One for all

How to transcribe Icelandic manuscripts for scientific research

Ludger Zeevaert

The Icelandic manuscripts are an indispensable source for the investigation of Icelandic language and literature. However, their scientific exploitation still poses a challenge for the application of methods typically used in digital humanities. Some peculiarities of pre-modern Icelandic writing make a diligent transcription rather time-consuming, and optical character recognition is not yet ready for use with medieval or early modern Icelandic script. Obstacles for digital methods are for example allographic variation (round/long s, r and r-rotunda, f and insular f etc.) and the, in comparison to continental manuscripts, excessive use of abbreviations.

Different research questions call for different types of text-representation. For historical, literary or folklore studies normalised text editions (or English translations) may be sufficient, but linguistic, philological or paleographic studies are dependent on a more accurate rendering of the manuscript texts (see Zeevaert 2014). The compilation of an automatically lemmatised corpus requires orthographically normalised transcriptions, whereas morphological and phonological variation can only be studied on a more manuscript-near representation of the spelling.

To avoid the multiple manual transcribing of the same manuscript on different levels of accuracy it would be advantageous if different representations of one text needed for different research questions could be produced automatically from one transcription that can be prepared with an affordable amount of time. The optimal approach would go out from a maximally accurate transcription that could then be downgraded step by step. In a project with limited financial and temporal resources, however, this is not feasible due to the amount of work and time needed for such a transcription. In the project "Gullskinna. Postmedieval transmission and reception of a lost medieval parchment-codex" we therefore go out from a semi-diplomatic transcription of different manuscripts of Njáls saga which can be used as a starting point for an electronic edition of the text. To tackle linguistic questions morphological and syntactic annotations are added.

Although a more detailed level of transcription is not needed for our research it would be useful for other scholars to have at their disposal a transcription that renders allographic variation and abbreviation signs. Such a type-facsimile transcription is for example favoured by the Medieval nordic text archive (www.menota.org). We therefore decided to find out whether our semi-diplomatic transcription could be used to compile a type-facsimile transcription automatically. Our approach is based on the consideration that the medieval and early modern Icelandic writing system is, in difference to the modern Scandinavian practice,
not simply an alphabetic script but a combination of a logographic, a syllabic and an alphabetic system (for examples see figures 1-3).

This system is not completely random but based on rules that can be shown on manuscripts like *AM 136 fol.* from around 1640, the main manuscript in the *Gullskinna*-project:

![Fig. 1](image1.png): Example for a logogram from manuscript *AM 136 fol.*: the abbreviation for því að 'because'

![Fig. 2](image2.png): Example for the abbreviation of combinations of letters from manuscript *AM 136 fol.*: the abbreviation for -ur in the word suður (ſudur) 'south' (in the manuscript no distinction is made between <ð> and <d>).

![Fig. 3](image3.png): Example for allographs from manuscript *AM 136 fol.*: different s-forms in different positions of the word sins (ſyn3) 'his/her' (the dashes above the n are part of the letter y).

Our semi-diplomatic transcription follows the orthography in the manuscript but expands abbreviations and reduces the allographic variation to a certain amount. In a first step words that are represented by logograms in the manuscript have to be replaced on the basis of a list with logograms and their expansions. In a second step letters and combinations of letters rendered by superscript abbreviation signs have to be transferred to their original shape. Expanded abbreviations are marked with a certain tag in the transcription, and even here a list of letter-combinations and their abbreviations has to be used. In a third step certain graphemes at certain positions in the word have to be replaced with the allographs used in the manuscript. This has to be done according to certain rules.

Literature:

AM 136 fol. (*Njáls saga*). Stofnun Árna Magnússonar í íslenskum fræðum, Handritasvið, Safni Árna Magnússonar, Reykjavík.
