Describing Inflectional Patterns of Nouns in Old Icelandic

Ellert Thor Johannsson [1] and Finnur Ágúst Ingimundarson [1]

[1] The Árni Magnússon Institute for Icelandic Studies, Dept. of Lexicography, University of Iceland, Laugavegi 13, IS-105 Reykjavík, Iceland.

Abstract

The Database of Old Icelandic Inflections (DOII) is a project at *The Árni Magnússon Institute for Icelandic Studies* (SÁM) at the University of Iceland with the goal to describe the inflectional patterns of Old Icelandic. DOII uses the same structure as the Database of Icelandic Morphology (DIM), which is an already developed digital resource. The linguistic data comes from *A Dictionary of Old Norse Prose* (ONP), a historical dictionary of the medieval language of Iceland and Norway. The first phase of the project focuses on simplex nouns. ONP lists over 2,400 simplex nouns with more than 10 citations, which give a broad representation of the noun system and can be processed in the DOII as an initial step to describe the inflectional patterns of Old Icelandic.

Keywords

Inflectional database, Old Icelandic, Morphology

1. Introduction

This paper accounts for the Database of Old Icelandic Inflections (DOII), a project that aims to describe the inflectional patterns of Old Icelandic through computer modeling, currently underway at the Árni Magnússon Institute for Icelandic Studies (SÁM) at the University of Iceland. The paper is structured as follows. First, we account for the background of the project, how the inflectional system of Old Norse has traditionally been described and why such a project is warranted. We then give an overview of the inflectional system and discuss the two main components of the project, namely the structure of the computer model and data we need to process. We also discuss the benefits of our approach as well as normalization principles. We then describe the long-term vision of the project before going into more detail of the initial phase and the methods used to input the morphological information about simplex nouns. We mention some of the challenges and issues we have had to consider and how we have chosen to approach them. Finally, there are some conclusions based on the work so far.

2. Background and approach

The discussion of the inflectional system of Old Norse has traditionally been based on the presentation of a system of rules with a few selected examples (cf. [1], [2]). This is understandable considering the medium of printed books aimed to comprehensively account for the entire grammatical system of the language. More recently one can find information on the morphological system in web resources, such as Wiktionary [3], which certainly have the capacity to expand beyond the paper, but are still quite limited, only displaying selective examples. The representative paradigms with cherry-picked examples of the inflectional system tend to be chosen on the basis of information gathered from a long tradition of Germanic historical linguistics. They are usually presented as rather straight forward, neatly grouped into clearly defined inflectional classes based on stem types, with some exceptions listed. The main problem with such presentation of the inflectional information is that it is not clear where the data is coming from, i.e., what sort of textual or philological work it is based on, and the integrity of the forms is often in doubt. Moreover, a comprehensive description is lacking.

It therefore makes sense to try another approach, which can be referred to as a "bottom-up approach", where the inflectional description and the paradigmatic classification is based on the actual data, i.e. the grammatical forms attested in the surviving texts. With this in mind we set out to build a database of attested forms to create a new description of the inflectional patterns of Old Norse. This is the idea behind the project: *Beygingarlýsing íslensks fornmáls* or *The Database of Old Icelandic Inflections*, BÍF or DOII for short.

The scientific value of such a project is twofold: It is an important teaching and research tool for all those interested in Old Norse and Icelandic literature, language, and culture, but at the same time the data on the vocabulary would be accessible for use in various language technology projects, from text analysis, spelling standardization, transcription of texts (e.g, normalized spelling of Old Norse texts), interconnection of databases and text, and material for search engines. The value of the project in lexicographic context is also unequivocal as very precise information about inflection is an important part of any lexical description.

The "bottom-up approach" to the inflectional description of Old Norse is greatly inspired and influenced by modern corpus linguistics and especially by earlier work on the inflectional system of Modern Icelandic. Work on describing the inflectional system of Icelandic based on corpus material dates back to the early 2000's when the visionary scholar, Kristín Bjarnadóttir, started her work on what later would become known as the Database of Icelandic Morphology (DIM) (Beygingarlýsing íslensks nútímamáls, BÍN [4]). Her work pioneered the approach of letting the data lead the way and set up inflectional patterns in accordance with the actual attested forms (see [5]). Today the DIM is free and available online and is continuously expanded with material from the Icelandic giga-corpus (cf. [6]), which is the largest corpus of Modern Icelandic containing over 1.6 billion tokens in its latest version.

The DOII project was initiated under the aegis of SÁM, the home of DIM and various other language technology resources, in collaboration with the *Dictionary of Old Norse Prose* (ONP) in Copenhagen. The authors of this article are the main contributors of the project, with four experts forming an advisory board: Guðrún Þórhallsdóttir and Haraldur Bernharðsson, both comparative linguists and associate professors in Icelandic at the University of Iceland, Kristín Bjarnadóttir, editor of DIM, and Tarrin Wills, editor and programmer of the digital version of ONP [7].

3. Old Norse and its Morphology

Old Icelandic together with Old Norwegian is usually considered a single language, commonly referred to as Old Norse. This language is the written language preserved in surviving manuscripts from the earliest records (dated around 1150) up until premodern times, usually set at 1540 when the first printed book appeared in Iceland (cf. [8]). The Old Norse period in Norway is somewhat shorter as linguistic developments in the 14th century reshaped the Norwegian language to such a degree that it can hardly be referred to as Old Norse beyond 1370–1400 (ibid.). Old Norse is the ancestor of Modern Icelandic and the dialects spoken in Norway today. One of the main characteristics of Old Norse is its morphology which includes a rich system of noun declensions and verb conjugations as well as a dynamic system of derivational affixes. Modern Icelandic is structurally very conservative and has preserved most of the morphological features of the earlier stage of the language. This sets Modern Icelandic apart from Modern Norwegian and other mainland Scandinavian languages, which have lost most inflectional forms. Modern Faroese is another descendant of Old Norse which also has preserved a rich inflectional morphology, although not to the same degree as Modern Icelandic.

The morphological system of Old Norse is quite complex. A database centering on Old Norse inflections has to account for all theoretical forms of each word. These are different for each part of speech. The declensional categories for nouns are case (nominative (nom.), accusative (acc.), dative (dat.) and genitive (gen.)), number (singular (sg.) and plural (pl.)) and definiteness (all forms can have a suffixed definite article). The inflectional categories for adjectives are the same as for the nouns in addition to different levels of gradation (positive, comparative and superlative). The verbs are conjugated in three persons, number (sg. and pl.), tense (present and preterite), mood (indicative and subjunctive) and voice (active and middle). The adverbs also show gradation patterns and the pronouns share many of the inflectional characteristics of nouns and adjectives with the addition of the dual number for some personal pronouns. The numerals from one to four inflect for person, case and number.

It is therefore clear that a description of the inflectional system has to be able to account for the morphological manifestations of all these different grammatical categories.

4. The structure of the project

The morphological features of Old Norse have survived into Modern Icelandic and so have most of the inflectional forms. Any database of Old Norse inflections can benefit from existing resources developed for Modern Icelandic, both regarding methodology, actual organization and programming, as well as how to base inflectional description on data from texts. The data also need to be extensive and truly reflect the language of the medieval texts. We therefore found it logical to base our project on some of the work that has already been done and seek collaboration with the architects of the Database of Icelandic Morphology (DIM), as well as the editors of the Dictionary of Old Norse Prose (ONP).

4.1. Database of Icelandic Morphology (DIM)

The infrastructure that has been developed for DIM is detailed and extensive (see description of DIM [9]). The first version, in the early 2000's, was a descriptive collection of inflectional paradigms in Modern Icelandic. Since then it has grown and expanded to contain much more data, i.e. on word formation and various grammatical features as well as information on the genre, style, domain and age of forms, making it partly prescriptive and essentially allowing for a more historical database of Icelandic morphology (ibid.). All of DIM's data is readily accessible in CSV format for language technology purposes under a CC BY-SA 4.0 license, whereas the data on DIM's website currently only contains descriptive inflectional paradigms and notes on usage, variants, etc. (cf. DIM's website [10]).

As the people involved in this project are affiliated with the same institute as we are (SÁM), we sought their collaboration on this new project on Old Icelandic. In this way we could receive advice and assistance from people that have had a long experience in developing a digital tool similar to the one we have in mind. Thanks to Kristín Bjarnadóttir, the editor of DIM, and Samúel Þórisson, DIM's main programmer, a prototype for an inflectional database based on the structure of DIM was prepared for this project. It contained no data, simply the bare bones, or the structure of the fully developed DIM. Since the inflectional system of Old Norse and Modern Icelandic is to a large degree the same, i.e. with the same system of gender, case, number and moods and tenses of verbs, no radical changes were needed to adapt DIM's system to Old Norse. It only needed new data.

4.2. Dictionary of Old Norse Prose

For data we turned to the ONP dictionary, which was established in 1939 and is the most extensive dictionary of Old Norse currently available. ONP is hosted by the University of Copenhagen and is currently published and produced in electronic form online. The website of the dictionary contains a wealth of information on Old Norse vocabulary and original sources with many linked additional resources. ONP contains about 65,000 headwords from Old Norse texts from the period 1150–1540 associated with over 800,000 example citations (cf. [8]). The citation collection of ONP covers a considerable part of all the textual material that is preserved or as much as 5–10% [11]. This collaboration allows us to have access to sufficient amounts of textual data as ONP possesses a large corpus of dictionary citations from all preserved prose text in Old Norse from the medieval period.

ONP is a dictionary of Old Norse, not only Old Icelandic, as some of the material comes from Norwegian manuscripts as well as the oldest preserved documents from the Faroe Islands. Icelandic manuscripts do however account for the vast majority of ONP's text material, at least 71% are from Icelandic medieval sources compared with only 6% of texts originating in Norwegian manuscripts. Additional texts are of mixed origin, i.e. Norwegian texts surviving in Icelandic manuscripts (6%) or from younger, Icelandic sources (see [12]). In the medieval period there were minimal differences between Norwegian and Icelandic texts and the inflectional system appears to be of a consistent character in the whole Old Norse area. Since the texts are overwhelmingly Icelandic and the scribes not necessarily from the geographic area where the text was produced, we decided to look at the medieval

language as a whole and refer to our project as a database of Old Icelandic inflections, although any regional differences in inflectional patterns will be commented on.

A typical entry in ONP consists of an entry head followed by a semantic tree where the citations are listed under each sense to demonstrate the use of the word and its meaning. The morphological form of the headword is marked and appears as underlined on the dictionary website. The form is not further tagged for grammatical information. Key morphological information about the lemma is found in the entry head. Immediately following the headword, we find abbreviations indicating part of speech, followed by further morphological information in square brackets. For verbs this includes the principal parts. In the case of nouns this information includes the gender of the noun along with certain identifying case forms in the square brackets, i.e. morphological forms that describe the inflectional pattern. An example of an entry from ONP Online is shown in Figure 1.

Figure 1: A screenshot from ONP showing a partial entry for a common noun *hestr* 'horse'. The morphological information is given in square brackets ([-*s;* dat. *-i; -ar*]). This shows that the genitive singular form is attested as *-s*, a dative singular form is attested showing an ending *-i* and following the semicolon that indicates that plural forms are attested we find that the nominative plural in *-ar* is recorded in the material. In the example citation the relevant word form is underlined.

It is important to note that the identifying forms are only displayed in ONP if they are found in the material, i.e the relevant form has to be recorded in an actual text. For nouns these forms are gen. sg. and nom. (or acc.) pl. Other forms are also shown if they have relevance for the inflection such as the dat. sg. of the strong masculine nouns, which is always indicated. In contrast the acc. and dat. sg. of the strong feminine nouns are only indicated where they are different from the nominative, as these three case forms tend to be the same for the majority of such nouns. A semicolon is used to separate singular from plural. If no semicolon is present, then one may assume that no plural forms are attested in the citation material.

If the identifying case forms are not attested in the material this is indicated with a minus sign, e.g. a headword followed by [−; -ar] would thus indicate that singular forms are not attested, only plural forms in -ar. The nominative (and accusative) are used as the identifying plural forms. If either of those forms are missing but other plural forms exist it is inferred that the word existed in the plural even though it so happens that the identifying forms are not attested. That way the user of the dictionary is informed that the word exists in the plural and has some plural forms although they are not the

identifying case forms. Further details about the significance of individual forms for certain inflectional categories can be found in the ONP Key [13].

By using the available data from ONP, the DOII project has significant textual material to base the inflectional description on, although it is not as extensive as the material found in many modern language corpora. The additional morphological information registered in the ONP database is also helpful when assigning a word to a particular inflectional pattern.

4.3.   The benefits of DOII

The work on DOII in close collaboration of SÁM and ONP will benefit both parties. The end result will be an independent database in the form of an inflectional description that will fit well with other digital lexicographic and linguistic resources, that are hosted and maintained by SÁM and published online (see the portal *malid.is* for an overview [14]). At the same time, the project benefits ONP, as additional information from DOII can be supplied to each headword in the dictionary, thereby providing the users of ONP with much more detailed information on the morphological forms and variation of whatever word they might be looking up.

4.4.   Normalization

The orthography of medieval texts is highly irregular with abbreviations and non-standard characters. When editing and publishing such texts we have several options on how to represent them. Most popular text editions choose to follow some sort of a standard orthography, where the texts are transcribed in a consistent manner. Scholarly editions tend to be more loyal to the source, by showing more details where the orthographical practices of each scribe are loyally rendered, commonly with many non-standard characters, as well as abbreviations. Some editions go to great lengths to reproduce every character, whereas others are less rigorous in this regard.

ONP adheres to strict philological principles when displaying example citations from the medieval sources and follows to a large degree the practices of scholarly editions, but normalizes headwords (cf. [8]). Since DOII is primarily intended as a reference tool we opted to follow a normalized orthography when presenting the inflectional data. There are several orthographic standards for Old Norse, but we decided to follow the ONP normalization standard. This standard is recommended by the *Medieval Nordic Text Archive* (*MENOTA*), which issues guidelines for publishing electronic editions of Old Norse texts (see [15]). The examples in DOII will also be linked to headwords in ONP and DIM by using a fixed ID assigned to each word, which is already a part of the ONP data. This way the orthographic details of each form can be found directly in the ONP data if needed. It is conceivable at a later stage that both the normalized and unnormalized form will be accessible directly in the DOII, but for the time being the inflectional patterns will be given in a normalized form.

5.   The project as a whole

In the previous section we have described the basic building blocks and certain principles of the DOII project. The long term goal is to identify all the different inflectional patterns that are found in the vocabulary preserved in Old Norse medieval sources, i.e. for all parts of speech, such as nouns, adjectives, verbs, pronouns etc. The extent of the project and complexity of the material has led us to divide it into several phases where in each phase we focus on different parts of speech. For all phases the initial reference point is the lemma list of ONP. Later we aim to add the additional words found in poetical sources (cf. [16]) as well as place names and proper names that are only partly represented in the aforementioned databases, but are listed in some glossaries and handbooks. The final result will be a comprehensive description of the inflectional patterns of all attested Old Norse vocabulary.

As part of the project, we aim to publish the inflectional information on an independent website, identical to the one hosting DIM. The results of the project will be published gradually as each phase is completed. The website will be open to everyone, in addition to which the data will be made accessible with an open license and stored at CLARIN [17], like other official Icelandic language technology data.

When available, the data from DOII could also be linked to DIM's data so that information about the historical development of inflectional patterns in Icelandic could easily be accessed. This would make it possible to examine how the traditional presentation of inflectional classes for Old Norse compares to the actual data. Such a resource could form the basis for a historical inflectional description of the Icelandic language that could be used by laymen and scholars from all over the world and become a source of countless studies of Icelandic historical morphology.

5.1. The first phase − simplex nouns

The current investigation falls under the first phase of the project; an account of the inflection of nouns where inflectional patterns of basic nouns will be described and defined. ONP has identified 7,337 basic nouns, i.e. uncompounded and unaffixed simplex nouns, which can be analyzed and entered into the database. The reason for starting with simplex nouns is primarily a practical one, this is a limited category of important lexical items that show enough complexity as to test the capabilities of the system.

One major advantage of focusing on simplex nouns (e.g. *maðr* 'man', *kona* 'woman', *hestr* 'horse') is that the database entry system subsequently can be used to suggest the correct inflectional pattern for compound words, where the last element of the compound is one of the simplex nouns already defined (e.g. *ójafnaðarmaðr* 'overbearing man', *draumkona* 'dream woman', *graðhestr* 'stallion'). Although we do not have estimates for how many of the compound nouns this applies to — ONP has registered over 30,000 compound nouns — we assume that this will account for a significant part of them and facilitate the work in describing the inflections of all attested nouns as they should cover a broad spectrum of inflectional patterns.

5.1.1. The practical approach

When we started investigating the material we decided to begin with those nouns where we have numerous citations from the material to demonstrate their use and enough forms to assign them to a particular inflectional pattern. Of the 7,337 uncompounded nouns found in ONP, 2,441 have more than 10 citations. In order to process them into the actual DOII framework ONP generated from its database a list of all simplex nouns where the following information was indicated: lemma, modern Icelandic equivalent, gender, inflection (i.e. identifying forms), number of citations, lemma ID and a list of attested forms.

Using the ONP wordlist we could then filter the data based on predefined criteria, such as number of citations and inflectional information, separating the ones with 10 or more citations. We then grouped those words according to the morphological information from the ONP database. A screenshot from this secondary list is shown in Figure 2. Once an inflectional pattern had been defined, all nouns belonging to that same pattern could be entered by modifying entry strings (cf. 5.1.2.).

lemma	modicel	gender	hodecl	cits	id	forms
mót	mót	n.	-s; dat. -um	1871	55300	moti (604), mot (331), móti (320), mote (307), mót (174), motz (54), mott (37),
hǫnd	hönd	f.	handar, dat. hendi;	1846	39633	hendr (357), hendi (262), handa (190), hond (153), hondum (99), hende (78), h
sǫk	sök	f.	sakar; sakar/sakir	1439	78738	sakir (455), saker (251), sakar (228), sacar (159), sok (63), sacir (57), saka (29), s
maðr	maður	m.		1288	51596	menn (406), maðr (161), manna (155), mann (108), monnum (63), madr (38), n
vegr	vegur	m.	-s/-ar, dat. -i/-; -ar/	1233	85237	veg (452), vega (139), ueg (105), vegna (84), vegar (61), vegh (35), uega (34), v
messa	messa	f.	-u; -ur	1048	53927	messo (493), messu (244), messa (125), messv (58), Messo (26), mœsso (24), r

Figure 2: A screenshot from the secondary list of ONP's data, showing the 6 most common nouns (based on the number of citations).

5.1.2. An example of a pattern

The inflectional patterns form the main pillars of the database and all headwords, regardless of the category they belong to, are assigned to one. Each inflectional pattern for nouns consists of two main features: stem types and inflectional endings. A case in point is the masculine noun *fjǫrðr* 'fjord', traditionally classified as belonging to masculine u-stems. Although it has 16 different inflectional

forms, 3 stem types (*fjǫrð, firð, fjarð*) suffice to account for all of them, i.e. with the addition of the inflectional endings, cf. Figure 3:


```
 1  NFET:X1+r
 2  NFETgr:X1+rinn
 3  PFET:X1+
 4  PFETgr:X1+inn
 5  PGFET:X3+i
 6  PGFETgr:X3+inum
 7  EFET:X4+ar
 8  EFETgr:X4+arins
 9  NFFT:X3+ir
10  NFFTgr:X3+irnir
11  PFFT:X1+u
12  PFFTgr:X3+ina
13  PGFFT:X2+um
14  PGFFTgr:X2+unum
15  EFFT:X4+a
16  EFFTgr:X4+anna
```

	Eintala			Fleirtala	
	án greinis	með greini		án greinis	með greini
Nf.	**fjǫrðr**	**fjǫrðrinn**	Nf.	**firðir**	**firðirnir**
Þf.	**fjǫrð**	**fjǫrðinn**	Þf.	**fjǫrðu**	**fjǫrðuna**
Þgf.	**firði**	**firðinum**	Þgf.	**fjǫrðum**	**fjǫrðunum**
Ef.	**fjarðar**	**fjarðarins**	Ef.	**fjarða**	**fjarðanna**

Figure 3: Two screenshots from the database, showing, on the left, the structure of the inflectional pattern in question and, on the right, the declension of *fjǫrðr* as an example of that same pattern.

The inflected forms derive from manually modified entry strings run through the database entry system in accordance with the inflectional structure on the left in Figure 3. These strings contain, in their most basic form, the following information for nouns: word, inflectional pattern, number (singular/plural or both), definite or indefinite, and stem types. The stem types correspond to the ones in the inflectional pattern in lines 1–16 in Figure 3. The forms in Figure 3 are merely a preview to showcase the system and do not entirely reflect the ONP data, as the only definite plural form attested is the accusative form *fjǫrðuna*. On the basis of this system all nouns, or headwords for that matter, following the same pattern can be automatically inflected as long as the stem types for each word are defined.

5.2.   Challenges and problematic issues

This initial phase of the project is a pilot phase, which has allowed us to get a better feel for the data and its inherent challenges that we need to address. We have needed to make some editorial decisions on how we present the data and what is the optimal way to display the information gathered. During this process we have encountered several challenges and problematic issues that we have had to find a way to resolve, and we will continually be revising our methods as we process more of the data.

A fundamental question is how to distinguish between truly attested forms and those that are theoretically possible but not attested in the data. Once we have entered the inflectional information the system will generate a full paradigm showing all possible forms. ONP typically only provides examples of a few forms, even when multiple citations are attested. An illustrative example would be the dative singular form of *Yggdrasill*, the world tree, known from Norse mythology, which is not attested. The data only records genitive *Yggdrasils*. This word also exists in Modern Icelandic and in DIM the dative form is given as *Yggdrasil*, with an editorial comment stating that it is a surmise — is *Yggdrasli* perhaps the correct form (cf. [5])? This latter hypothesized form would not be unfathomable as other masculine words, such as *lykill* 'key', have similar dative forms attested, e.g. *lykli*. As the system is set up now all theoretical forms will be generated. An editorial procedure at a later stage will involve checking off the forms that are attested in the data in order to be able to present them clearly.

In other cases, the hypothesized forms are more straight-forward, such as definite forms like *firðirnir*, where the suffixation of the article is completely regular. Similar examples are plural forms of weak nouns *afi* and *amma* 'grandfather, grandmother'. Both are only attested in the singular. Here the plural can be generated without hesitation (*afar* and *ǫmmur* respectively) as there is no other possible inflectional pattern for these kinds of singular forms. A slightly different case is the hypothesized singular form *mæðga* in ONP, which is generated from plural only forms *mæðgur*. This word only exists in the plural as the meaning is 'mother and daughter', so a singular form is not really warranted. In such cases the hypothetical singular forms would not be displayed in DOII.

Sometimes there aren't enough morphological clues in the attested forms so more than one inflectional pattern is possible. In such cases it is possible to consider several alternative criteria to determine the most likely pattern, such as the occurrence of the noun in question in compounds. We

can also try to find additional examples in other corpora and resources, which also contain vocabulary from the period, such as the concordance of MENOTA's text archive and Lexicon Poeticum (LP). An example would be the masculine noun *rekkr* 'man', which is only attested in plural in ONP's data (nom.pl. *rekkar*). Based on the plural forms alone the nominative singular could just as well be *rekki* as *rekkr*. In this case the compounds where this word occurs as the last element do not give any clues. However, this word is attested in poetry and data from LP contains singular forms confirming that the nominative singular form is the latter variant. Such editorial decisions will be noted in a comment.

We have encountered other cases where the data is compatible with more than one inflectional pattern. In such cases we look to the sources and put greater weight on the earliest attestations. The same can be said about variant forms where we try to order them chronologically. The oldest form is taken as the best representation and then the younger forms are considered variants of the old form. Problems involving the forms for *oxi* 'ox, bull' illustrate this and the borderline between inflectional variations and separate patterns or lemmas. ONP lists three different lemmas for this particular word: *oxi, uxi, yxi* as well as three different nom.pl. forms: *yxn, øxn, oxar*. It would be possible to list the variant forms of *oxi* as a separate paradigm under the lemmas *uxi* or *yxi*. Here we have chosen the solution to look at the most archaic forms as basic, based on information on the sources and dating of the forms in ONP, and list the variant forms as part of a single inflectional pattern. This particular word shows great variation of forms and stands alone in the Old Norse inflectional system. Up to three variants can be displayed for the same case in each paradigm (e.g. masc.pl.nom. *menn/mennr/meðr* 'men'), which is in most cases sufficient to do variant forms justice, although there are several exceptions.

There are also several cases of homographs in ONP's data, such as the feminine nouns *lykkja* 'loop' or 'piece of land' and *vaka* 'awake' or 'fluid', which have two separate entries in ONP because of differing etymology and meaning. Even though one would consider these words as two separate words in a lexical description this distinction is not necessary when looking at the forms. Both homographs follow the exact same inflectional pattern, and no distinction is therefore required in the DOII. If certain inflectional forms might be associated with a particular meaning this will be commented on.

Orthographic variation is also a challenge as we have some variation in almost every single form attested more than once in our data. This is to be expected as the orthography of the source material is not standardized and highly irregular. This makes it very difficult to automate the process of checking off the forms that are attested in the data. Accounting for orthographic variations will wait until subsequent phases, where the classification system and analysis from the DIM system need to be adapted to better cover the historical dimension. The distribution of orthographic variants can be commented on as needed in DOII, following the practice of DIM with the use of free text fields.

6. Conclusions

The database of Old Icelandic inflections is a work in progress. We have only taken the first small step to establish the principles and begin the work of describing the elaborate inflectional system of Old Norse. In this article we have discussed the background and origin of the project and the essential building blocks that we intend to build upon as the project progresses.

At the end of this first phase, we hope to have created a usable prototype of the tool and at the same time acquired the knowledge and understanding of DOII's internal system necessary to be able to further develop and expand it in subsequent phases.

When finished, the DOII will be a valuable addition to the available resources for studying and working with this important literary language. The basic structure adapted from DIM demonstrates that the methodology functions well for this sort of historical linguistic description, although there are some problematic issues that need to be addressed and resolved. In principle the method and approach that are discussed in this article do not exclusively apply to Icelandic or Old Norse but can be used in relation to many other languages as well. A similar database, based on DIM, is currently being developed for Modern Faroese. The core functions of the DIM are applicable for describing the morphology of any inflectional language. This applies especially to other Indo-European languages, such as Latin or Russian, where the paradigms are similar to the ones we find in Icelandic and Old Norse and in the case of nouns, are primarily based on stem types and inflectional endings.

7. Acknowledgements

8. References

[1] Adolf Noreen. 1923. *Altnordische Grammatik I: Altisländische und altnorwegische Grammatik (Laut- und Flexionslehre) unter Berücksichtigung des Urnordischen*. 4th edn.. Halle: Niemeyer

[2] Iversen, Ragnvald. 1990. *Norrøn grammatikk*. 7. utg. revidert ved E.F. Halvorsen. Oslo: Tano.

[3] Wiktionary https://en.wiktionary.org/wiki/Category:Old_Norse_lemmas (February 2022).

[4] BÍN = *Beygingarlýsing íslensks nútímamáls.* Kristín Bjarnadóttir (ed.). Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum. <https://bin.arnastofnun.is/> (February 2022).

[5] Bjarnadóttir, Kristín. 2014. Beygingarlýsing íslensks nútímamáls: Regluverk eða beygingardæmi. *Orð og tunga* 16:123–140.

[6] Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. *Proceedings of LREC 2018*, p. 4361–4366. Myazaki, Japan.

[7] ONP Online = *Ordbog over det norrøne prosasprog/A Dictionary of Old Norse Online.* <onp.ku.dk> (February 2022).

[8] Johannsson, Ellert Thor & Simonetta Battista. 2016. "Editing and Presenting Complex Source Material in an Online Dictionary: The Case of ONP" in Tinatin Margalitadze & Georg Meladze. (eds*.) Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity, 6–10 September 2016, Tbilisi,* pp. 117–128.

[9] Bjarnadóttir, Kristín, Kristín Ingibjörg Hlynsdóttir & Steinþór Steingrímsson. 2019. DIM: The Database of Icelandic Morphology. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pp. 146–154. (NoDaLiDa 2019, Turku, Finland).

[10] DIM = *Database of Icelandic Morphology*, see [4] above.

[11] Tarrin Wills og Ellert Thor Johannsson. 2019. "Reengineering an Online Historical Dictionary for Readers of Specific Texts". In: Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S. & Tiberius, C. (eds*..). Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*. 1–3 October 2019, Sintra, Portugal, pp. 116–129. Lexical Computing CZ, s.r.o, Brno.

[12] Johannsson, Ellert Thor & Simonetta Battista 2018. "Middelaldertekster som sproglig ressource" in Ásta Svavarsdóttir & Helga Hilmisdóttir (eds.) Nordiske Studier i Leksikografi 14: Rapport fra 14. Konference om Leksikografi i Norden, Reykjavík 19.–22. maj 2017, 152–161.

[13] Helle Degnbol, Bent Chr. Jacobsen, James E. Knirk, Eva Rode, Christopher Sanders & Þorbjörg Helgadóttir (eds..): *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose.* ONP Registre (1989). ONP 1: a–bam (1995). ONP 2: ban–da (2000). ONP 3: de–em (2004). Key (2004). København: Den Arnamagnæanske Kommission.

[14] malid.is = *The language portal málið.is* <www.malid.is> (February 2022).

[15] Haugen, Odd Einar (Gen. ed.) 2019. *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources. Version 3.0.* Bergen: Medieval Nordic Text Archive.

[16] LP = *Lexicon Poeticum* <lexiconpoeticum.org> (February 2022).

[17] CLARIN = Common Language Resources and Technology Infrastructure. <https://www.clarin.eu/> (February 2022).