Exercises for 19.12.2017

**Exercise 1. Induction of verb classes**

Step 1. Open the file tenseaspect.txt in R as a data frame.

Step 2. Swap the rows and columns.

Step 3. Create a correlation matrix of individual verbs based on their frequencies of occurrence in specific tenses. Use Kendall’s tau.

Step 4. Make a distance matrix by subtracting the correlation coefficients from 1.

Step 5. Perform hierarchical cluster analysis, compare different methods.

Step 6. Interpret the results.

**Exercise 2. Dutch causatives**

Step 1. What is the difference in the relative frequencies of causative verbs *doen* and *laten* in the Dutch and Belgian corpora (dataset *doenLaten*)? Make a bar plot. Compute the proportions and odds ratio. Is the difference statistically significant?

Step 2. Create a conditional inference tree with Aux as the response variable and Country, Causation and EPTrans as predictors. Interpret the results.

Step 3. Grow a random forest with 500 trees and investigate the conditional variable importance.

Step 4. Compare the accuracy of predictions by the individual tree and by the forest. Which of the methods provides the better fit for the data?