Kristín Bjarnadóttir
The Árni Magnússon Institute for Icelandic Studies, Reykjavík
kristinb@hi.is

The Case for Normalization: Linking lexicographic resources for Icelandic

The purpose of this project is linking data from two longstanding lexicographic projects on Icelandic, i.e., the Dictionary of Old Norse Prose (ONP) in Copenhagen, and the work on a historical dictionary of Icelandic (as manifested in Written Language Archive (WLA)), in Reykjavik. The main obstruction is a difference in spelling, but lemmatization also plays a part. Linking these lexicographic resources would be of great benefit, both in lexicography, and in language research in general, as the data created would facilitate access to words and word forms, irrespective of variants.

The ONP and the WLA were started in 1939 and 1944, respectively, and they are independent of each other, both in scope and methods. The line of demarcation between the projects is 1540, i.e., the year of publication of the 1st printed book in Icelandic. The split into two periods at that point is unwarranted, as far as the cohesion of Icelandic throughout the history of the language is concerned. The changes in morphology and spelling of Icelandic are regular enough to make normalization with a spell checking application based on a noisy channel model possible (Jón Friðrik Daðason et al. 2014). The case for using Modern Icelandic as a basis of the normalization rests on the standardization of the modern language, and on the use of existent LT tools and data.

The immediate result could be a flow of information between ONP and WLA. A simple case in point is the dating of individual lemmas, such as the 14th C dagdrykkja (‘day drinking’) from ONP (Orkneyinga saga, 1387-1395), dated 1980 in WLA. The data from both ONP and WLA can also serve as training material for further work on the automatic normalization of Icelandic texts from all periods.

Bibliography

Websites
Ordbog over det norrøne prosasprog: http://onp.ku.dk/
The Written Language Archive: http://www.arnastofnun.is/page/gagnasofn_ritalssafn