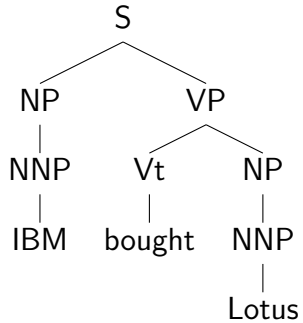
Weaknesses of Probabilistic Context-Free Grammars

Michael Collins, Columbia University

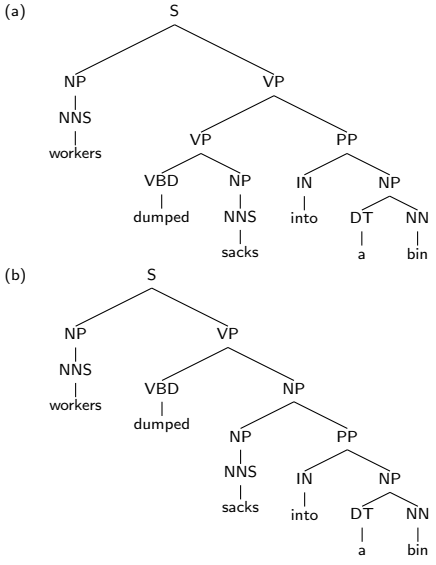
Weaknesses of PCFGs

- Lack of sensitivity to lexical information
- Lack of sensitivity to structural frequencies



$$
\begin{aligned}
\text{p(t)} = \quad & q(\text{S} \rightarrow \text{NP VP}) && \times q(\text{NNP} \rightarrow IBM) \\
& \times q(\text{VP} \rightarrow \text{V NP}) && \times q(\text{Vt} \rightarrow bought) \\
& \times q(\text{NP} \rightarrow \text{NNP}) && \times q(\text{NNP} \rightarrow Lotus) \\
& \times q(\text{NP} \rightarrow \text{NNP})
\end{aligned}
$$

Another Case of PP Attachment Ambiguity



(a)

Rules
S → NP VP
NP → NNS
VP → VP PP
VP → VBD NP
NP → NNS
PP → IN NP
NP → DT NN
NNS → workers
VBD → dumped
NNS → sacks
IN → into
DT → a
NN → bin

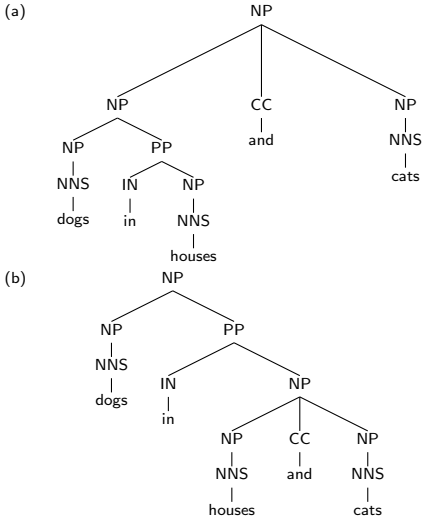
(b)

Rules
S → NP VP
NP → NNS
NP → NP PP
VP → VBD NP
NP → NNS
PP → IN NP
NP → DT NN
NNS → workers
VBD → dumped
NNS → sacks
IN → into
DT → a
NN → bin

If $q(\text{NP} \rightarrow \text{NP PP}) > q(\text{VP} \rightarrow \text{VP PP})$ then (b) is more probable, else (a) is more probable.

Attachment decision is completely independent of the words

A Case of Coordination Ambiguity



(a)

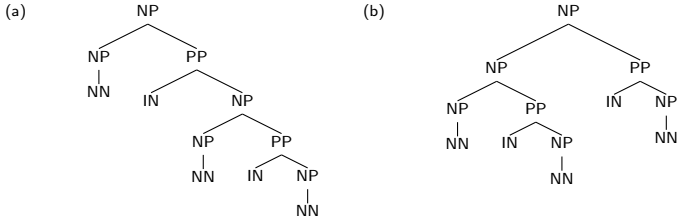
Rules
NP → NP CC NP
NP → NP PP
NP → NNS
PP → IN NP
NP → NNS
NP → NNS
NNS → dogs
IN → in
NNS → houses
CC → and
NNS → cats

(b)

Rules
NP → NP CC NP
NP → NP PP
NP → NNS
PP → IN NP
NP → NNS
NP → NNS
NNS → dogs
IN → in
NNS → houses
CC → and
NNS → cats

Here the two parses have identical rules, and therefore have identical probability under any assignment of PCFG rule probabilities

Structural Preferences: Close Attachment



- Example: president of a company in Africa
- Both parses have the same rules, therefore receive same probability under a PCFG
- "Close attachment" (structure (a)) is twice as likely in Wall Street Journal text.

Structural Preferences: Close Attachment

Previous example: John was believed to have been shot by Bill

Here the low attachment analysis (Bill does the *shooting*) contains same rules as the high attachment analysis (Bill does the *believing*), so the two analyses receive same probability.