

Week 4 : POS tagging with transfer learning and cross lingual learning

Work in groups of 2. Hand in a 1 page report (not counting the table) per group presenting your ideas, features and and results on French, English, Finnish and Estonian in a table.

1. Use your own CoNLL-U parser, accuracy measuring methods and so on.
2. On UD's website download the Finnish FTB and Estonian EDT treebanks.
3. Train POS tagging linear models (Perceptron) on French, English, Finnish and Estonian (one language at a time) and test them on each dev sets.
4. Train POS taggers on each pair of languages and test them on all four languages (keep separate scores though).
5. Repeat 3 and 4 but with a non linear model this time (KNN or Label propagation)

Ideas of features :

1. Simple target word based features,
2. + left and right neighbours,
3. + sentence level features,
4. Combinations of simple target word based features,
5. + left and right neighbours,
6. Cooccurence matrix reduction,
7. Feature matrix reduction ?