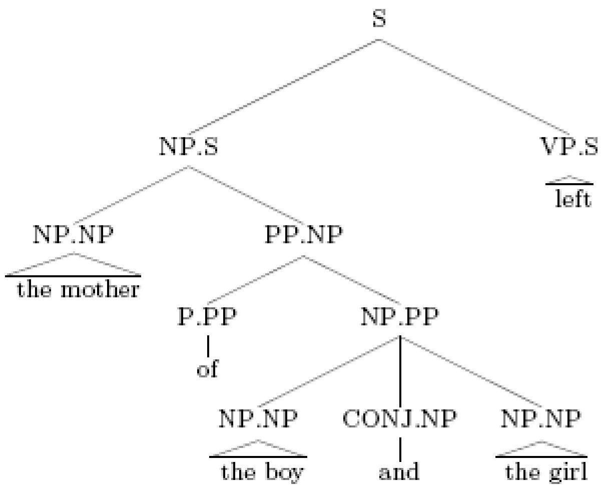
Workbook

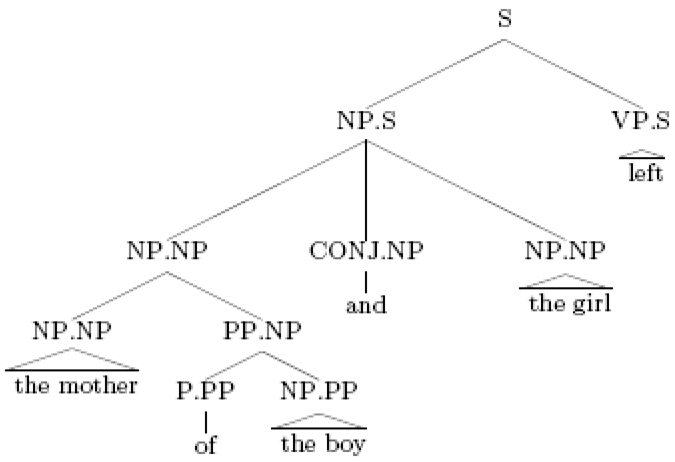
Solution 1

$$p(\mathbf{VP} \longrightarrow \mathbf{V\ NP\ PP\text{-}ON}) \quad < \quad p(\mathbf{VP} \longrightarrow \mathbf{V\ NP}) \cdot p(\mathbf{NP} \longrightarrow \mathbf{NP\ PP\text{-}ON})$$
$$p(\mathbf{VP} \longrightarrow \mathbf{V\ NP\ PP\text{-}WITH}) \quad > \quad p(\mathbf{VP} \longrightarrow \mathbf{V\ NP}) \cdot p(\mathbf{NP} \longrightarrow \mathbf{NP\ PP\text{-}WITH})$$

Workbook

Solution 2 : Parent Encoding (Grammar Transformation)





In this example, we will choose the analysis at the left-hand side, if

$$p(\,\mathbf{NP.S} \longrightarrow \text{NP.NP PP.NP}\,) \cdot p(\,\mathbf{NP.PP} \longrightarrow \text{NP.NP CONJ.NP NP.NP}\,) \cdot p\left(\begin{array}{c}\mathbf{NP.NP}\\ \triangle \\ \text{the boy}\end{array}\right)$$

is more likely than

$$p(\,\mathbf{NP.NP} \longrightarrow \text{NP.NP PP.NP}\,) \cdot p(\,\mathbf{NP.S} \longrightarrow \text{NP.NP CONJ.NP NP..NP}\,) \cdot p\left(\begin{array}{c}\mathbf{NP.PP}\\ \triangle \\ \text{the boy}\end{array}\right)$$

Workbook

Another consequence of recursion: Non-local dependencies

The **bird** with the nice brown eyes and the beautiful tail feathers **catches** a worm.

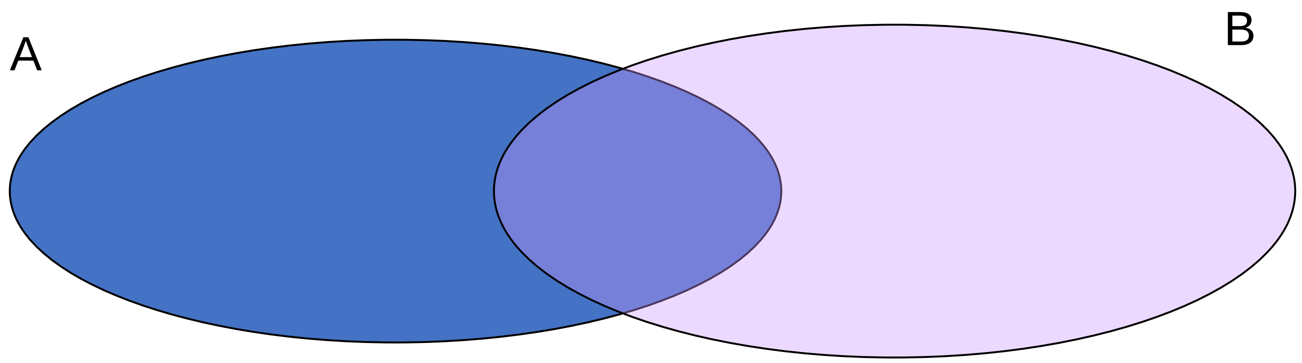
Workbook

How do we get the PCFG rule probabilities?

- Next class

Workbook

Evaluation: Labeled Precision and Labeled Recall



- |A|: total number of non-terminal constituents the parser postulated.
- |B|: total number of non-terminal constituents in the tree-bank version
- |A∩B| : the number of correct constituents found by the parser (summed over all sentences in the test set)

- Precision = |A∩B | / |A|
- Recall = |A∩B | / |B|
- Important: Non terminals pertaining to part of speech tags (POSTs) are not considered

Charniak's AI Mag article

Precision and Recall

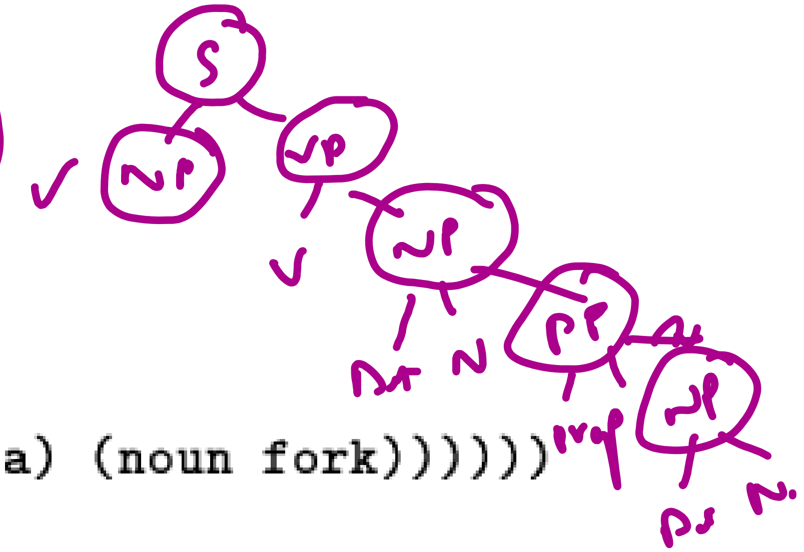
Recall $\left(\frac{5}{6}\right)$

Correct :

```
(s (np (det The) (noun stranger))
   (vp (verb ate)
       (np (det the) (noun doughnut))
       (pp (prep with) (np (det a) (noun fork)))))
```

Parser output :

Precision $\left(\frac{5}{6}\right)$




```
(s (np (det The) (noun stranger))
   (vp (verb ate)
       (np (det the) (noun doughnut)
           (pp (prep with) (np (det a) (noun fork))))))
```

Refer: Charniak's AI Mag article

State-of-the-art?

http://homepages.herts.ac.uk/~comqpcl/downloads/phdthesis.pdf

Homework

- Please go through the AI Magazine paper by Charniak (shared) and find more on the problem of Part of Speech tagging, as explained in the paper

- You can ignore details on HMM and lexicalized parsing for now

References

- Workbook excerpt is from “A Tutorial on the Expectation-Maximization Algorithm Including Maximum-Likelihood Estimation and EM Training of Probabilistic Context-Free Grammars” by Detlef Prescher (Presented at the 15th European Summer School in Logic, Language and Information (ESSLLI 2003))

- Draft chapter from textbook “A First Course in Artificial Intelligence” by Prof. Deepak Khemani (McGraw Hill Education; 1st edition (2017))

- Charniak, E. (1997). Statistical Techniques for Natural Language Parsing. *AI Magazine*, *18*(4), 33. https://doi.org/10.1609/aimag.v18i4.1320