Mörkum Njálu! An Annotated Corpus to Analyse and Explain Grammatical Divergences between 14th-Century Manuscripts of *Njáls saga*.

Ludger Zeevaert
Stofnun Árna Magnússonar í íslenskum fræðum
Árnagarði við Suðurgötu, 101 Reykjavík, Iceland
E-mail: ludger@zeevaert.de

Abstract

The work of the research project "Variance of Njáls saga" at the Árni Magnússon Institute for Icelandic Studies in Reykjavík relies mainly on an annotated XML-corpus of manuscripts of *Brennu-Njáls saga* or 'The Story of Burnt Njál', an Icelandic prose narrative from the end of the 13th century. One part of the project is devoted to linguistic variation in the earliest transmission of the text in parchment manuscripts and fragments from the 14th century. The following article gives a short overview over the design of the corpus that has to serve quite different purposes from palaeographic over stemmatological to literary research. I will focus on features important for the analysis of certain linguistic variables and the challenge lying in their implementation in a corpus consisting of close transcriptions of medieval manuscripts and give examples for the use of the corpus for linguistic research in the frame of the project that mainly consists of the analysis of different grammatical/syntactic constructions that are often referred to in connection with stylistic research (narrative inversion, historical present tense, indirect-speech constructions).

Keywords: Old Icelandic, linguistic annotation, historical corpora

1. Background of the Project

With an extent of ca 100000 words, *Njáls saga* is not only the longest, but also the most favoured of the Icelandic family sagas, a fact that is documented in the unusually large number of 18 medieval and 45 post-medieval manuscripts. *Njáls saga* is the subject of a research project located at the Árni Magnússon Institute for Icelandic Studies in Reykjavík. The project *The variation of Njáls saga (Breytileiki Njáls sögu)* aims at

a) a revision of the stemma of the manuscripts which was outlined by Einar Ólafur Sveinsson some 60 years ago and
b) an investigation of variation in the manuscripts from different scientific angles (material philology, linguistics, stylistics, literary studies).

Part of this research is the investigation of synchronic linguistic variation in the oldest manuscripts of the saga from the beginning of the 14th century. In the following I would like to give a short outline of how the corpus is compiled and prepared for linguistic research and describe our methods for the analysis and explanation of linguistic variation in our corpus.

2. The Project Corpus

2.1 Extent and Levels of Transcription

The corpus of the project consists of XML-transcriptions of the oldest fragments that cover about half of the text of the saga and the corresponding parts in the eight extant parchment codices from the 14th and 15th century.[1] It contains a total of ca 400000 words. The manuscript texts are transcribed in three parallel versions or levels (<facs>, a type-facsimile transcription, <dipl>, a diplomatic transcription and <norm>, a normalised transcription). This approach is suggested in the guidelines for transcriptions for the Medieval Nordic Text Archive (www.menota.org), an established standard for digital publishing of Scandinavian texts from the Middle Ages. The different transcription levels make it possible to use the corpus for a variety of tasks.

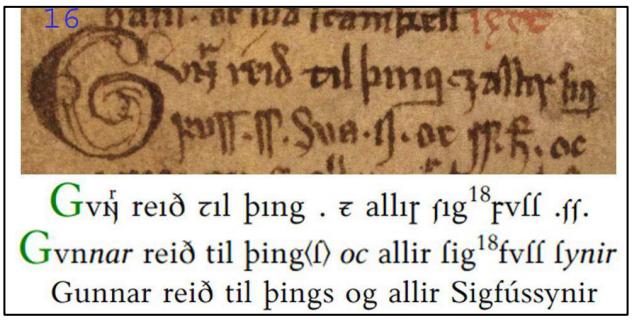

Figure 1: 'Gunnar and all the Sigfussons rode to the Thing', beginning of ch. 51 of *Njáls saga* in ms. AM 162 B fol. δ, type-facsimile -, diplomatic - and normalised transcription.

The type-facsimile transcription, i.e. a transcription of the manuscript text reproducing letter forms, abbreviation signs, line and page breaks etc., was e.g. successfully used for a systematic description of the abbreviation system of the *Njáls saga* manuscript AM 162 B fol. δ (cf. Zeevaert, 2013b), the normalised level is useful e.g. for machine-based collations of manuscripts (cf. Zeevaert, 2013a). It is already freed from orthographical variation which is irrelevant for stemmatological questions but an obstacle for collation software which has difficulties to distinguish between stemmatologically important and unimportant variants.

[1] Currently the corpus is extended to 17th century paper manuscripts of the saga, cf. Zeevaert (in prep.).

Corpus linguistics aims at working with large corpora that can be assumed to be representative for an entirety of texts in a certain language. For several reasons, however, the corpus of our project is only a far cry from standard corpora of modern languages. Not only is the amount of Medieval Icelandic texts rather limited (less than 1000 manuscripts, most of them only fragments containing not more than a few leaves, and many of them containing the same texts, cf. Jónsson 2003: 12), but also the representativeness of the surviving texts for the Old Icelandic language as a whole.
In addition to this, our corpus is not designed especially for an analysis of the Icelandic language at a certain time or of linguistic developments in certain periods. Rögnvaldsson & Helgadóttir (2011, pp. 67ff.) and Rögnvaldsson et al. (2012) describe tagged corpora of Old Icelandic that were designed for such tasks. *The Old Icelandic Corpus* contains 1651398 words (tokens) and uses mainly texts from editions in Modern Icelandic spelling. It was successfully applied in research on syntactic changes from Old to Modern Icelandic (cf. Rögnvaldsson & Helgadóttir, 2011, pp. 70ff.). The *Sögulegi íslenski trjábankinn (IcePaHC)* contains ca 375000 words (tokens) from medieval Icelandic texts in Modern Icelandic spelling (and additionally ca 625000 words from post-medieval texts, cf. Rögnvaldsson et al. 2012: 335ff.).
The editions used for *The Old Icelandic Corpus* and *IcePaHC* are based on rather reliable transcriptions of one manuscript that usually aim at a general public and therefore give the text in Modern Icelandic spelling, do not give variants from other manuscripts, expand abbreviations tacitly and correct errors. However, for the tasks of the *Njáls saga* project that include not only a detailed study of certain aspects of Icelandic prose style in the 14th century, but also a revision of the textual relationships between the 63 different manuscripts and eventually the preparation of a critical edition of the saga, it is necessary to account for the text in the different manuscripts more thoroughly.

2.2 Transcription of Manuscripts

In most cases the oldest surviving manuscripts of medieval texts are esteemed to be the most valuable sources for philological and historical linguistic questions, but they are at the same time usually the most difficult ones to decipher (cf. Figure 2). We nevertheless decided against a restriction of the corpus to easier accessible sources because this would have required a modification of our research questions. It goes without saying that the preparation of a corpus built on manuscript transcriptions requires many times the amount of time and work needed for a corpus put together from electronic versions of texts or from sources that can be digitised with the help of OCR.
Due to the differences in the preservation of the single manuscripts it is difficult to make exact calculations of the amount of time needed for the transcription of the whole corpus. For a complete transcription from scratch on all three levels with a complete annotation according to the Menota-guidelines one word per minute is from our experience a quite realistic calculation of the average speed, although parts of manuscripts like *Oddabók* (AM 466 4to) or the fragment AM 162 B fol. β (cf. Figure 2) slow down the process considerably.

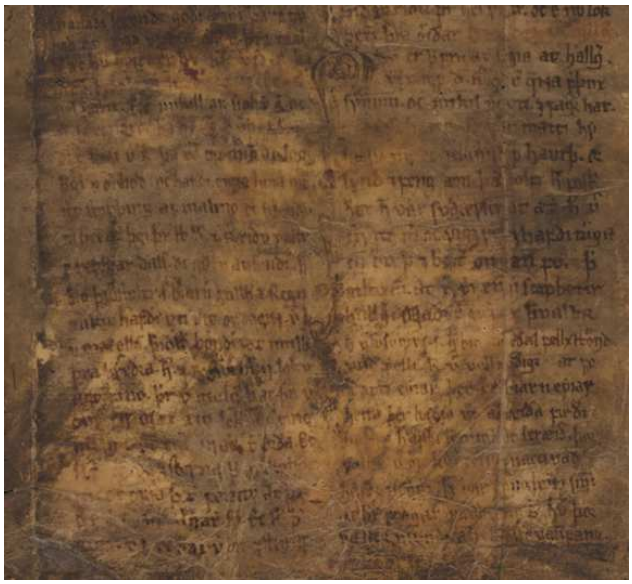

Figure 2: Fragment AM 162 B fol. β (ca 1300)

Under this circumstances a complete transcription and annotation of the 14th-century corpus during the funding period would require the complete working time available in the project with no time left for the analyses. Thus a sensible use of work power in our project is even more crucial than in projects dealing with less labour-intensive ways of text input.
One way to achieve this objective is to avoid unnecessary repetitions of working steps. By basing the transcriptions on a prefabricated document containing part of the XML-mark up and a normalised transcription of one of the main manuscripts the amount of time needed for the transcription could be reduced considerably. We were also able to profit from work done by students in the frame of transcription courses and final theses.
Given the limited amount of extant Old Icelandic sources it is quite natural that different approaches to research on Old Icelandic language, literature and culture rely by and large on the same corpus of texts. It thus seems reasonable to aim at a corpus design that would be able to serve the various needs of these different approaches.[2]

2.3 Successive Enrichment

It is not necessary, though, to implement all features that might be useful for all thinkable kinds of future research before the corpus can be used as long as the structure allows for a successive enrichment of the information contained in the XML-files.

[2] Cf. Sahle (2013: 133) who argues for multidisciplinarity as one key feature of digital editing and questions the necessity to produce different types of editions for different research interests: "Es gibt keinen vernünftigen Grund, warum z.B. eine mittelalterliche volkssprachige Urkunde einmal für die Belange der historischen Sprachforschung und einmal für die allgemeine Geschichtsforschung ediert werden sollte."

E.g. a stemma, i.e. a (graphic) reconstruction of the relations between the different manuscripts (exemplars and copies) in the form of a family tree, can be retrieved with rather basic information.

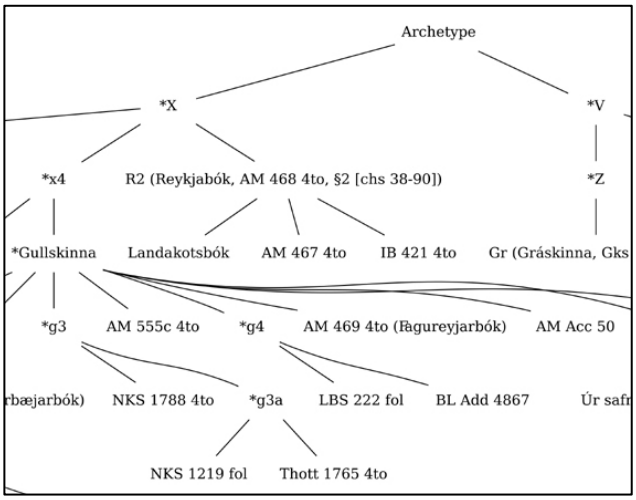

Figure 3: Part of a stemma of chapter 86 of *Njáls saga*

Stemmatological work builds mainly on an evaluation of (intentional or unintentional) changes of text by a scribe copying a manuscript. Changes in the orthography are considered to be less important (a 17th century scribe copying a 14th century manuscript would usually adopt the spelling to his own scribal habits), a diplomatic transcription together with a segmentation into comparable textual units but without any further annotation would suffice to produce the basis for a calculation and evaluation of differences between the manuscripts and the construction of a stemma.[3] This means that part of the work in the project can be performed without more detailed levels of transcription and grammatical mark up.

The normalisation in the project is based on the modern Icelandic norm. This creates a certain distance from the language of the older manuscripts, although the distance between contemporary and medieval Icelandic is, compared to other European languages, rather small. What speaks for the modern norm is the easy applicability and complete documentation with dictionaries and grammars which helps to avoid errors and inconsistencies. In addition to this, historical normalisations, e.g. the one used for the Old Norse dictionary ONP (*Ordbog over det norrøne prosasprog*, onp.ku.dk), usually render the language from around 1200, that is 100 years before the earliest and more than 600 years before the youngest manuscripts of *Njáls saga* were written down.

2.4 Transcription Conventions

Most appropriate for linguistic analysis is a diplomatic transcription that expands abbreviations, corrects errors and normalises allographs without phonological value.[4]

For a linguistic analysis it is necessary, though, to identify both corrections and expansions with suitable tags that e.g. enable a typographical differentiation in the output (e.g. by using angle brackets and italics, cf. Figure 1). Not in all cases abbreviations can be expanded without ambiguity. The most striking examples are abbreviated forms of verbs that are often ambiguous with respect to tense, and analyses that include grammatical tense, e.g. comparisons of the use of historical present tense in different manuscripts, have to be able to exclude such examples (cf. paragraph 3.3 below).

Also the distinction between unintentional scribal errors and intentional stylistic variants is not always unambiguous. As an example, grammatical agreement or the use of non-finite verb forms in medieval manuscripts do in many cases not follow the rules of Modern Icelandic grammar, and what seems to be a grammatical error from a modern point of view was in many cases obviously correct language use from the point of view of a 14th-century scribe.

3. Linguistic Variation in 14th-Century Manuscripts

3.1 Tagging Linguistic Variables

Linguistic research is only one among several research focuses of the *Njáls saga* project but of particular interest in the frame of this publication.

Typical linguistic variables that showed up in comparisons of smaller textual units of the earliest manuscript fragments of *Njáls saga* and are thus of larger interest for research on synchronic variation were grammatical/stylistic features like the position of the finite verb (verb-first or verb-second order), the order of noun and modifier (modifier before noun or noun before modifier), but also other grammatical phenomena such as pronominal reference, definiteness, agreement, the use of present tense vs. past tense in the narrative parts of the saga (historical present tense) or the use of AcI-constructions or conjunctional subordinate clauses in indirect speech.[5]

To enable a systematic analysis of these variables, a mark up of certain grammatical information has to be added to the XML-transcriptions. It consists of two components: a lemmatisation and part-of-speech tagging and a tagging of syntactic entities.

The limited resources of the project do not allow for a

[3] Cf. Hall et al. (in prep.) for a description of a stemmatological approach built on normalised transcriptions of one chapter of *Njáls saga*. This approach builts not least on scibal errors and divergent spellings of placenames common to multiple manuscripts so that a complete normalisation of the text is not a viable option.

[4] The conventions applied in the project differ partly from those common in printed editions of Old Norse texts. The edition practice of the manuscript institutes in Reykjavík and Copenhagen renders partly the, now obsolete, technical restrictions of metal type printing which also influences the distinction of letter forms, and generally the question which letter forms have to be considered as phonologically distinctive is not at all undisputed, cf. Zeevaert (2013a).

[5] Phonological/morphological variation is of certain interest for the study of diachronic linguistic variation due to the, in comparison to other linguistic features, rather pronounced changes in this domain from Old to Modern Icelandic. In the frame of the *Njáls saga* project diachronic change is investigated by Haraldur Bernharðsson.

complete morphological and syntactical tagging of the whole corpus. For an investigation of the above-mentioned variables, a tagging of relevant parts of speech (nouns, adjectives, finite verbs, conjunctions) and a partial morphosyntactic annotation is sufficient (cf. Zeevaert, 2008 for an application of this method to subordinate clause word order).

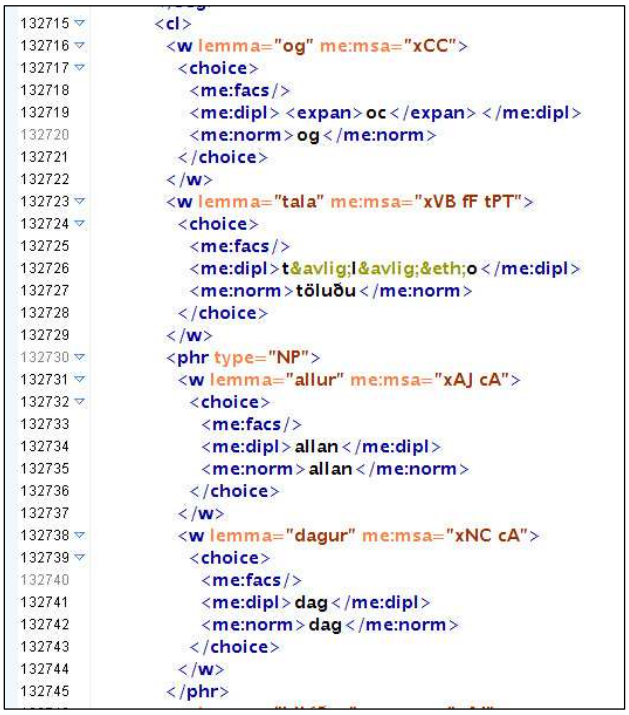

Figure 4: Transcription of GKS 2870 4to ...

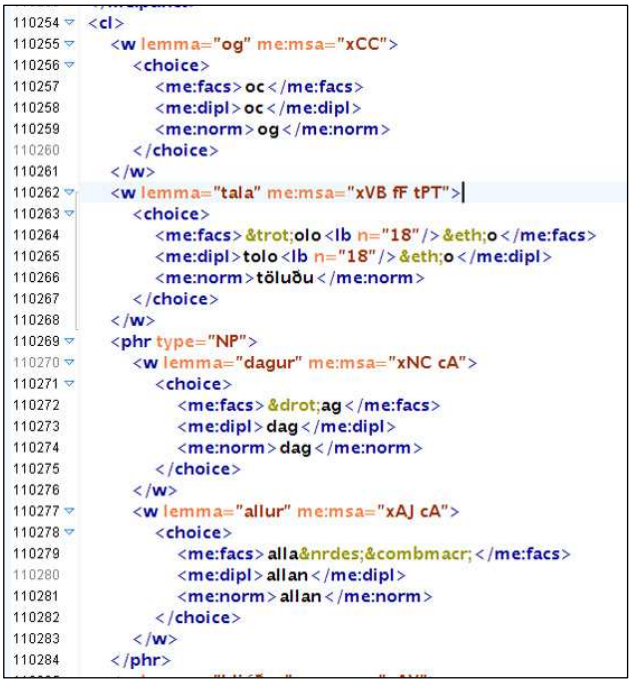

Figure 5: ... and AM 162 B fol. δ (both ca 1300) with different noun-phrase word order

An extension of the tagging for further research questions is unproblematic and can be based on the already realised tagging at a later stage.[6]

For the variables involving word order a tagging of syntactical units (phrases and clauses) is needed in addition to the POS-tagging. To determine the word order in NPs, information about case of nouns has to be added to distinguish between heads and modifiers. Subordinate-clause constructions in indirect speech (conjunctional clauses or accusatives with infinitive) need a marking of clause-type and a grammatical tagging of the parts of speech decisive for the construction (finiteness, mode).

To find instances of historical present tense a tagging of the tense of verbs is self-evident, but also a marking of direct speech to distinguish instances of historical present tense in the narrative parts of the text from (non-historical) present tense used in dialogues is necessary.

Currently transcriptions of all eight existing 14th-century fragments (AM 162 B fol. β, γ, δ, ε , ζ, η, ϑ and κ) and the corresponding parts in three of the five 14th-century codices (AM 468 4to, AM 133 fol., GKS 2868 4to) are finished as well as larger parts of the remaining two 14th-century codices (GKS 2870 4to, AM 132 fol.). A grammatical annotation that is suitable for analyses of word order, use of tense and indirect speech constructions was implemented in five chapters of the saga in six of the fragments and four of the codices.

The resources of the project do not allow for tailor-made complex software-solutions for an automatic analysis of the transcriptions. We therefore opted for a simple low-cost solution with maximal efficiency, i.e. XSLT-style sheets containing XPath-expressions that can be used to find, count and display the above mentioned structures in the manuscript transcriptions. However, to be able to make clear statements about variation in manuscripts with regard to these structures it is necessary to find corresponding chunks of text in manuscripts that do not exhibit the structure in question. We decided to implement a serially numbered segmentation common to all transcriptions which is based on the smallest self-contained textual unit, i.e. the sentence, in one of the main text-witnesses, *Reykjabók* (AM 486 4to). A similar system (chapter and verse) is used very successfully to identify corresponding textual units in different versions of the Bible. In cases where transpositions, omissions and additions of text change the order of the segments in different text-witnesses, corresponding sentences can still be identified by means of the segment numbers.

3.2 Word Order

Some of the above mentioned typical linguistic variables that were identified by a comparison of parts of the text in different 14th-century manuscripts play an important part in descriptions of a typical Icelandic saga style (cf. e.g. Hauksson & Óskarsson 1994, pp. 273ff.). This accounts especially for word order variation and historical present tense.

In the context of Old Scandinavian texts the order of constituents is interesting for two reasons: Constructions

[6] The technical requirements for an automatic tagging of TEI-XML-transcriptions of medieval Icelandic manuscripts are not yet at hand, but generally the use of POS-taggers optimised for Old Icelandic seems to be a promising option (cf. Zeevaert in prep.).

like the verb-initial word order in declarative sentences (narrative inversion) are used to determine the style and thereby the author of single texts (cf. Hallberg, 1968). Word-order patterns, e.g. the position of head and modifier in the noun phrase, are used to document long-term developments in the change of the Scandinavian languages from OV- to VO-languages (cf. Zeevaert, 2012).
Before the tagging of NPs in our corpus is finished, reliable quantitative analyses of the word order in the noun phrase are not possible. Preliminary observations on the already tagged material indicate that the order of noun and modifier is rather stable between most manuscripts. However, single manuscripts like AM 162 B fol. δ from around 1300 seem to deviate in a quite systematic way and prefer the order noun-modifier in contrast to modifier-noun in other manuscripts from about the same time (c.f. example (3) below).
A comparison of five chapters in seven 14th-century manuscripts from our corpus resulted in considerable differences in the use of narrative inversion. The counting was based on the output from a search for clause-initial finite verbs ("//s[.//cl[*[1]=w[contains(@me:msa, 'xVB fF')]]]"), non declarative sentences (interrogative and imperative sentences) were disregarded.
The largest difference for the same chapter was found between two mss. that were both written around 1300, AM 162 B fol. β (narrative inversion in 5% of the sentences) and GKS 2870 (*Gráskinna*, narrative inversion in 12.5% of the sentences). If the single examples are compared the differences in the use of this stylistic device between individual scribes are even more striking: Only 26,5% of the sentences exhibiting narrative inversion in at least one of the manuscripts in the corpus did so in all manuscripts, in the remaining 73,5% of the cases at least one manuscript had the unmarked verb-second word order. From our point of view this emphasises that both stylistic surveys of Old Icelandic texts and research on typological change would profit from research based on a comprehensive quantitative evaluation of all manuscripts of a text.

3.3 Historical Present Tense

Previous quantitative approaches to the frequency of historical present tense in Icelandic family sagas[7] show huge differences between different sagas. Sprenger (1951, p. 48) in a counting based on a normalised edition found 60% historical present tense in the Icelandic family saga *Eyrbyggja*, but a considerably lower percentage in younger sagas (although she does not quantify the difference). Hallberg (1968, p. 207) found between 3.2% and 78% instances of historical present tense in 40 Icelandic family sagas. Torgilstveit (2001, pp. 78-79), who examines three manuscripts of the sagas of Norwegian kings, found between 3% and 50% usage of historical present tense in the same part of the text in the different manuscripts.
Sprenger and Torgilstveit explain the huge differences in their material with a development of saga style over time; Hallberg's results do not show a correlation between estimated age of a text and the use of tenses, he thus assumes that the individual style of the authors is responsible for the observed variation.
There are, however, strong indications to methodological problems in Sprenger's and Hallberg's approaches. Hallberg, using the same edition of *Eyrbyggja saga* as Sprenger (Sveinsson & Þórðarson Eds., 1935), counts only 3.2% of historical present tense, an astonishingly large deviation from Sprenger's 60%. We therefore decided to examine the use of tenses in a limited sample from our corpus.
A complete evaluation of the corpus is impeded by the fact that the oldest manuscript fragments of *Njáls saga* cover mostly different parts of the text (altogether the 14th-century fragments cover 77 chapters of the saga, but only 27 chapters are represented in more than one ms. and only 4 chapters in at least three mss.). Research on historical present tense (HPT) in narratives (cf. Thoma, 2011) was able to show that its use is at least partly dependent on discourse function (for a different hypothesis in the framework of markedness theory cf. Torgilstveit, 2007). This means that an equal distribution over the different chapters of the saga cannot be expected.
By analysing 15 chapters it is possible to at least indirectly compare all 14th-century manuscript witnesses and get a fairly representative picture of differences in the usage of tenses between them.
Verbs in the present tense are identified by the corresponding POS-tag (with the attribute "me:msa" in the Menota-conventions, "xVB" stands for 'verb', "fF" for 'finite' and "tPS" for 'present tense'):

```
<w lemma="ríða" me:msa="xVB fF tPS">
  <choice>
    <me:facs>r&inodotsup;<unclear>&eth;</unclear>&rrot;</me:facs>
    <me:dipl>ri<unclear>&eth;</unclear>r</me:dipl>
    <me:norm>ríður</me:norm>
  </choice>
</w>
```

Figure 6: Tagging of tense

Parts of this occurence of the verb *ríða* are illegible in the manuscript but the readable part suffices to determine the tense (otherwise the value of the attribute would have been "tU" for 'tense unknown').

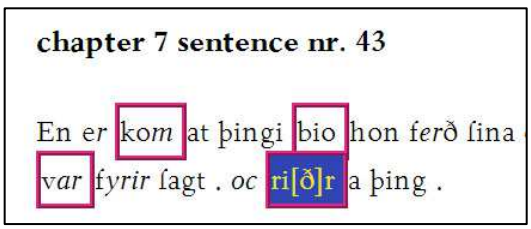

Figure 7: HTML-display of the example in Figure 6.

With an XPath expression

```
//s[.//cl//w[contains(@me:msa,'tPS') and count(ancestor::q)=0]]
```

[7] Icelandic family sagas (*Íslendingasögur*), in distinction to other types of sagas, is the name for a group of prose narratives composed in the 13th and 14th century and describing mainly events from the time of the Icelandic free state (930-1262). The Family sagas are widely assumed to be a genuinely Icelandic literary product (in contrast to e.g. Chivalric sagas or Saints' sagas that depend largely on foreign originals) and are therefore of special interest for historical linguistic research.

all instances of present-tense verbs outside of clauses containing direct speech (<q>) can be found and, if integrated in an appropriate style sheet, counted, transformed to HTML, PDF or a text format and displayed (e.g. with dark background colour to distinguish them from verbs in past tense, cf. Figure 7) together with their sentence number.
Verbs contained in clauses rendering direct speech (the part of the sentence inside the quotations marks) and verbs where the tense is indeterminable can be excluded from the output or assigned a different appearance, cf. Figure 8. As the angle brackets indicate, the verb form *for* 'went' (from *fara* 'go') is not recognisable in the text beyond any doubt, and the verb form *mællti* 'spoke' (from *mæla* 'speak') is truncated after the *m*, i.e. the italicised part which contains the grammatical information about tense is not present in the manuscript and has to be added by the reader. For the transcription past tense was assumed to be the unmarked tense, but for an analysis of tense such examples have to be excluded.

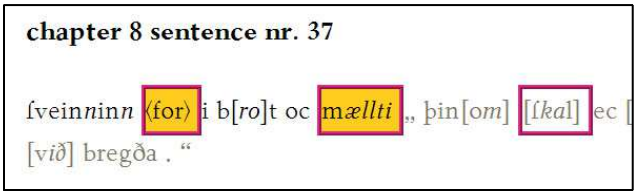

Figure 8: HTML-display of disregarded verb forms

A study of five chapters from nine different manuscripts that was conducted in our project gave between 2.38% and 14.17% instances of historical present tense. All manuscripts are dated to the first half of the 14th century. A chronological explanation for the differences between the manuscripts is therefore unlikely.

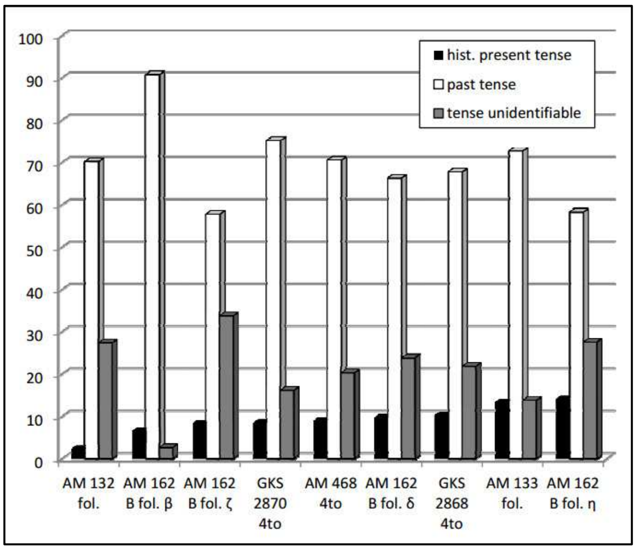

Figure 9: Distribution of HPT in 9 different mss. (PS=present tense, PT= past tense, U=tense unidentifiable)

Hallberg´s (1968) results for instances of historical present tense in *Njáls saga* (13%) come up to the results from the manuscripts with the highest amount of HPT in our corpus. What comes as a surprise is that the manuscript used for the edition utilised by Hallberg, AM 132 fol. (*Möðruvallabók*, Sveinsson Ed., 1954) is the one with the lowest amount in our sample (2.38%), a quite considerable deviation from Hallberg's results.
Two factors seem to be mainly responsible for the discrepancies between the different countings:

a) Hallberg's study was only based on a sample of verbs he assumed to be representative for Icelandic verbs as a whole.
b) As in the following example (1) from *Möðruvallabók* (AM 132 fol.), frequent verbs are very often truncated in the manuscripts and do then not allow for a determination of tense (ſ. G. = *segir* (say-3PRS) *Gunnarr* or *sagði* (say-3PST) *Gunnarr*):

(1) hůgı mun ek ꝼ͛ ſ. G. ⁊ ſua vıllða ek at þu g͛ðer. (AM 132 fol. ca 1330-1370, fol. 27r[8])

In the normalised editions used by Sprenger and Hallberg abbreviations are silently expanded:

(2) „Hvergi mun ek fara," segir Gunnarr, „ok svá vilda ek, at þú gerðir." (Sveinsson Ed., 1954, p. 183)

"I will not go away any whither," said Gunnar, "and so I would thou shouldest do too." (Dasent Trans., 1971, p. 131)

In addition to this, a considerable amount of variation in the use of tenses can be found between manuscripts from the same time:

(3) NU eggiar (egg on-3SG.PRS) St*arkaðr* ſina m*enn* (AM 162 B fol. δ, ca 1300, fol. 11 v)
⟨S⟩iþan egiaði (egg on-3SG.PST) ſtark*aðr* m*enn* ſina (GKS 2870 4to, ca 1300, fol. 40 r)
After that Starkad egged on his men, ... (Dasent Trans., 1971, p. 111)

This means that the use of tenses as it is found in a normalised edition is neither representative for the language of a certain period in language history nor for the individual style of a certain author or scribe but is heavily influenced by the stylistic preferences of the 20th-century editors. With a corpus consisting of (strictly) diplomatic transcriptions of manuscripts those problems can be avoided.

4. Conclusion

In this article I gave a short description of a corpus of the earliest text transmission of *Njáls saga*, an Old Icelandic prose narrative composed shortly before 1300.
The corpus is built from TEI-XML-transcriptions of the manuscripts. The TEI-XML-format facilitates on the hand to enrich the transcriptions with additional information successively (different levels of transcription, codicological, semantic, morphological, syntactic etc. information) and on the other hand to filter this information for certain tasks (displaying only the type-facsimile level for palaeographic research or only certain syntactical structures for linguistic analyses etc.).
The XML-format makes it possible to add all kinds of

[8] Folio 27 recto, i.e. the front side of folio 27.

information to the transcription and thus to extend its usability to a variety of research questions. The corpus is not primarily designed for linguistic research, and in difference to existing digital corpora of medieval Icelandic texts it is not targeting an exhaustive overview over a certain period in language history or general historical developments in Icelandic but a detailed study of variation between single contemporaneous linguistic sources.

The corpus is not capable of replacing such larger corpora built on easier accessible sources. However, the first analyses of grammatical/stylistic structures show considerable differences between contemporaneous manuscripts, and we are convinced that a more prominent consideration of the manuscript tradition would be a valuable complement to research based on modern editions of historical texts and might in some cases lead to a revision of its results.

5. Acknowledgements

The project “Breytileiki Njáls sögu/Variance of Njáls saga” (principal investigator: Dr. Svanhildur Óskarsdóttir) is funded by Rannsóknarmiðstöð Íslands/The Icelandic Centre for Research (Rannís). I would like to thank Ulrike Henny (Cologne Center for eHumanities) and Dr. Kai Wörner (Hamburger Zentrum für Sprachkorpora) for their advice in connection with the designing of XSLT-style sheets and Dr Emily Lethbridge (Háskóli Íslands, Miðaldastofa) and three anonymous reviewers for valuable suggestions for an improvement of the article.

6. References

Dasent, G. W., Sir (Trans.) (1971). The Story of Burnt Njal. London, New York: Dent, Dutton (1st ed. Edinburgh: Edmonston & Douglas 1861).

Hall, A. et al. (in prep.). A new stemma of *Njáls saga*.

Hallberg, P. (1968). Stilsignalement och författarskap i norrön sagalitteratur (Nordistica Gothoburgensia, 3). Stockholm, Göteborg, Uppsala: Almqvist & Wiksell.

Hauksson, Þ. & Óskarsson, Þ. (1994). Íslensk stílfræði. Reykjavík: Mál og menning.

Jónsson, M. (2003). Megindlegar handritarannsóknir. In E. Ornato, Lofræða um handritamergð. Hugleiðingar um bóksögu miðalda (Ritsafn Sagnfræðistofnunnar, 36). Reykjavík: Sagnfræðistofnun Háskóla Íslands, pp. 7-34.

Rögnvaldsson, E. & Helgadóttir, S. (2011). Morphosyntactic tagging of Old Icelandic texts and its use in studying syntactic variation and change. In C. Sporleder, A. van den Bosch, K. Zervanou (Eds.), Language Technology for cultural heritage. Selected papers from the LaTeCH workshop series. Heidelberg, Dordrecht, London, New York: Springer, pp. 63-76.

Rögnvaldsson, E., Ingason, A. K., Sigurðsson, E. F., Wallenberg, J. C. (2012). Sögulegi íslenski trjábankinn. Gripla 23, pp. 331-352.

Sahle, P. (2013). Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik. Norderstedt: BoD.

Sprenger, U. (1951). Praesens Historicum und Praeteritum in der altisländischen Saga. Ein Beitrag zur Frage Freiprosa-Buchprosa (Basler Studien zur deutschen Sprache und Literatur, 11). Basel: B. Schwabe.

Sveinsson, E. Ó. (Ed.) (1954). Brennu-Njáls saga (Íslenzk fornrit, 12). Reykjavík: Hið íslenzka fornritafélag.

Sveinsson, E. Ó. & Þórðarson, M. (Eds.) (1935). Eyrbyggja saga. Brands þáttr Qrva. Eiríks saga rauða. Grœnlendinga saga. Grœnlendinga þáttr (Íslensk fornrit, 4). Reykjavík: Hið íslenzka fornritafélag.

Thoma, C. (2011). The function of the Historical Present tense: Evidence from Modern Greek. Journal of Pragmatics 43, pp. 2373-2391.

Torgilstveit, T. (2001). Historisk presens i et utvalg islandske tætter. Hovedfagsavhandling i norrøn filologi. Universitetet i Bergen, Nordisk institutt.

Torgilstveit, T. (2007). Historisk presens på norrønt. Maal og minne 2007,1, pp. 29-50.

Zeevaert, L. (2008). The development of subordinate clauses in Old Swedish – an example of contact-induced language change? In A. Zanchi (Ed.), Skáldamjöðurinn. Selected proceedings of the UCL graduate symposia in Old Norse literature and philology, 2005-2006. London: Centre for Nordic Research. University College London, pp. 181-225.

Zeevaert, L. (2012). Low German influence and typological change in Swedish: some results from a research project. In L. Elmevik & E.-H. Jahr (Eds.), Contact between Low German and Scandinavian in the Late Middle Ages. 25 Years of Research (Acta Academiae Regiae Gustavi Adolphi, 121). Uppsala: Kungl. Gustav Adolfs Akademien för svensk folkkultur, pp. 171-190.

Zeevaert, Ludger (2013a): Axes, halberds, bows or stones? Tools to get to grips with linguistic variation in the manuscripts of Njál’s saga. Working paper, Rannís-project “Breytileiki Njáls sögu/Variance of Njáls saga“.

Zeevaert, L. (2013b): The abbreviation system of Njál’s saga manuscript AM 162 B fol. δ. Working paper, Rannís-project “Breytileiki Njáls sögu/Variance of Njáls saga”.

Zeevaert, L. (in prep.): IceTagging the “Golden Codex”. Using language tools developed for Modern Icelandic on a corpus of Old Norse manuscripts. Submitted for the LRT4HDA-workshop on the LRE-conference, Reykjavík, 26. May 2014.