Assignment 5

L665/B659: Applying Machine Learning Techniques in CL; Sandra Kuebler

- For this assignment, you will go back to the English lexical sample data from the SensEval-3 competition.

- For the current assignment, we will also concentrate on 3 words out of the set: `arm.n`, `difficulty.n`, and `interest.n` and the same features and train/test sets as in assignment 3.

- But this time, we will use an SVM classifier, `svmlight`, or more specifically the multi class version: `http://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html`.

- This means, you need to convert your timbl features into a set of binary features.

- The data format for SVMs is the following: One line per instance. The first item is the class, then you get a sparse matrix, in which you list the non-negative features with their values. E.g., if word 2, 5, and 7 occur in the instance, you get the following SVM representation: class 2: 1 5:1 7:1

- Then train a classifier each for the 3 words and run it on your test sets.

- Report the timbl and SVM results. How do they differ?

- Work in groups of 2-3 people.

- **DUE DATE**: Monday, April 6 before class. Please submit your report via canvas.