Structured predictions: trees and graphs

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

- Syntactic parsing
  - Phrase structure (or constituent) trees
  - Dependency trees
- Graph-based semantic parsing
  - Abstract Meaning Representations
- Needs a grammar that dictates admissible or non-admissible trees for a sentence
- Requires an efficient decoding algorithm (e.g., CKY), like sequence labeling
- Same statistical or neural models extended to parsing, e.g., Naïve Bayes, Perceptron, Neural networks.

Syntactic Parsing

https://powcoder.com

- Syntactic parsing is the task of assigning a syntactic structure (typically in the form of a tree) to a sentence
- The grammar formalism dictates what a tree looks like syntactically
- The main technical challenge is ambiguity (like many other NLP problems), the fact that multiple trees are possible for a sentence given a grammar
- The general solution to the problem is to build a model that assigns a "score" to each possible tree, and using a decoding (search) algorithm to find the best tree (or top-k trees)
- Syntactic parsing is one of the core problems in NLP, and syntactic structures are useful for a wide range of NLP applications

Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Two broad schools of thought on the syntactic structure

https://powcoder.com

- Phrase structures
  - The building blocks of phrase structures are “phrases” that are consecutive groups of words that make sense linguistically
  - Each phrase receive a syntactic category (e.g., “NP”, “VP”)
  - The phrases are hierarchically organized into a tree structure
- Dependency structures
  - The fundamental building blocks of a dependency structure are relations between words in a sentence that may or may not be consecutive (They are non-consecutive more often than not)
  - Dependency relations are thus naturally lexicalized
  - The relations are hierarchically organized
  - The syntactic relations between the words are typically labeled
- Each category of syntactic structures also have different flavors

Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

https://powcoder.com

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

Phrase-structure parsing based on context-free grammars

https://powcoder.com

Add WeChat powcoder

Context-free grammars

https://powcoder.com

Assignment Project Exam Help

Context-free languages are specified by context-free grammars, which are tuples of (N, Σ, R, S) consisting of:

Add WeChat powcoder

Assignment Project Exam Help

- a finite set of **non-terminals** N;
- a finite alphabet Σ of **terminal symbols**;
- a set of **production rules** R, each of the form $A \rightarrow \beta$, where $A \in N$ and $\beta \in (\Sigma \cup N)*$
- a designated start symbol S

https://powcoder.com

Add WeChat powcoder

Example CFG: mathematical operations

https://powcoder.com

Assignment Project Exam Help

Assignment Project Exam Help

Add WeChat powcoder

$S \rightarrow \textsc{S Op S}|\textsc{Num}$

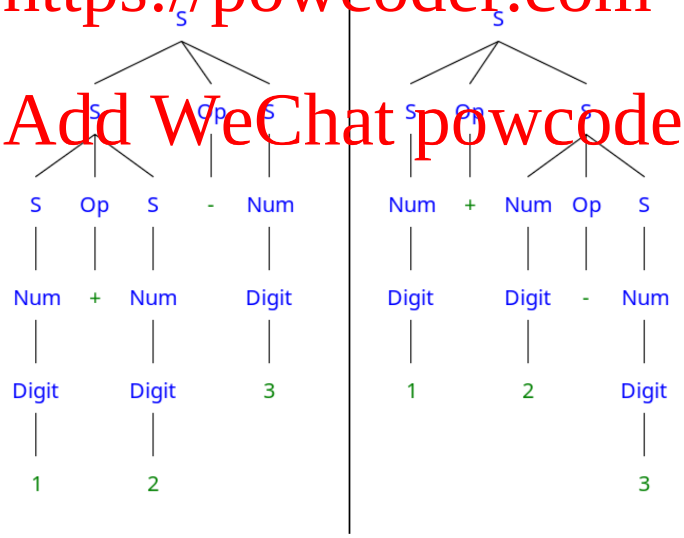
$\textsc{Op} \rightarrow +|-|\times|\div$

$\textsc{Num} \rightarrow \textsc{Num Digit}|\textsc{Digit}$

$\textsc{Digit} \rightarrow 0|1|2|\cdots|9$

https://powcoder.com

Add WeChat powcoder



Is natural language context-free?

https://powcoder.com

- Context-free grammars can be used to represent syntax, which is the set of rules that determine whether an utterance is judged to be grammatical.

Assignment Project Exam Help

- If this representation were perfectly faithful, then a natural language such as English could be transformed into a formal language, consisting of exactly the (infinite) set of strings that would be judged to be grammatical by a fluent English speaker.

Assignment Project Exam Help
Add WeChat powcoder
https://powcoder.com
Add WeChat powcoder

- Contemporary theories generally do not consider natural languages to be context-free, yet context-free grammars are widely used in natural language parsing.
- The reason is that context-free representations strike a good balance: they cover a broad range of syntactic phenomena, and they can be parsed efficiently.

A phrase-structure grammar

https://powcoder.com

A phrase-structure grammar is one in which sentences are broken down into constituents (phrases), which are contiguous sequences of words that function as coherent units for the purpose of linguistic analysis.

Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help

Phrases are labeled with a type that are determined by their **heads**: noun phrases (NP), verb phrases (VP), etc.

https://powcoder.com
Add WeChat powcoder

S → NP VP
NP → NP PP |we|sushi|chopsticks
NP → NP CC NP
VP → V NP |VP PP
PP → IN NP
V → eat
CC → and

Chomsky Normal Form

https://powcoder.com

Assignment Project Exam Help

- In **Chomsky Normal Form (CNF)**, the right-hand side of every production includes either two non-terminals, or a single terminal symbol.

Add WeChat powcoder
Assignment Project Exam Help

- CNF can be parsed efficiently in cubic time.

https://powcoder.com

- But some CFG productions are not naturally CNF, e.g., NP → NP CC NP.

Add WeChat powcoder

- The general practice in syntactic parsing is to convert CFGs to CNF during training and decoding.

Ambiguity of different types

https://powcoder.com
Assignment Project Exam Help
Assignment Project Exam Help
Add WeChat powcoder
https://powcoder.com
Add WeChat powcoder

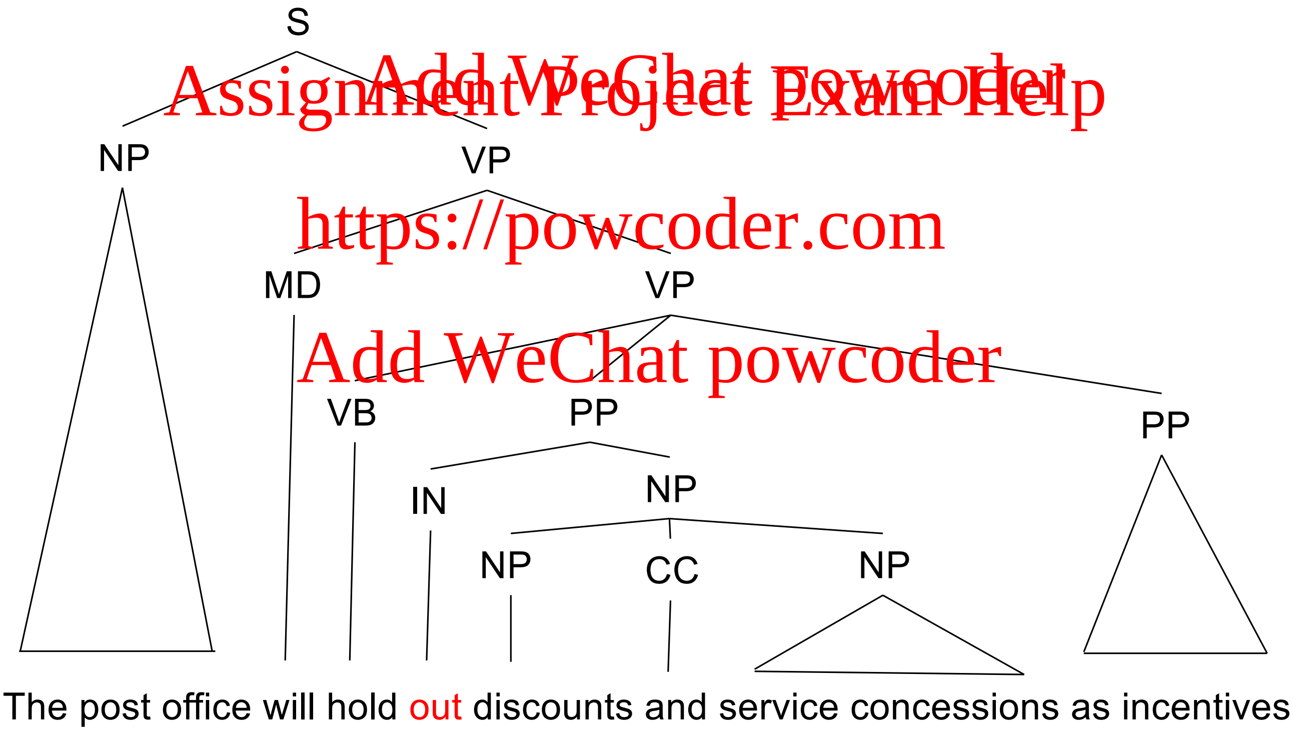
Even with a grammar, multiple parses (trees) for a sentence are often possible due to ambiguity:

- Attachment ambiguity: e.g., We eat sushi with chopsticks, I shot an elephant in my pajamas.
- Modifier scope: e.g., southern food store, plastic cup holder
- Particle vs preposition: e.g., The puppy tore up the staircase.
- Complement structure: e.g., The students complained to the professor that they didn't understand.
- Coordination scope: e.g., “I see,” said the blind man, as he picked up the hammer and saw.

Spurious ambiguity

In practice, a grammar often allows parses that do not correspond with human intuition at all. Such cases are called **spurious ambiguity**

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder


S
NP
VP
MD
VP
VB
PP
PP
IN
NP
NP
CC
NP
The post office will hold out discounts and service concessions as incentives

Weighted context-free grammars

The major challenge in syntactic parsing is not just to find a possible tree for a sentence, but to find the best tree among all possible trees. One way to do that is to have weighted grammar so we can weight the trees...

https://powcoder.com
Assignment Project Exam Help
Assignment Project Exam Help
Add WeChat powcoder
https://powcoder.com
Add WeChat powcoder

S → NP VP	S → NP VP	-3
NP → NNP	NP → NNP	-1
NP → NP CC NP	NP → NP1 NP	-0.5
	NP1 → NP CC	-0.5
NP → ADJP NP	NP → ADJP NP	-0.5
NP → NNS	NP → NNS	-1
NP → NN	NP → NN	-1
NP → NP PP	NP → NP PP	-0.3
VP → VBZ NP	VP → VBZ NP	-0.5
VP → VBD NP	VP → VBD NP	-0.5
VP → VP PP	VP → VP PP	-0.2
PP → IN NP	PP → IN NP	0
ADJP → JJ	ADJP → JJ	0
NNP → John	NNP → John	0

Probabilistic context-free grammars

And one special form of weighted context-free grammars is **probabilistic context free grammars** where the sum of all rules with the same left hand side is 1:

S → NP VP	S → NP VP	1.0
NP → NNP	NP → NNP	0.1
NP → NP CC NP	NP → NP1 NP	0.2
[illegible]	NP1 → NP CC	1.0
NP → ADJP NP	NP → ADJP NP	0.2
NP → NNS	NP → NNS	0.1
NP → NN	NP → NN	0.1
NP → NP PP	NP → NP PP	0.3
VP → VBZ NP	VP → VBZ NP	0.2
VP → VBD NP	VP → VBD NP	0.3
VP → VP PP	VP → VP PP	0.5
PP → IN NP	PP → IN NP	1.0
ADJP → JJ	ADJP → JJ	1.0
NNP → John	NNP → John	1.0
NNS → apples	NNS → apples	0.4
NNS → oranges	NNS → oranges	0.4
NNS → noodles	NNS → noodles	0.2
NN → gravy	NN → gravy	1.0
JJ → green	JJ → green	1.0
IN → with	IN → with	1.0
VBZ → likes	VBZ → likes	1.0
VBD → ate	VBD → ate	1.0

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

The probability model of a phrase structure tree

Independence assumptions:

https://powcoder.com

- Place invariance: Location of a subtree does not impact the probability of the subtree

Assignment Project Exam Help

- Context Free-ness: Probability of a subtree not impacted by words outside of the subtree

Add WeChat powcoder

Assignment Project Exam Help

- Ancestor Free-ness: Probability of a subtree does not depend on non-terminals outside of the subtree

https://powcoder.com

Given these assumptions, the probability of a tree (the joint probability of the tree and the input sentence) is:

Add WeChat powcoder

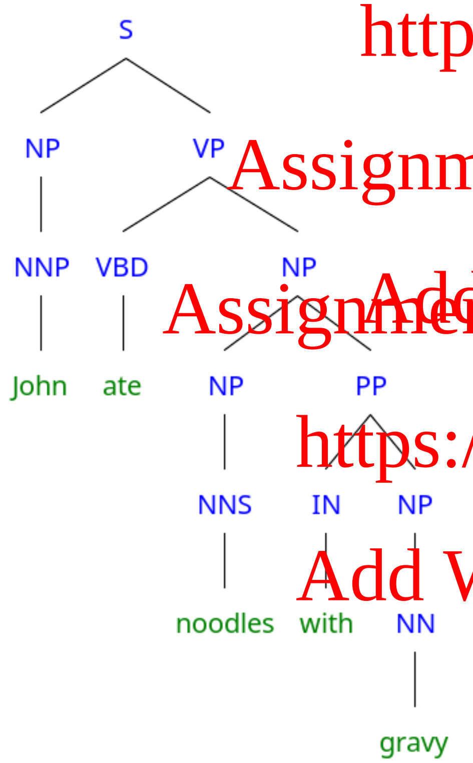
$$P(\mathcal{T}) = \prod_{i=1}^{m} P(RHS_i | LHS_i)$$

The decoding process is to find the tree that has the highest probability:

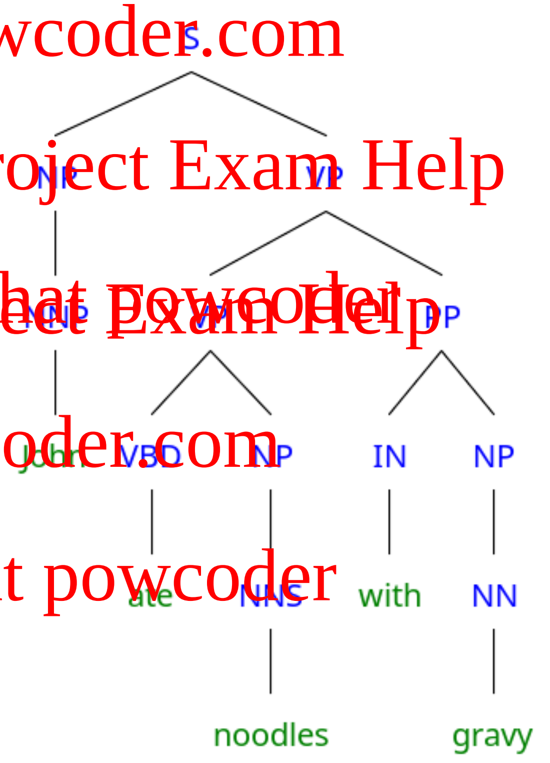
$$\hat{\mathcal{T}}(s) = \underset{\mathcal{T} s.t. s=yield(\mathcal{T})}{\text{argmax}} P(\mathcal{T})$$

Probability of an example tree

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



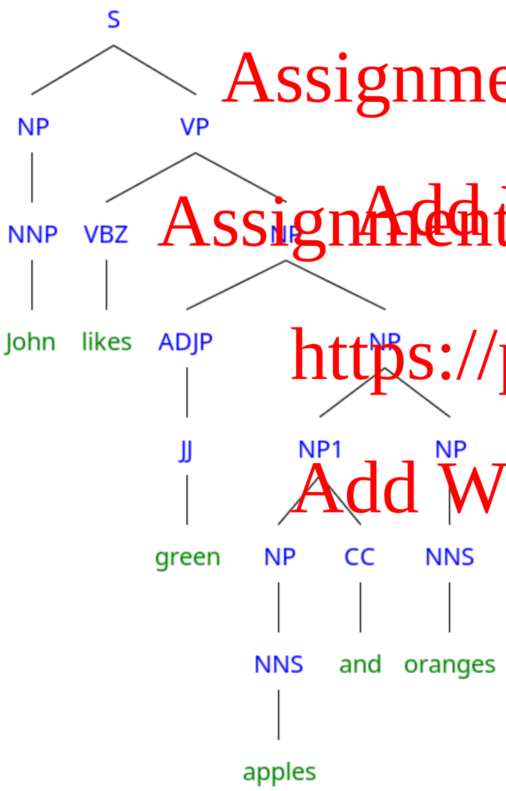
$P(T_1) = 1.0 \times 0.1 \times 1.0 \times 0.3 \times 1.0 \times 0.3 \times 0.1 \times 0.2 \times 1.0 \times 1.0 \times 0.1 \times 1.0 = 0.000018$

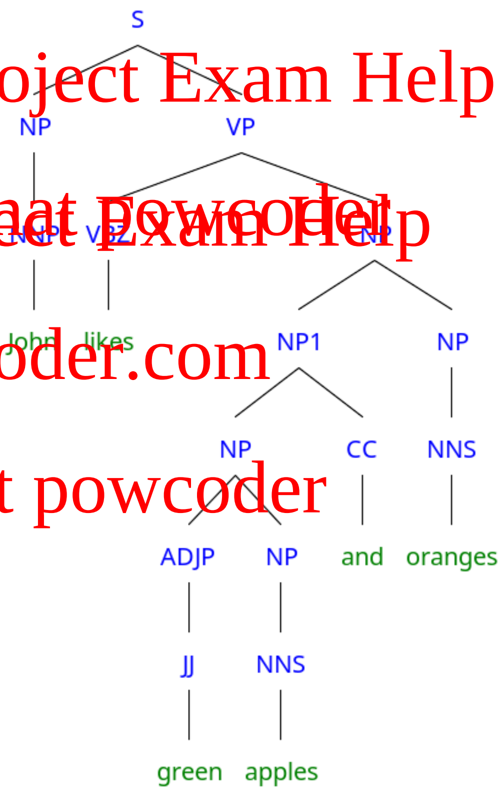


$P(T_2) = 1.0 \times 0.1 \times 1.0 \times 0.5 \times 0.3 \times 1.0 \times 0.1 \times 0.2 \times 1.0 \times 1.0 \times 0.1 \times 1.0 = 0.00003$

Probability of an example tree

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder





Parsing with PCFGs: the CKY algorithm

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

```
procedure WCKY($\boldsymbol{w}, G = (N, \Sigma, R, S)$)
    for all $i, j, X$ do
        $t[i, j, X] \leftarrow 0$
        $b[i, j, X] \leftarrow \emptyset$
    for $m \in \{1, 2, \cdots, M\}$ do
        for all $X \in N$ do
            $t[m, m+1, X] \leftarrow \psi(X \rightarrow w_m, (m, m+1, m))$
    for $\ell \in \{2, 3, \cdots, M\}$ do
        for $m \in \{0, 1, \cdots, M - \ell\}$ do
            for $k \in \{m+1, m+2, \cdots, m+\ell-1\}$ do
                $t[m, m+\ell, X] \leftarrow \max_{k,Y,Z} \psi(X \rightarrow Y\ Z, (m, m+\ell, k)) + t[m, k, Y] + t[k, m+\ell, Z]$
                $b[m, m+\ell, X] \leftarrow \text{argmax}_{k,Y,Z} \psi(X \rightarrow Y\ Z, (m, m+\ell, k)) + t[m, k, Y] + t[k, m+\ell, Z]$
    return TraceBack(S,0,M,b)
```

CKY example

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

		1	2	3	4	5
0	John	NNP:1.0 NP:0.1				
1		eats	VBD:1.0			
2			noodles	NNS:0.2 NP:0.02		
3				with	IN:1.0	
4					gravy	NN:1.0 NP:0.1

CKY example

https://powcoder.com

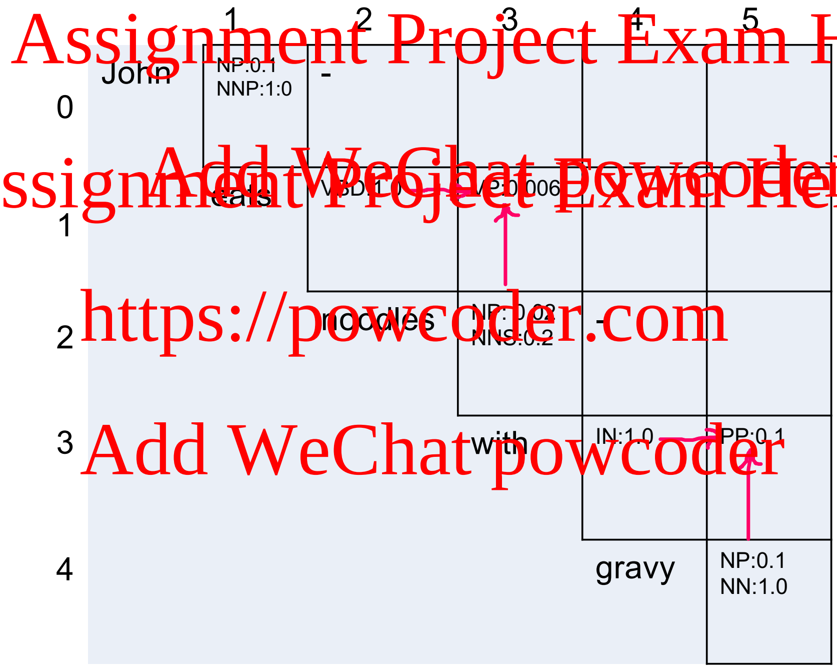
Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

https://powcoder.com

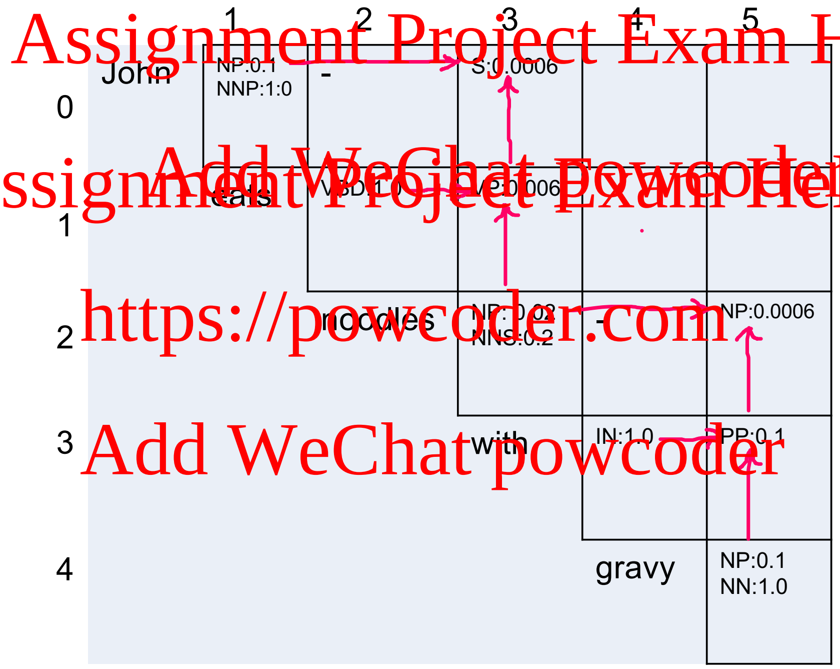
Add WeChat powcoder



		1	2	3	4	5
0	John	NP:0.1 NNP:1:0	-			
1		eats	VBD:1.0	VP:0.006		
2			noodles	NP:0.02 NNS:0.2	-	
3				with	IN:1.0	PP:0.1
4					gravy	NP:0.1 NN:1.0

CKY example

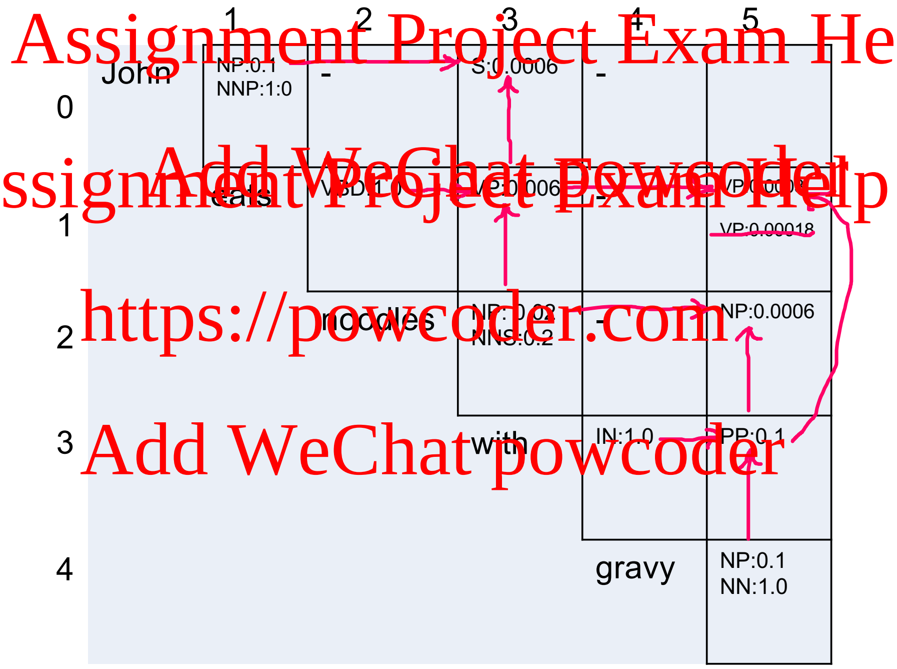
https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



		1	2	3	4	5
0	John	NP:0.1 NNP:1:0	-	S:0.0006		
1		eats	VBD:1.0	VP:0.006		
2			noodles	NP:0.02 NNS:0.2	-	NP:0.0006
3				with	IN:1.0	PP:0.1
4					gravy	NP:0.1 NN:1.0

CKY example

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

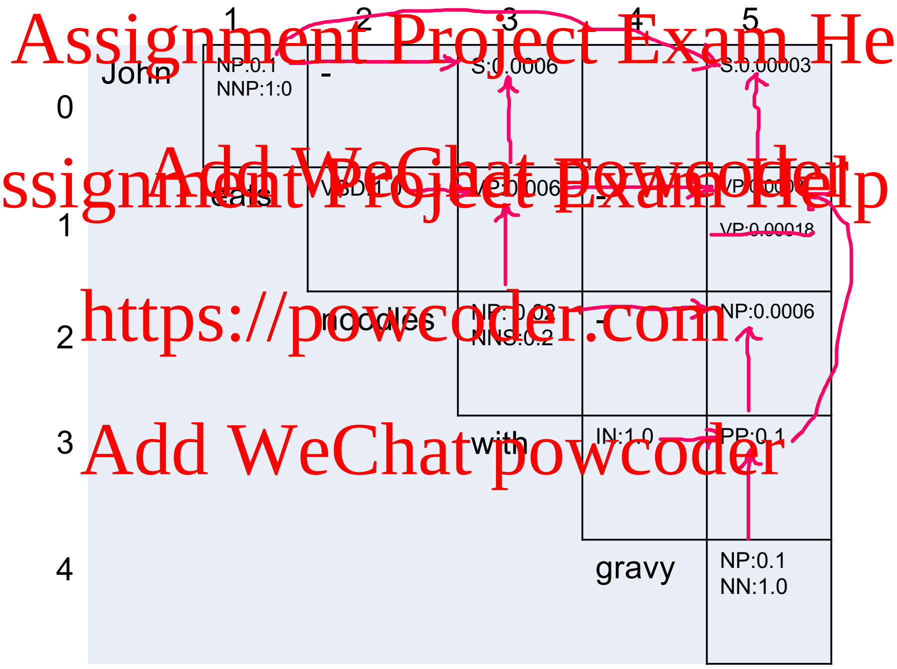


		1	2	3	4	5
0	John	NP:0.1 NNP:1:0	-	S:0.0006	-	
1		eats	VBD:1.0	VP:0.006	-	VP:0.0008 VP:0.00018
2			noodles	NP:0.02 NNS:0.2	-	NP:0.0006
3				with	IN:1.0	PP:0.1
4					gravy	NP:0.1 NN:1.0

CKY example

https://powcoder.com

Assignment Project Exam Help



		1	2	3	4	5
0	John	NP:0.1 NNP:1:0	-	S:0.0006		S:0.00003
1		eats	VBD:1.0	VP:0.006	-	VP:0.0003 ~~VP:0.00018~~
2			noodles	NP:0.02 NNS:0.2	-	NP:0.0006
3				with	IN:1.0	PP:0.1
4					gravy	NP:0.1 NN:1.0

Add WeChat powcoder

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Learning PCFGs

https://powcoder.com

Assignment Project Exam Help

Parameters in probabilistic context-free grammars can be estimated by relative frequency, as with HMMs:

Add WeChat powcoder

Assignment Project Exam Help

$$\psi(X \to \alpha) = \log P(X \to \alpha)$$

https://powcoder.com

$$\hat{P}(X \to \alpha) = \frac{count(X \to \alpha}{count(X)}$$

Add WeChat powcoder

E.g., the probability of the production NP → DET NN is the corpus count of this production, divided by the count of the non-terminal NP. This applies to terminals as well.

Grammar Refinement

https://powcoder.com

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

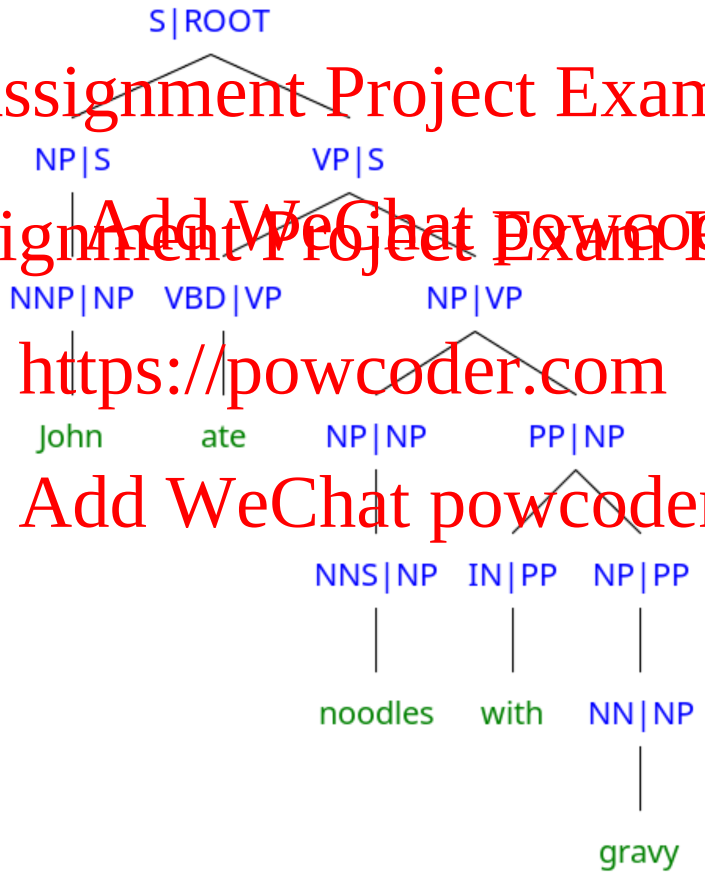
- Grammars extracted from treebanks (e.g., the Penn TreeBank) are often sensitive to ambiguities in the parses, even with the weighted productions

https://powcoder.com

- There are various attempt to augment with the vanilla PCFG with more expressive productions
    - Parent annotation
    - Lexicalization

Add WeChat powcoder

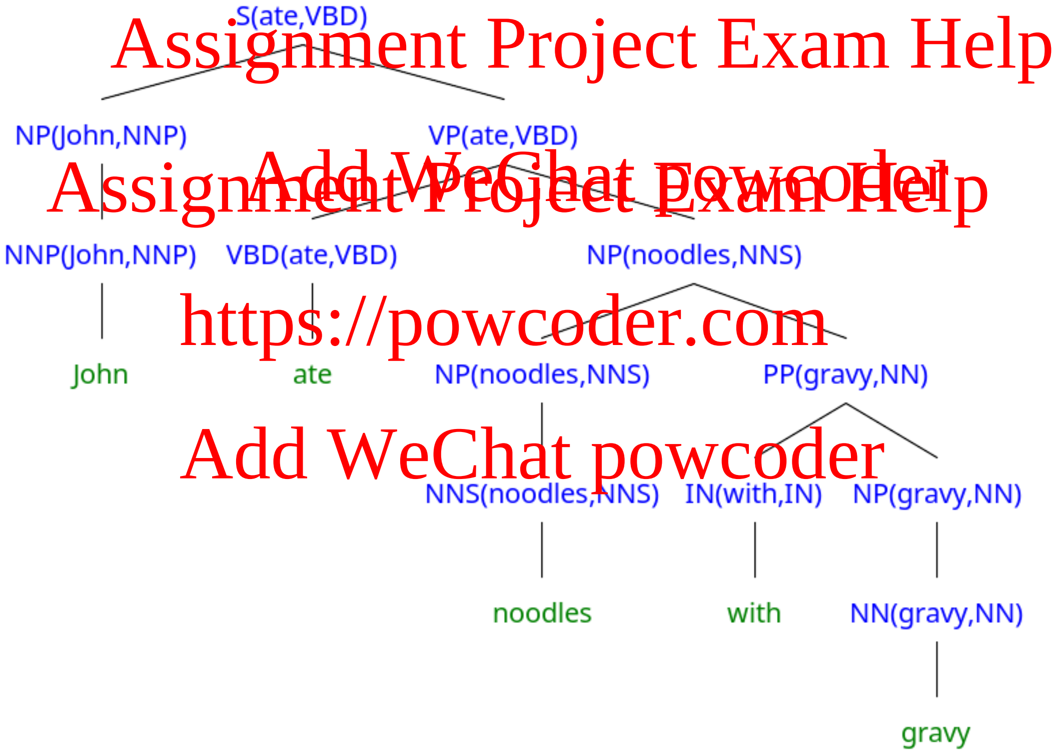
Parent annotation



https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Lexicalized CFGs

https://powcoder.com
Assignment Project Exam Help
Add WeChat powcoder
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder


S(ate,VBD)
NP(John,NNP)
VP(ate,VBD)
NNP(John,NNP)
VBD(ate,VBD)
NP(noodles,NNS)
John
ate
NP(noodles,NNS)
PP(gravy,NN)
NNS(noodles,NNS)
IN(with,IN)
NP(gravy,NN)
noodles
with
NN(gravy,NN)
gravy