This is the README for Assignment 3 of Introduction to NLP (Fall 2019),

and contains some information regarding the data used for the assignment.

The datasets in Domain1 and Domain2 were taken from the Brown corpus section of the Penn Treebank.

The data was selected from different domains: non-fiction for the first domain,

and western and adventure fiction for the second domain; in order to prevent

training size bias, Domain1 was then adjusted to match the size of Domain2.

These datasets are not very large (~40000 tokens), so don't expect accuracies

nearing state-of-the-art, which often train on many millions of tokens.

The English Learner data is a lexicalized version of the TLE/UD-English-ESL (<https://github.com/UniversalDependencies/UD_English-ESL>) treebank v2.3.