Data-driven Dependency Parsing
Project part A, INF5830, Fall 2013

Deadline: October 30th

In this assignment you will perform dependency parsing experiments using the MaltParser software. We will work with the data sets released for the CoNLL-shared task on multilingual dependency parsing, which provides treebanks for several different languages. You will investigate the influence of various parameters of the parser, such as parse algorithm, machine learner, feature models, etc., on parse results and report on these results in a research report. Note that you will be performing experiments that may take a while to run so please start your work well ahead of the deadline.

The requirement for this assignment is to submit a written report of 2-3 pages which provides details on your experiments (described in section 3 below). The initial exercises in 1-2 below describe how to download and set up the software and data, and should not be a part of your report. The report should be submitted in Devilry before the deadline.

1. **Obtain and run MaltParser** MaltParser is freely available from `http://maltparser.org`. Follow the instructions from the Download-page and make sure you have a running version of the parser before you proceed. Familiarize yourself with the software and consult the online documentation when necessary.

   (a) Work through the *User guide* section entitled *Start using MaltParser*, where you train a (very small) parse model and apply it to the text taken from the Swedish treebank Talbanken05 that accompanies the software.

   (b) Examine the training and test data available in the examples folder. What do the different columns signify? Can you work out the dependency graph from this representation?

(c) Make use of the `info`-flowchart to obtain information about the parse model you trained. Which parse algorithm was employed? And which machine learner?

2. **Obtain the CoNLL-data** Download the data sets that are freely available from the CoNLLwebpage: `http://ilk.uvt.nl/conll`, i.e. the data sets for Danish, Dutch, Portuguese and Swedish. The data sets have been split into training and testing sets, make sure you obtain all splits for all languages.

3. **Parse experiments** *The following exercises should provide the basis for your project report. Please describe your experiments in a coherent text.*

   We will start out by training and evaluating parsers for each of the languages, employing MaltParser "out of the box", i.e., without doing any optimization of different settings.

   Evaluation should be performed using the evaluation script `eval.pl` which is available from the CoNLLwebpage, you should report both unlabeled and labeled accuracy. Please present your results in one or more table(s) and refer appropriately in the text (*Table X presents the results for . . . , we see from Table X that . . .*)

   (a) train a parser for each language and evaluate it on the corresponding test data. Training may take a while, so it might be worth looking into a command like `nohup`, if you are not already familiar with it, which will enable commands to keep running after you have logged out.

   (b) examine the output of the evaluation script: which dependency relations are difficult for the parser? how is this related to frequency in the training data?

   We will now go on to examine various aspects of parser performance and we will train parsers which differ from the baseline parsers trained in the previous assignment.

   For all of the below assignments you should compare the performance of the new parser(s) with the baseline parser and report the results. You should furthermore assess the statistical significance of any differences in parser performance using Dan Bikel's Randomized Parsing Evaluation Comparator (available from the course web page). This script (`compare.pl`) requires that you first run the evaluation script with the `-q -b` flags for each of the systems you wish to compare. These evaluation files are then input to

`compare.pl`. For these experiments, you may assume that p-values below 0.05 indicate a statistical significant differences.

Please choose *two* of the following assignments and report on your results.

(c) Parse algorithm: Investigate the various parsing algorithms available with the MaltParser software (consult the documentation).

   i. provide short descriptions of each algorithm
   ii. choose one language and evaluate the different algorithms
   iii. are there any (statistically significant) differences between the different parsers?
   iv. how does the choice of parse algorithm affect parsing time?

(d) Machine learner: MaltParser comes with two machine learning algorithms, which are both based on so-called Support Vector Machines: LIBSVM and LIBLINEAR. Investigate the effect of choosing one or the other on parse results

   i. provide short descriptions of the two machine learners
   ii. choose one language and evaluate the two learners
   iii. are there any (statistically significant) differences between the different parsers?
   iv. how does the choice of machine learner affect parsing time?

(e) Pseudo-projective parsing: some of the data sets contains non-projective structures. Investigate the effect of so-called pseudo-projective parsing for all the languages.

   i. employ the built-in facility for projectivization and deprojectivization in order to projectivize the input data and deprojectivize the output data
   ii. what is the effect of this process in terms of results?
   iii. are there differences between the languages?
   iv. are there any (statistically significant) differences between the baseline and the pseudo-projective parsers for each language?

(f) Feature model: the feature model provides a specification of the features deemed relevant for parsing. Investigate the effect of varying the feature model in terms of parse results.

   i. choose one language and investigate the effect of removing information of the word form from the feature model. How does this delexicalization affect results?

ii. examine the default feature model and consult the description of the feature model and the different feature functions in the User Guide at `http://maltparser.org` . Experiment with the addition of at least three new features to the feature model. How does this affect the results?

iii. the Danish data set contains information on morphology which is not currently employed in the default feature model. Experiment with the addition of this extra information to the feature model. What are the effects?

iv. are there any (statistically significant) differences between the different parsers?