

## Birmingham Corpus Linguistics Summer School Overview

- More complicated and less inspiring than the one in Lancaster
- Programming with R (R-Studio)
- Cluster Analysis - family of techniques for classifying objects into groups based on a set of measurements. The goal is to group objects into clusters so that objects in a cluster are similar to each other and different to objects in other clusters.
- Binary Logistic regressions:
- What do you do with binary logistic regression?
- you analyze a case of binary variation: one binary dependent variable, several explanatory factors.
- What kind of data?
- A data set of examples of language use where speakers have a choice between A and B. The examples are annotated for a range of explanatory factors
- What does the analysis do with that data?
- It calculates how well the data can be classified into A and B on the basis of the explanatory factors. What factors are more important than others? Which ones can be left out? How much variability is explained?
- Corpus annotation and analysis with the UAM Corpus Tool – one of the brightest sessions.
- Basic Concordancing by Michael Barlow – Vaclav's teacher, a potential partner for us.
- Concordance shows instances of language. What is the relation between those instances and (a) grammar (b) social interactions.
- When comparing two corpora uses log-likelihood values (the more the better).
- Paul Thomson – customized tagging using Notepad++. Editor of Journal of English for academic purposes: the corpora are not comparable, use the percentage.