

EMNLP Tutorial 3

Philipp Koehn

This tutorial deals with syntactic parsing.

Given the following data set:

Jim visits Hoboken

Jim is from Hoboken

John and Jim are happy

Happy Jim visits John

The city is Hoboken

John is in the city of Hoboken

1. Context-free grammar, phrase structure grammar:

- (a) Tag the sentences using part-of-speech tags of the Penn treebank (*NNP*, *NN*, *CC*, *VBZ*, *DT*, *IN*).
- (b) Write a grammar using context-free rules that parses the sentences in the data set. Use the phrase tags *NP*, *VP*, *PP*.
- (c) Perform chart parsing for the sentence
Jim visits John from Hoboken
- (d) Update your grammar, so that the following sentences will not be parsed. Add additional phrase or part-of-speech tags, if necessary.
 - * *Jim visits*
 - * *Hoboken visits John*
 - * *John is from the Hoboken*

2. Probabilistic context-free grammar:

- (a) Build parse trees for all the sentences in the data set using the grammar you wrote for Question 1b.
- (b) Given these parse trees, collect statistics for rules and compute conditional probabilities with maximum likelihood estimation.
- (c) Compute probabilities for parse trees for the following sentences:
 - John visits the city
 - Jim and Hoboken is from John

3. Lexical dependency grammar:

- (a) Add lexical heads to the parse trees you built in Question 2a
- (b) Collect statistics for the dependency relations in the lexicalized parse trees for head-head dependencies (parent tag/word, head child tag) and head-nonhead dependencies (parent tag/word, child tag/word, direction)