

Table des matières

[Tasks we had to do 2](#_Toc511842050)

[Results 3](#_Toc511842051)

*Assignment 2*

Dependency parsing & Part-of-speech tagging

**Mourad Akandouch**

**INFOY004**

**Academic year: 2017-2018**

**ULB - VUB**



**Section** 1

# *Tasks we had to do*

The aim of our first assignment was to write language models based on a training set. After that, regarding the generated language models, we had to detect in which varieties of English some sentences are written.

The classification was based on *perplexity.* The language models we had to build are based on trigrams letter and the training set consists of a corpus of three varieties of English: British, Australian and American.

In addition, we had to generate a random output based on the generated language models for each varieties of English. Those generated strings have a fixed length *k.* In this report, I will put a string of length *k = 200* as mentioned in the assignment.

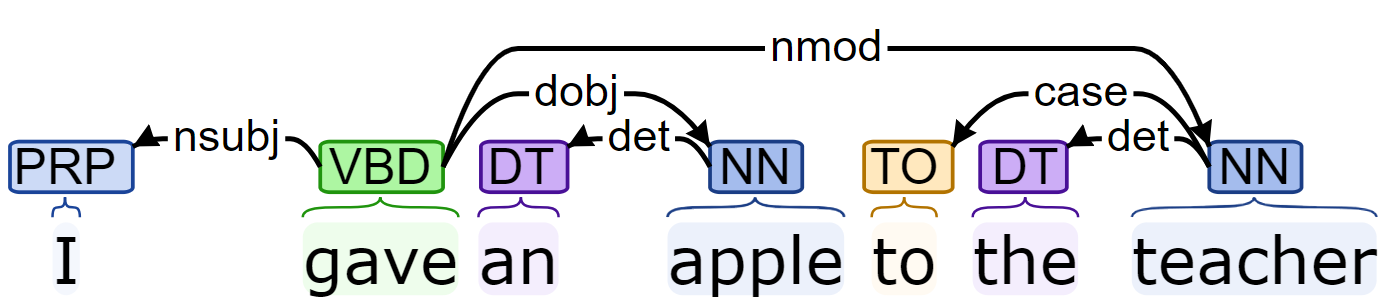
After that, we had to score each sentence in the test set with each of our generated language models.

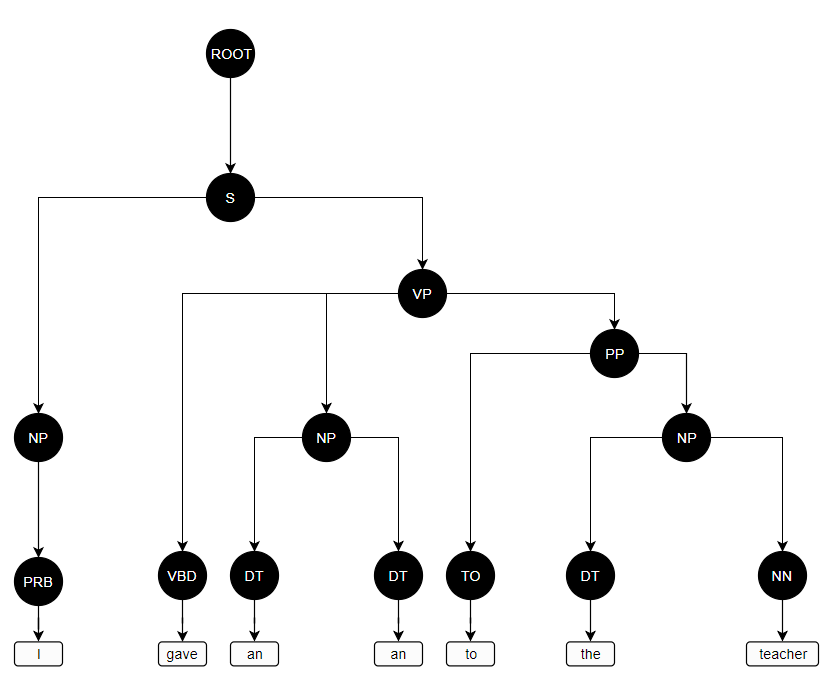
One important thing to mention is that we had to “clean” the training set by removing all non-alphabetical symbols and replacing all spaces by two underscores. I also added one underscore at the beginning and ending of the training text. I did so because the underscore’s role on our assignment is the detection of the beginning/ending of a word. A very easy way to clean the text is simply using regular expression.

**Section** 2

# *Part 1 - Annotation of toy sentences with dependency relations*

In this section, I will first discuss about the way I cleaned the text. Then, I will put a random output generated by each language models. One interesting thing is that the output generates words that *could* be in the English vocabulary. After that, I will put an excerpt of the language model for each of the varieties of English.





## Question 1

*Are there ambiguous sentences with multiple possible parses? If yes, what form does the ambiguity take?*

## Question 2

Which sentence contains non-projective dependency relations? Which relations exactly and why are they non-projective?