

# Coping with variation in the Icelandic Diachronic Treebank

**Eiríkur Rögnvaldsson, Anton Karl Ingason and Einar Freyr Sigurðsson**

**University of Iceland**

{eirikur, antoni, einasig}@hi.is

At the University of Iceland, work on a diachronic treebank for Icelandic is just commencing. This is a project which is partly funded by the Icelandic Research Fund and the University of Iceland Research Fund. The main emphasis will be on modern Icelandic, but the treebank will also contain texts from all periods in the history of Icelandic. The treebank will be made in close cooperation with Professor Anthony Kroch and his research team at the University of Pennsylvania. We will build on the annotation scheme for the Penn Parsed Corpora of Historical English (<http://www.ling.upenn.edu/hist-corpora/>) which we will adapt as necessary to deal with the morphology and syntax of all stages of Icelandic. We will use the open source CorpusSearch software (<http://corpussearch.sourceforge.net/index.html>) which was designed to use with the English historical corpora, and adapt it to the Icelandic annotation scheme.

Since neither the morphology nor the syntax of Icelandic has undergone dramatic changes in the history of the language, it should be possible to combine texts from all stages of the language in one treebank, and search all these texts simultaneously in order to study language variation and change. Crucially, this means that both our annotation scheme and our search software must be able to deal with considerable variation. For instance, we need to be able to search the whole corpus for syntactic constructions which occur at different stages in the language, even though the actual manifestation of these constructions may change over time. This might mean that we will have to develop an extra abstract layer of annotation. Our research in the coming months will focus on finding solutions which enable us to carry out diverse and efficient searches having different purposes.

We will work on the annotation scheme and the search software during the summer. At the workshop in September, we hope to be able to present a preliminary version of our annotation scheme, together with a modified version of the CorpusSearch software.