LT2214 Computational Syntax
Student: Stas Hakkarainen

Report Card

English:

For this part of the assignment, a txt file with sentences in English was used. First, the file was tuened into a .conllu format so it would be possible to use the UD format for the rest of the assignment. This file was considered as a gold standard.

Afterwards, the following modifications were done:

1. gfud extract-pos-words: this command produced a conllu file that had a pos-tagged text (Generated file: “POS.conllu”).

2. gfud dbnf English.dbnf Utt: this command produced an utterance conllu file that also had UD trees and the weight of each sentence. Later on, this file was used to compare it to the gold standard that had been created before (Generated file: utteng.conllu)

3. gfud parse2pdf: this command produced a pdf file that showed graphical figures of sentences with dependency labels (Generated file: EngHuman.pdf).

The English txt file was also used to produce a conllu file by a computer to conduct an experiment if the computer was better at annotating the sentences that the user. The following part of the assignment produced the following files:

1. Eng_Test.conllu
2. ENGComp.pdf

Finnish:

For this part for the assignment, a txt file with sentences in Finnish was used. First, the file was converted into a .conllu format so that it would be possible to get the UD trees to use for the rest of the assignment. This file was considered a gold standard.

Afterwards, the following modification were done:

1. gfud extract-pos-words: this command produced a conllu file that had a post-tagged text. (Generated file: “posFI.conllu).

2. gfud dbnf Englishdbf Utt: this command produced an utterance conllu fiule that also had UD trees and the weight of each sentence. Later on, this file was used to comare it to the gold standard that had been created before (Generated file: “uttFI.conllu”).

3. gfud parse2pdf: this command produced a pdf file that showed graphical figures of sentences with dependency labels (Generated file: FiNHuman.pdf).

The Finnish txt file was also used to produce a conllu file by a computer to conduct an experiment if the computer was better at annotating the sentences that the user. The following part of the assignment produced the following files:

1. FiComp.pdf

Having evaluated the contents of the files, the initial results were low; therefore, a few modifications were introduced.

1. The gold standard conllu file was inspected and some UD trees were corrected to show the right structure and dependency. Afterwards, the evaluation was conducted again. This procedure was done both to the English and the Finnish versions. The following manipulation increased the number of the evaluation but not by a lot.

2. Several manipulations were made to the English.dbnf file so that it would be able to cover mor instances of the language structure in Finnish.

Results (English):

1. English before modifications:
   1.1. Micro: 0.3548387096774194
   1.2. Macro: 0.3612250672033281

2. English after modifications:
   2.2.Micro: 0.5384615037383681
   2.3.Macro: 0.5725146805235728

Results (Finnish);

1. Finnish before modifications:
   1.1. Micro: 0.3224637681159425
   1.2. Macro: 0.3174056743909683

2. Finnish after modifications:
   2.2. Micro: 0.5527428539471648
   2.3. Macro: 0.5772659823626524

Comparing the hand-written and the generated trees, it because possible to see that in some examples, the computer marked different words as roots; therefore, the tree itself looked a bit different. However, it is worth noticing that most of the time the human and the computer-generated trees looks either almost or completely the same. It is also worth mentioning that the same sentences were used in Finnish, and, in some instances, the computer picked different roots for those sentences.