

## Assignment 2

L715: Data manipulation for parser improvement

Due: Monday, October 10, 2011

You are going to be working with an annotation conversion tool, namely LTH ([http://nlp.cs.lth.se/software/treebank\\_converter/](http://nlp.cs.lth.se/software/treebank_converter/)).

1. (a) Split the WSJ corpus into sections 2-10 for training, 00 for development and 23 for testing
- (b) Convert these subcorpora into dependencies
- (c) Train both MaltParser and MSTParser

I want you to manipulate the different linguistic options, creating different dependency corpora, hypothesizing how they will affect parsing accuracy and then evaluating resulting accuracy. I encourage you to use something like MaltEval for evaluation (and visualization): <http://w3.msi.vxu.se/~jni/malteval/>

I expect an approximately two-page write-up, including quantitative and qualitative evaluation.