Univerzita Karlova v Praze

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Filologie – Anglický jazyk

Ondřej Tichý

Nástroj na tvaroslovnou analýzu staré angličtiny

Morphological Analyser of Old English

Disertační práce

vedoucí práce - Prof. PhDr. Jan Čermák, CSc.

2014

„Prohlašuji, že jsem dizertační/disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu."

Abstrakt

Práce popisuje sestavení a testování elektronického nástroje na automatickou tvaroslovnou analýzu staré angličtiny.

Uvádí metodologická a materiální východiska projektu ve vztahu k současnému stavu poznání na poli strojové analýzy staré angličtiny a na základě přehledu staroanglického tvarosloví. Podrobně rozvádí zvolenou metodologii a konkrétně popisuje postup implementace projektu: od sběru vstupních dat a volby technologie, přes programování až k testování výsledků. Výslednou návratnost 95% je možné pokládat za úspěch, práce však na konkrétních výsledcích ukazuje, jak by bylo možné návratnost zlepšit. Zabývá se také dalším využitím sestrojeného analyzátoru, především možnostmi disambiguace výsledků.

Výsledky práce dávají naději na brzkou možnost částečně automatizovaného značkování staroanglických textů.

Abstract

The paper describes the construction and testing of an electronic application for automatic morphological analysis of Old English.

It introduces resources and methodologies at our disposal based on the state of the art in the field of electronic analysis of Old English and on an overview of Old English morphology. A detailed account of the chosen methodology is offered and a specific description of the implementation is provided: from the acquisition and preparation of the input data and choice of technology to the programming and testing of the results. The resulting recall of 95% can be seen as a success of the project, however, the paper also shows how the recall may be improved. It also discusses further use of the analyser, especially the disambiguation of its results.

The paper makes a future semi-automatic morphological tagging of Old English texts a real possibility.

Contents

List of Tables

List of Abbreviations

<u>General</u>

OE	Old English
ME	Middle English
PDE	Present-day English

<u>Grammatical</u>

Sg.	Singular
Pl.	Plural
M.	Masculine
F.	Feminine
N.	Neutral
Nom.	Nominative
Acc.	Accusative
Gen.	Genitive
Dat.	Dative
Ins.	Instrumental
Pos.	Positive
Comp.	Comparative
Super.	Superlative
1. p.	1. person
2. p.	2. person
3. p.	3. person
Inf.	Infinitive
Ind.	Indicative
Subj.	Subjunctive
Imp.	Imperative
Pres.	Present
Pret.	Preterit
Part.	Participle
Infl.	Inflected

<u>Sources</u>

W or Wright	Joseph Wright's OE Grammar
C or Campbell	Alistair Campbell's OE Grammar
BT	Bosworth-Toller's *An Anglo-Saxon Dictionary*
DOE	*Dictionary of Old English*
DOEC	*Dictionary of Old English Corpus*
OED	*Oxford English Dictionary*

Notes to readers:
Paragraphs dealing with technical matters only are indented to the right and marked by a sans serif font.

Glossary of Basic Technical Terminology

(Only the specific areas in which the terms are used in this paper are covered.)

array   a data structure of indexed set of elements (variables), or more simply a list of "items"

array of arrays   an array whose elements contain (link to) other arrays, a list of lists, or more simply a table

client-side   running directly on the side of a client, e.g. on the user's computer

hash   an array with named indices (individual elements are not indexed by numbers, but by names)

server-side   running on the server (client gets only results)

source code   the set of computer instructions as written by the programmer rather than on the server

Glossary of Basic Technical Terminology

1. Introduction

1.1.   Definition

The automatic morphological analyser this dissertation aims to implement is to be a computer program that receives Old English (OE) forms on input, evaluates them automatically and outputs all morphological functions the forms can conceivably carry (within our understanding of OE grammar). Together with the functions, the analyser will also provide the users with all dictionary forms (lemmata) from which the input forms may be plausibly derived by OE inflectional morphology and with links to their definitions (e.g. links to respective dictionary entries).

1.2.   Aims and Motivations

The motivation to create the program is manifold, but mainly pedagogical and scientific.

Let us first focus on the pedagogical. When students are confronted with OE texts today, they are usually not expected to actively master what is, in a sense, a dead language[1]. They are expected to master some skills that would allow them to parse written texts and understand them with the help of grammars and dictionaries. Therefore, one of the most important skills they need is a skill to look up words from the texts in the reference books (both dictionaries and grammars). With non-standardized texts written in a fusional[2] language that is far from trivial. It requires from the reader not only a firm grasp of the morphological system of the language, but also some expertise in identifying all kinds of variation that can affect the written forms. This knowledge comes with experience – and

---

[1] In the sense that there are no living native speakers today that would readily understand it.
[2] An inflectional language with discontinuous or overlapping morpheme boundaries.

that is hard to come by, with smaller and smaller portions of the curriculum left for the study of "dead" languages. If beginners had a tool that could help them look up forms from the OE texts they are required to read in the dictionaries or grammars and also suggest their grammatical properties, understanding of OE texts could come easier and some of the initial barriers that many students fail to overcome might be lowered.

From the opposite angle – teachers today usually overcome the difficulties described above by supplying the students with texts glossed specifically for their needs. Assembling a good glossary may be enlightening even for the teacher, but to a degree, it is always a somewhat exhausting job and a tool that could generate a basic glossary requiring only some fine-tuning by the teacher would save precious time.

Part of the motivation is scientific. Diachronic linguistics has lately been more and more driven by corpus methodology. With inflectional languages, much of this methodology relies on the existence of lemmatised corpora – without those, some of the basic concepts of corpus linguistics – such as a word count, a type/token ratio or collocations are impracticable or at least incomparable with results from other types of languages or with results from lemmatised corpora. Word count and type/token ratio would be out of proportion due to the number of inflectional forms of the same word, collocation calculations would be much less meaningful if each form of the same word counted separately etc. In fact, the whole concept of a corpus-driven research is impracticable for OE, since the researcher needs to know exactly what he or she is looking for (i.e. all forms of particular lexical items), before he or she even begins the search (Miranda García, Calle Martín, Moreno Olalla, & Muñoz González, 2006, p. 82).

Since there is no lemmatised corpus of OE,[3] it is high time that some effort be expended to change that. If the process could be at least semi-automatized, the lemmatisation could be sped up significantly, but the tools in existence today are more or less unusable (see chapter 1.3.).

The main research questions that motivate the creation of this tool are then as follow:

- Automatic morphological analysers were successfully developed for a number of living languages, but they have been regularly applied only to highly standardised texts. Can similar approaches be adapted to texts with high level of variation?

- What can the application of these approaches, the necessary modifications to them and even their very possible failure tell us something about the morphology or the nature of variation of the analysed languages?

- And can the task be actually done in an efficient way with a reasonable amount of precision?[4]

The dissertation therefore aims to describe and implement a tool that would:

a) simplify the process of finding entries in reference books – specifically, the tool will be embedded into the online version of *The Anglo-Saxon Dictionary* (*BT*);

b) automatically generate all inflectional forms for a given paradigm (for corpus search and educational purposes);

c) automatically create template glossaries that would simplify the manual creation of glosses to OE texts;

---

[3] Of any practical size – see chapter 1.3. for several small sized manually lemmatised corpora.

[4] The "reasonable amount of precision" is an obviously lacking definition for a scientific enterprise. Yet it is very difficult to define, since different applications of the tool may require different level of precision and perhaps even a different method of its calculation. We will discuss the precision in chapter 9.1.3. when we survey the actual results of the tool.

d) make a semi-automatic (or assisted) lemmatisation and morphological analysis of large amounts of texts feasible;

e) attempt to implement and if necessary modify frameworks already employed in automatic morphological analysis of typologically similar languages.

The solutions proposed for these aims will of course need to reflect the scope of this paper.

1.3.   State of the Art

The idea to automatically analyse and/or lemmatise OE texts is not new and several projects have been proposed, attempted or completed over the years. We will describe those that have traits in common with this project in a roughly chronological order:

- *Verbix,* a project of Tord Lindberg's since 1995, features a free web application as well as a downloadable commercial product that can conjugate verbs in several languages including OE. The operation of the programme is simple: the user enters a base form of a verb and the application returns a list of inflected forms together with their grammatical categories. Though we were unable to determine the background processes of the programme,[5] it seems quite clear that a simple list of standard (and only standard) inflected forms with their lemmata and functions forms the backbone of the project. However, the application does not allow the reverse procedure of looking up the lemma based on an inflected form and its use is thus limited to educational purposes or text production – which in OE is clearly limited (Lindberg, 2014).

---

[5] There are no publications concerning *Verbix* and we were not successful in contacting the author through e-mail.

- *Morphological Analyser of Old English Texts* was a project carried out at University of Málaga between 1997 and 2001. The project was based on a small annotated corpus of ca. 77.000 words consisting among others of the text of the *Old English version of the Apollonius of Tyre*. The corpus was manually tagged for POS information, morphological categories and simple definitions based on Clark Hall's *Dictionary*. The Analyser was developed during the annotation process to speed it up by identifying untagged items with the already tagged lemmata taking some common vowel variation into account and using probabilistic methods to match the items. The project produced a complete annotation of the corpus and developed a special *Old English Concordancer* to manage it, but the tool for lemmatisation did not progress further, since the team moved on to "compilation and tagging of Middle English scientific manuscripts" (Calle Martín, J., personal correspondence, November 2006). No data are available on the precision of the tool within or outside the text it was developed on. (Miranda García, Triviño Rodríguez, & Calle Martín, 2000), (Calle Martín & Triviño Rodríguez, 1998), (Miranda García, Calle Martín, Moreno Olalla, & Muñoz González, 2006)

- *The Nerthus Project* has been in development under the direction of Javier Martín Arista at University of La Rioja since 2004. The project attempts to create a lexical database based on Clark Hall. It is focused on derivational morphology, but it contains some information on inflected forms of strong verbs too. Although the project is not geared toward a grammatical morphological analysis, its database could certainly be used to greatly facilitate any automated annotation procedure, once it is finished. (Martín Arista, 2014)

- In 2007 Jason Adams created an experimental *Morphological Analyzer for Old English Verbs* by taking 143 strong OE verbs and following standard grammatical descriptions to generate all their conjugations including some variant forms. In this fashion he created a list of 2.431 valid forms. The experiment however never proceeded any further, e.g. to matching the forms with an authentic OE text or to adding further lexical classes. (Adams, 2007)

- Based on our initial data from the digitization of *An Anglo-Saxon Dictionary* and on manually lemmatised text of OE *Daniel*, Scott Kleinman at California State, Northridge developed a script between 2009 and 2012 that lemmatises tokens of OE texts. It matches the tokens using the edit distance from the forms in the manually tagged *Daniel* and from the *BT* wordlist we have provided. While the data or methods of the projects had never been published and most of the links on the project's website no longer work, there is a list of lemmata and corresponding forms derived by the script from the text of OE *Genesis* A and *B at* http://www.csun.edu/english/match.php. (Kleinman, 20012)

- The most ambitious project in automatic lemmatisation or morphological analysis of OE has started recently at Radboud University Nijmegen as a continuation of an ongoing development of OE corpora. In 2000, the team under Ans Van Kemenade at Nijmegen carried out manual machine-assisted POS tagging of the *Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English* and the *York-Toronto-Helsinki Parsed Corpus of Old English* (*YCOE*) using software by Onno Huber that recognised already tagged forms and allowed the researches to pick from POS information and grammatical categories already associated with the same form (Van Kemenade, Ans, personal communication, January 2014). Since November 2013 Erwin R. Komen has been working on a

project to lemmatise YCOE and the other parsed parts of the *Helsinki Corpus of English Texts*. The project is based on a small manually lemmatised corpus which has been converted into a list of forms, lemmata and POS tags. This list was enriched by lemmata and some inflected forms from Clark Hall and *BT*. When the forms in the list were grouped under a corresponding POS–lemma pair, it was possible to derive certain rewrite rules, i.e. string replacement rules that, if applied on the inflected form in question, would turn it into its lemma. Such rewrite rules were then used to generate inflected forms from other lemmata with the same POS information.[6] Once the list is fully generated, POS tagged tokens from YCOE were matched against forms in the list with the same POS information using Levenshtein string distance to calculate the probability of the match. This process has so far yielded excellent results matching 98,88% tokens to lemmata.[7] (Komen, E.R., personal correspondence, January–February 2014)

1.4.   Choice of Methodologies

A common factor of all the projects noted so far is that the identification of a lemma or grammatical categories is not based on an analysis of an input form per se (i.e. on parsing the string to identify individual morphemes), but on matching the input forms against a list of pre-synthesised items complemented by the morphological information. The primary choice is therefore whether the items should be synthesised by generation or derivation and the secondary choice is the method of matching the pre-synthesised forms with those that are to be analysed.

---

[6] Deriving the generation rules from the material itself is an inspiring concept that our project could also benefit from. Unfortunately, we learned about it only when the programming phase as Nijmegen was over. Also, this process requires a manually pre-lemmatised corpus that we lack.

[7] See chapter 10.1.1. on the proposed cooperation between the two projects.

The choices the projects mentioned above had to make seem to have been based mainly on three factors:

a) Is the tool to be a simple lemmatiser or an analyser working with a number of morphological categories?

b) What is to be analysed? Is there a set text/corpus or should the tool be universal, i.e. able to analyse any OE text?

c) What are the resources at the authors' disposal? Is there a lemmatised/analysed text to use for "learning",[8] a lexical database or a set of morphological rules ready at hand for the project, or is any of these to be created from scratch?

Our aim has been to both lemmatise and analyse the input and we have aimed at allowing any OE on the input. At the same time, given the background of our project (see below), while we can make use of the rich lexical database of the *Germanic Lexicon Project*, we have no lemmatised or morphologically suitably annotated corpus or any set of morphological rules in a digital form beyond digitised texts of standard OE grammars. Neither can we count on a team of OE experts to undertake a time consuming job like manual corpus lemmatisation/annotation on our behalf. Similarly, manual compilation of all attested forms and their classification is not an option for us. We do not have the resources to carry out such a task and a long term project of this type is already under way (and probably will be for a foreseeable future, see chapter 6.1.1.).

Therefore, we have to come up with a methodology that is based on a lexical database and rules of a standard OE grammar (i.e. on generation, see chapter 4.2. below) rather

---

[8] E.g. derivation of rewrite rules.

than on automated learning of the rules or forms (i.e. derivation) from manually annotated texts.

Before we choose a theoretical framework for our project, we should consider all the basic characteristics of OE morphology and all the important peculiarities of OE lexical material (including the standards of its record that we will work with).

2. Overview of Old English Inflectional System

We do not intend to cover the whole OE inflectional system in detail here, we will only describe those parts of the structure necessary for the operation of the proposed analyser.

In doing so, we may diverge from the traditional categories and paradigms that originate in historical linguistics and whose aim is to show the continuity of OE as an offspring of the Germanic languages. We may also diverge from more recent simplified descriptions aimed at contemporary students.

Instead we will choose the level of detail pragmatically – lower for adjectives, higher for verbs, but our unifying principle will be efficiency regarding the computational processing and feasibility with respect to the resources at our disposal.

Therefore, if we deem that the differences between some traditional paradigms are smaller than the variation or oscillation[9] of their members between the paradigms, or if a merger of the paradigms does not cause problems in analysing OE text, we will merge them. If, on the other hand, the number and character of exceptions from a paradigm warrants creation of new paradigms, we will establish them. It is important to remember that due to the requirements of the project we are "dictionary form oriented", we try to define the paradigms so that with the dictionary forms, with a set of endings and with as few replacement rules as possible we can inflect any member of a paradigm.

In case of nouns and adjectives (and classes inflected accordingly), we usually define one paradigm per set of endings. In the case of verbs, we usually have a paradigm for every set of infixes and endings.

---

[9] Be it scribal variation, variation based on sound change or analogical shifts in class paradigm membership.

The measure of efficiency of the proposed description can only be measured by the accuracy and success of the project and may be, therefore, tweaked and adjusted, once the analyser has been used and extensively tested for some time.

As long as there are no reasons to the contrary, we will keep to Wright's and Campbell's descriptions, because we can make good use of their extensive lists of words assigned to their paradigms and because it may be useful to be able to refer users of the analyser back to the traditional and widely accepted reference books for further details on forms and functions. We will also note our major departures from these traditional descriptions here.

One common departure consists in sometimes assigning a grapheme to an ending, where in some traditional grammars, it is understood to belong to the stem or vice versa (e.g. *fré-a* or *wite-0* rather than the traditional *fréa-0* or *wit-e*). The reason for this is to keep the stems as uniform as possible and make their computational processing simpler. The theoretical problem of assigning the graphemes unambiguously into the stem or ending has no practical impact on our project and thus does not concern us here.

The paradigms used by the project are printed in **bold** and we give their full declension in the tables below.

2.1.   Nouns

OE noun is inflected for four cases (nom., acc., gen., and dat.) and two numbers (sg. & pl.). The distinction of nom. and acc. pl. is, however, mostly theoretical/historical, since formally, the two have fallen together in all paradigms. There are thus seven "functional slots" for every noun, though the number of forms to fill them may be both lower (due to

common homography of endings)[10] as well as higher (due to variation). The individual forms are distinguished by endings and rarely by stem changes (infixes). The stem change is most commonly a syncope of the stem-final vowel, assimilation or loss of the stem-final consonant, syncope of the stem pre-final vowel, stem-final gemination of consonants or change in quality of the root or stem vowel.

Every noun is of one (or rarely two and even three) three genders (masc., neut. and fem.) and gender is the major decisive factor in assigning the noun within a particular stem/class. The other factor is the phonological (e.g. root vowel or syllabic structure) and the morphological structure (e.g. derivational suffix) of the noun. But often the factors that have "caused" a noun to inflect in a particular way were already obscured by the OE time and since the analyser will only operate on synchronic OE data, the factors are for our purposes opaque (i.e. it is impossible to predict the inflected forms from the base form alone). Most nouns decline according to one paradigm, which is traditionally classed as strong or weak. Some nouns can, however, be fully or partly declined according to multiple paradigms.

The details on the inflections of the individual paradigms follow.[11]

	singular	plural
Nom.	*stán-0*	*stán-as*
Acc.	*stán-0*	*stán-as*
Gen.	*stán-es*	*stán-a*
Dat.	*stán-e*	*stán-um*

Table 1 Declension of *stán.*

---

[10] We use the term homography of endings or grammatical homography where some authors may use homonymy of endings.

[11] See chapter 6.2.2.1 for the decision making process in assigning nouns to their paradigms.

Most strong masculine nouns are declined like ***stán*** ("stone")[12] in Table 1.[13] Under the same paradigm we subsume words traditionally following paradigms of *cyning* ("king"), *gist* ("yeast"), *secg* ("man") and with the following modifications also:

- *ende* ("end") and *wine* ("friend") – the final –*e* is deleted in all cases except in nom./acc. sg. as in: nom. sg. *wine*, but nom. pl. *winas*;

- *mearh ("horse")* – the final –*h* is deleted in all cases except in nom./acc. sg. as in: nom. sg. *mearh,* but nom. pl. *méaras*;

- *engel* ("angel") – the pre-final vowel is syncopated in all cases except in nom./acc. sg. and pl. as in: nom. sg. *engel,* but nom. pl. *englas*;

- *heofon* ("heaven") – the stem-final –*on* changes to –*en* in plural as in: nom. sg. *heofon,* but nom. pl. *heofenas*;[14]

- and even neuter[15] *cyn* ("kin") – gemination of the final consonant in all cases except nom./acc. sg. & pl. as in: nom. sg. *cyn,* but gen. sg. *cynnes*.

Most strong neuter long stems[16] decline like ***word*** ("word") in Table 2 (but note several long stem paradigms subsumed under the *hof* paradigm below). Under the same paradigm we

	singular	plural
Nom.	*word-0*	*word-0*
Acc.	*word-0*	*word-0*
Gen.	*word-es*	*word-a*
Dat.	*word-e*	*word-um*

Table 2 Declension of *word*.

---

[12] Almost all paradigm example words are based on Wright.

[13] The judgments on class membership refer to our model, not necessarily to the attested OE morphological system.

[14] Note that in addition to these modified variants, variants based on the paradigm example itself are also generated for all members of the paradigm – while *heofenas* is the historical form given by Wright, the analogical form *heofonas* is more than twice as common according to the *DOEC*.

[15] Words that inflect like and already established paradigm are subsumed under it, even though some grammatical properties would call for a distinct paradigm to be established, e.g. here, neuter *cynn* declines almost like masculine *stán*. For us, grammatical properties like gender play no role in determining paradigm membership, unless the properties are distinctly formally expressed.

[16] I.e. stems with long stem syllable. "A long syllable has a long vowel or long diphthong, or it ends with at least one consonant." (Baker, Pronunciation, 2012).

subsume words traditionally following paradigms of *wæter* ("water") and with the following modifications also:

- *tungol* ("star") –the pre-final vowel is syncopated in all cases except in nom./acc. sg. as in: nom. sg. *tungol,* but gen. sg. *tunglas*.

Most strong neuter short stems decline like **hof** ("dwelling") in Table 3. With the following modifications we also subsume under this paradigm words traditionally following the paradigms of:

- *héafod* ("star") –the pre-final vowel is syncopated in all cases except in nom./acc. sg. as in: nom. sg. *tungol,* but nom. pl. *tunglas*;

	singular	plural
Nom.	*hof-0*	*hof –u / hof-o*
Acc.	*hof -0*	*hof –u / hof-o*
Gen.	*hof -es*	*hof -a*
Dat.	*hof -e*	*hof -um*

Table 3 Declension of *hof.*

- *wíte* ("punishment") and *spere* ("spear") – the final –*e* is deleted in all cases except in nom./acc. sg. as in: nom. sg. *spere*, but nom. pl. *speru*;

- *wésten* ("desert") – gemination of the final consonant occurs in all cases except nom./acc. sg. & pl. as in: nom. sg. *wésten,* but gen. sg. *wéstennes*.

Most strong feminine stems decline like **ár** ("honour") in Table 4. Under the same

	singular	plural
Nom.	*ár-0*	*ár-a / ár-e*
Acc.	*ár-e*	*ár-a / ár-e*
Gen.	*ár-e*	*ár-a / ár-na / ár-ena*
Dat.	*ár-e*	*ár-um*

Table 4 Declension of *ár.*

paradigm we subsume words traditionally following paradigms of *firen* ("crime"), *leornung* ("learning"), *gird* ("rod"), cwén ("queen") and with certain modifications also:

- *gifu* ("gift") and *strengþu* ("strength") – the final –*u* is deleted in all cases except in nom./acc. sg. and pl. as in: nom. sg. *gifu*, but gen. sg. *gife*;

- *sáwol* ("soul") – the pre-final vowel is syncopated in all cases except in nom./acc. sg. as in: nom. sg. *sáwol,* but gen. sg. *sáwle*.

- *hen* ("hen") and *byrþen* ("burden") – gemination of the final consonant occurs in all cases except nom./acc. sg. & pl. as in: nom. sg. *hen,* but gen. sg. *henne.*

Two paradigms, **dæg** (masc., "day") and **fæt** (neut., "vat") mark the number by a change of their stem vowel as a result of the so-called Anglo-Frisian Brightening.

	singular	plural	singular	plural
Nom.	*dæg-0*	*dag-as*	*fæt-0*	*fat-u / fat-o*
Acc.	*dæg-0*	*dag-as*	*fæt-0*	*fat-u / fat-o*
Gen.	*dæg-es*	*dag-a*	*fæt-es*	*fat-a*
Dat.	*dæg-e*	*dag-um*	*fæt-e*	*fat-um*

Table 5 Declensions of dæg and fæt.

The masc. paradigm **bearu** ("grove") and neut. paradigm **bealu** ("woe") are similar to the paradigms of *þéo* ("servant") and *cnéo* ("knee") respectively with the exception of the final –*u/w* deletion in the latter two paradigms.

	singular	plural	singular	plural
Nom.	*bear-u / bear-o*	*bear-was*	*beal-u / beal-o*	*beal-u / beal-o*
Acc.	*bear-u / bear-o*	*bear-was*	*beal-u / beal-o*	*beal-u / beal-o*
Gen.	*bear-wes*	*bear-wa*	*beal-wes*	*beal-wa*
Dat.	*bear-we*	*bear-wum*	*beal-we*	*beal-wum*

Table 6 Declensions of *bearu* and *bealu*.

The rest of the strong declensions and all the weak declensions directly used by the analyser do not exhibit any changes in stem either through the incorporation of other paradigms or through sound changes. The declensions of the strong paradigms **feld** (masc., "field") followed also by *hand* (fem., "hand"), **sunu** (masc. "son") followed also by *duru* (fem., "door") and the weak paradigms **guma** (masc. "man"), **fréa** (masc., "lord"), **tunge** (fem., "tongue"), **béo** (fem., "bee"), **éage** (neut., "eye"), **wígend** (masc., "warrior") are listed below:

	singular	plural	singular	plural	singular	plural
Nom.	*feld*	*feld-a*	*sun-u / sun-o*	*sun-a*	*gum-a*	*gum-an*
Acc.	*feld*	*feld-a*	*sun-a*	*sun-a*	*gum-an*	*gum-an*
Gen.	*feld-a*	*feld-a*	*sun-a*	*sun-a*	*gum-an*	*gum-ena*
Dat.	*feld-a*	*feld-um*	*sun-a*	*sun-um*	*gum-an*	*gum-um*

Table 7 Declensions of *feld*, *sunu* and *guma*.

	singular	plural	singular	plural	singular	plural
Nom.	*fré-a*	*fré-an*	*tung-e*	*tung-an*	*béo-0*	*béo-n*
Acc.	*fré-an*	*fré-an*	*tung-an*	*tung-an*	*béo-n*	*béo-n*
Gen.	*fré-an*	*fré-ana*	*tung-an*	*tung-ena*	*béo-n*	*béo-na*
Dat.	*fré-an*	*fré-aum / fré-am*	*tung-an*	*tung-um*	*béo-n*	*béo-m*

Table 8 Declensions of fréa, tunge, béo.

	singular	plural	singular	plural
Nom.	*éag-e*	*éag-an*	*wígend-0*	*wígend-0 / wígend-e / wígend-as*
Acc.	*éag-an*	*éag-an*	*wígend-0*	*wígend-0 / wígend-e / wígend-as*
Gen.	*éag-an*	*éag-ena*	*wígend-es*	*wígend-ra*
Dat.	*éag-an*	*éag-um*	*wígend-e*	*wígend-um*

Table 9 Declensions of éage and wígend.

The rest of the "minor" nominal paradigms were deemed of too small a membership and of too complex inflection to be treated in terms of paradigms, or patterns. Rather, it proved easier to list all the inflected forms of all the paradigms manually according to Wright rather than to organise them in groups and define their common inflectional patterns. These are namely the traditional paradigms of *beadu* ("battle"), *mǽd* ("meed"), *fót* ("foot"), *bóc* ("book"), *hnutu* ("nut"), *hæleþ* ("warrior"), *mónaþ* ("month"), *mægþ* ("maid"), *ealu* ("ale"), *fæder* ("father"), *bróþor* ("brother"), *módor* ("mother"), *fréond* ("friend"), *lamb* ("lamb"), *cealf* ("calf"), *ǽg* ("egg") and *man* ("man").

2.2.   Adjectives

Adjectival declension is similar to the nominal system. The major difference is that every adjective, apart from being inflected according to all three genders (in agreement with its head) also inflects according to both strong and weak inflection (depending on its role in the phrase). In addition to the nominal cases, strong masc. and neut. inflection also contained the vestiges of the instrumental case.

While there is just one weak paradigm, we keep with the traditional nine strong paradigms as described by Wright. The assignment of adjectives into the paradigms is described in detail in chapter 6.2.2., but works mostly on morphological grounds (syllabic structure, stem vowels, base form endings and suffixes play their role).

The ten adjectival paradigms are listed below:

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*til-0*	*til-e*	*til-0*	*til-u / til-o*	*til-u / til-o*	*til-a / til-e*
Acc.	*til-ne*	*til-e*	*til-0*	*til-u / til-o*	*til-e*	*til-a / til-e*
Gen.	*til-es*	*til-ra*	*til-es*	*til-ra*	*til-re*	*til-ra*
Dat.	*til-um*	*til-um*	*til-um*	*til-um*	*til-re*	*til-um*
Instr.	*til-e*	–	*til-e*	–	–	–

Table 10 Declension of *til* ("good").

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*glæd-0*	*glad-e*	*glæd-0*	*glad-u / glad-o*	*glad-u / glad-o*	*glad-a / glad-e*
Acc.	*glæd-ne*	*glad-e*	*glæd-0*	*glad-u / glad-o*	*glad-e*	*glad-a / glad-e*
Gen.	*glad-es*	*glæd-ra*	*glad-es*	*glad-ra*	*glæd-re*	*glæd-ra*
Dat.	*glad-um*	*glad-um*	*glad-um*	*glad-um*	*glæd-re*	*glæd-um*
Instr.	*glad-e*	–	*glad-e*	–	–	–

Table 11 Declension of *glæd* ("glad").[17]

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*blind-0*	*blind-e*	*blind-0*	*blind-0*	*blind-0*	*blind-a / blind-e*
Acc.	*blind-ne*	*blind-e*	*blind-0*	*blind-0*	*blind-e*	*blind-a / blind-e*
Gen.	*blind-es*	*blind-ra*	*blind-es*	*blind-ra*	*blind-re*	*blind-ra*
Dat.	*blind-um*	*blind-um*	*blind-um*	*blind-um*	*blind-re*	*blind-um*
Instr.	*blind-e*	–	*blind-e*	–	–	–

Table 12 Declension of blind ("blind").

---

[17] The variation of the root vowel is historically based on the same sound change mechanism as in the case of the noun *dæg* (see above), but here it does not clearly mark any one grammatical category.

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*héa-h*	*héa-0*	*héa-h*	*héa-0*	*héa-0*	*héa-0*
Acc.	*héa-nne / -ne*	*héa-0*	*héa-h*	*héa-0*	*héa-0*	*héa-0*
Gen.	*héa-s*	*héa-rra / -ra*	*héa-s*	*héa-rra / -ra*	*héa-rre / -re*	*héa-rra / -ra*
Dat.	*héa-um / -m*	*héa-um / -m*	*héa-um / -m*	*héa-um / -m*	*héa-rre / -re*	*héa-um / -m*
Instr.	*héa-0*	–	*héa-0*	–	–	–

Table 13 Declension of héah ("high").

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*þweor-h*	*þwéor-0*	*þweor-h*	*þwéor-0*	*þwéor-0*	*þwéor-0*
Acc.	*þwéor-nne / -ne*	*þwéor-0*	*þweor-h*	*þwéor-0*	*þwéor-0*	*þwéor-0*
Gen.	*þwéor-es*	*þwéor-ra / -a*	*þwéor-es*	*þwéor-ra / -a*	*þwéor-re / -e*	*þwéor-ra / -a*
Dat.	*þwéor-um*	*þwéor-um*	*þwéor-um*	*þwéor-um*	*þwéor-re / -e*	*þwéor-um*
Instr.	*þwéor-0*	–	*þwéor-0*	–	–	–

Table 14 Declension of *þweorh* ("cross"). [18]

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*manig-0*	*manig-e*	*manig-0*	*manig-0*	*manig-0*	*manig-a / -e*
Acc.	*manig-ne*	*manig-e*	*manig-0*	*manig-0*	*manig-e*	*manig-a / -e*
Gen.	*manig-es*	*manig-ra*	*manig-es*	*manig-ra*	*manig-re*	*manig-ra*
Dat.	*manig-um*	*manig-um*	*manig-um*	*manig-um*	*manig-re*	*manig-um*
Instr.	*manig-e*	–	*manig-e*	–	–	–

Table 15 Declension of *manig* ("many").

[18] Notice the change of quantity of the root vowel in the forms that drop the final *–h*.

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*hálig-0*	*hálg-e*	*hálig-0*	*hálig-u / -o*	*hálig-u / -o*	*hálg-a / -e*
Acc.	*hálig-ne*	*hálg-e*	*hálig-0*	*hálig-u / -o*	*hálg-e*	*hálig-a / -e*
Gen.	*hálg-es*	*hálig-ra*	*hálg-es*	*hálig-ra*	*hálig-re*	*hálig-ra*
Dat.	*hálg-um*	*hálg-um*	*hálg-um*	*hálg-um*	*hálig-re*	*hálg-um*
Instr.	*hálg-e*	–	*hálg-e*	–	–	–

Table 16 Declension of *hálig* ("holy").

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*wild-e*	*wild-e*	*wild-e*	*wild-u / -o*	*wild-u / -o*	*wild-a / -e*
Acc.	*wild-ne*	*wild-e*	*wild-e*	*wild-u / -o*	*wild-e*	*wild-a / -e*
Gen.	*wild-es*	*wild-ra*	*wild-es*	*wild-ra*	*wild-re*	*wild-ra*
Dat.	*wild-um*	*wild-um*	*wild-um*	*wild-um*	*wild-re*	*wild-um*
Instr.	*wild-e*	–	*wild-e*	–	–	–

Table 17 Declension of *wilde* ("wild").

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*gear-u / -o*	*gear-we*	*gear-u / -o*	*gear-u / -o*	*gear-u / -o*	*gear-wa / -we*
Acc.	*gear-one*	*gear-we*	*gear-u / -o*	*gear-u / -o*	*gear-we*	*gear-wa / -we*
Gen.	*gear-wes*	*gear-ora*	*gear-wes*	*gear-ora*	*gear-ore*	*gear-ora*
Dat.	*gear-wum*	*gear-wum*	*gear-wum*	*gear-wum*	*gear-ore*	*gear-wum*
Instr.	*gear-we*	–	*gear-we*	–	–	–

Table 18 Declension of *gearu* ("ready").

	Masc.		Neut.		Fem.	
	singular	plural	singular	plural	singular	plural
Nom.	*blind-a*	*blind-an*	*blind-e*	*blind-an*	*blind-e*	*blind-an*
Acc.	*blind-an*	*blind-an*	*blind-e*	*blind-an*	*blind-an*	*blind-an*
Gen.	*blind-an*	*blind-ra / -ena*	*blind-an*	*blind-ra / -ena*	*blind-an*	*blind-ra / -ena*
Dat.	*blind-an*	*blind-um*	*blind-an*	*blind-um*	*blind-an*	*blind-um*

Table 19 Weak adjectival declension of *blinda* ("blind").[19]

2.2.1. Comparative Degree

The comparative degree of adjectives is formed agglutinatively by inserting an *–r–* affix between the stem and the ending.[20] The endings used for the comparative inflection are those of the weak inflection (see *blinda* in Table 19 above), e.g. gen. sg. masc. positive *wildes*, gen. sg. masc. comparative *wildran*. Some adjectives may also exhibit i-mutation of their stem vowel (see the description of i-mutation in chapter 7.1. below), e.g. nom. sg. masc. positive *héah*, nom. sg. masc. comparative *híerra*. However, due to the operation of analogy, it is difficult to predict precisely, which adjectives will have the mutated vowel.

Adjectives *gód* ("good"), *yfel* ("evil"), *micel* ("much") and *lytel* ("little") form their comparatives by suppletion. The respective suppletive forms: *betra/sélra, læssa, mára, wiersa* are supplied manually to the analyser.

---

[19] The syncope of the stem vowel exhibited e.g. in the paradigm of *hálig* producing e.g. gen. sg. masc. *hálges* is also sometimes exhibited in the weak declension yielding both *hálgan* and *háligan* as gen. sg. masc.

[20] Note that this a strictly synchronic OE interpretation, we are not concerned here with a more complex historical basis of the form.

2.3.　Superlative Degree

Similar to the comparative degree, superlatives are formed agglutinatively by inserting an *–ost–* or *–est–* (or rarely *–ust–* or *–ast–*)[21] affix between the stem and the ending. This addition of extra syllable is often compensated for by syncopating the initial vowel of the superlative affix, thus producing superlative *–st–*. The endings used for the superlative inflection can be either of the strong inflectional paradigm *blind* (see Table 12 above) or of the weak inflection (see *blinda* in Table 19 above), e.g. gen. sg. masc. positive *hálig*, gen. sg. masc. superlative *hálgostes* or *hálgostan*. Some adjectives may also exhibit i-mutation of their stem vowel (see the description of i-mutation in chapter 7.1. below), e.g. nom. sg. masc. positive *héah*, nom. sg. masc. superlative *híehst*. However, due to the operation of analogy, it is difficult to predict precisely, which adjectives will have the mutated vowel.

Adjectives *gód, yfel, micel, lytel* form their comparatives by suppletion. The respective suppletive forms: *betst/sélest*[22]*, lǽst, mǽst, wierst* respectively and these are supplied manually to the analyser.

2.4.　Pronouns

2.4.1. Personal Pronouns

Personal pronouns inflect for four cases, three persons and – in the first and second person – for three numbers (singular, dual and plural). In the third person, the inflection distinguishes only two numbers, but in singular, three genders are distinguished. Forms of the personal pronouns are highly suppletive and need to be supplied manually to the

---

[21] The historically "double" superlative formed synchronically by an affix *–mest–* as in pos. *éast* ("east"), sup. *éastmest* is in most of its forms highly irregular and may need to be supplied manually for the attested forms.

[22] Missing in Wright's description, but well attested in the DOEC.

analyser. They are listed here with only the most numerous variants for clarity and because they give raise to some of the possessives below. Note that it is mostly a matter of an arbitrary and/or practical choice to what form(s) personal pronouns are lemmatised. We may pick 1. p. sg. nom. *ic* as the lemma for all personal pronouns, but we may equally pick the nominative form of each person as a lemma of that person, or even the nominative of each number and person. We have chosen the middle way and consider nominatives of each person singular to be the lemmata of the particular persons of personal pronouns and we list them accordingly.

1. person   2. person

	singular	dual	plural	singular	dual	plural
Nom.	ic	*wit*	*we*	*þu*	*git*	*ge*
Acc.	*mec, me*	*unc*	*úsic, ús*	*þec, þe*	*incit, inc*	*éowic, éow, íow*
Gen.	*mín*	*uncer*	*úser, úre*	*þín*	*incer*	*éower, íower*
Dat.	*me*	*unc*	*ús*	*þe*	*incer*	*éow, íow*

Table 20 Declensions of the 1. and the 2. person of personal pronouns.

3. person

	Singular			Plural
	Masc.	Neut.	Fem.	all genders
Nom.	hé	*hit*	*hío, héo*	*híe, hí*
Acc.	*hine, hiene*	*hit*	*híe*	*híe, hí*
Gen.	*his*	*his*	*hiere, hire*	*hiera, hira, heora*
Dat.	*him*	*him*	*hiere, hire*	*him*

Table 21 Declension of the 3. person of personal pronouns.

2.4.2. Possessive Pronouns

The core possessives *mín*, *þín and sín* (all genders) for the singular of 1., 2. and 3. persons respectively follow the strong adjectival paradigm of *blind* (see Table 12). However, *sín*, was often came to be replaced with the genitive forms of the 3. p. of the personal pronouns: masc. *his,* neut. *his* and fem. *hire*, which are indeclinable.[23]

The derived dual and plural forms of the possessives *uncer* and *úser* for 1. p. and *incer*, *éower* for 2. p. follow in their inflections the strong adjectival paradigm of *hálig*, but *úre*, the variant of *úser*, follows *wilde*.

2.4.3. Demonstrative Pronouns

Demonstrative pronouns inflect for five cases (although instrumental is only formed in the neuter), two numbers and in the singular for three genders. Like personal pronouns, demonstrative pronouns are highly suppletive and have to, therefore, be supplied manually to the analyser. We list them here with only the more numerous variants for the sake of clarity. The demonstratives have been lemmatised to *se* ("that") and *þes* ("this").

	Singular			Plural
	Masc.	Neut.	Fem.	all genders
Nom.	*se*	*þæt*	*sío / séo*	*þá*
Acc.	*þone*	*þæt*	*þá*	*þá*
Gen.	*þæs*	*þæs*	*þǽre*	*þára, þǽra*
Dat.	*þǽm / þám*	*þǽm, þám*	*þǽre*	*þǽm, þám*
Instr.	–	*þý, þón*	–	–

Table 22 Declension of the demonstrative *se*.

---

[23] Interestingly, this is noted by neither Wright nor Campbell; nor do they note the plural 3. p. possessive *hiera*.

	Singular			Plural
	Masc.	Neut.	Fem.	all genders
Nom.	*þes*	*þis*	*þíos / þéos*	*þás*
Acc.	*þisne*	*þis*	*þás*	*þás*
Gen.	*þisses*	*þisses*	*þisse*	*þisse*
Dat.	*þissum*	*þissum*	*þisse*	*þisse*
Instr.	–	*þís*	–	–

Table 23 Declension of the demonstrative *þes*.

2.4.4. Interrogative Pronouns

We work with three basic forms of OE interrogative pronouns (and their variant forms): masc. *hwa* ("who") and neut. *hwæt* ("what"); *hwilc* ("which") and *hwæþer* ("which of two").

The first one declines for five cases (though for instr. only in neuter), two genders (masc. and neut.) and only one number – singular. Its endings are similar to those of the demonstrative pronouns:

Singular

	Masc.	Neut.
Nom.	*hwa*	*hwæt*
Acc.	*hwone*	*hwæt*
Gen.	*hwæs*	*hwæs*
Dat.	*hwǽm, hwám*	*hwǽm, hwám*
Instr.	–	*hwí*

Table 24 Declension of *hwa, hwæt*.

While Wright and Campbell agree that *hwilc* and *hwæþer* are declined like strong adjectives, corpus evidence shows also weak declension (i.e. forms like *hwilcan* and

hwǽþeran). We decline them both as strong (following *blind* and *manig*) and weak adjectives (see the appropriate declensions in Tables 12, 15 and 19 above).

2.4.5. Indefinite Pronouns

Indefinite pronouns are declined on the pattern of weak adjectives, see Table 19 above.

2.5. Numerals

2.5.1. Cardinal Numerals

Of the numerals 1–3, each has a specific set of inflections. The rest of the numerals inflect like the plurals of nominal declensions of *stán*, *hof* and *ár* (see Tables 1, 3 and 4 above), depending on gender (in which they agree with the head of the phrase), or remain undeclined.

Numeral *án* ("one") follows the singular inflection of strong adjectival *blind* or the weak adjectival inflection (see Tables 12 and 19 above).

Numerals *twégen* ("two") and *bégen* ("both") decline their plural forms as follows:

	Masc.	Neut.	Fem.
Nom.	*twégen*	*tú / twá*	*twá*
Acc.	*twégen*	*tú / twá*	*twá*
Gen.	*twégea / twégra*	*twégea / twégra*	*twégea / twégra*
Dat.	*twǽm, twám*	*twǽm, twám*	*twǽm, twám*

Table 25 Declension of twégen.

Numeral *þrí* ("three") declines its plural forms as follows (notice that the only distinction between the genders is in the nom./acc. forms of the masc.):

	Masc.	**Neut.**	**Fem.**
Nom.	þrí / þríe	þrío / þréo	þrío / þréo
Acc.	þrí / þríe	þrío / þréo	þrío / þréo
Gen.	þríora, þréora	þríora, þréora	þríora, þréora
Dat.	þrim	þrim	þrim

Table 26 Declension of *þrí*.

2.5.2. Ordinal Numerals.

Forma, formest and *ǽrest* (all three basic forms of "first") and *óþer* ("second") decline like cardinals. [24] The rest of the ordinals (including *æfterra*, which is another base form meaning "second") follow the weak adjectival inflection (see Table 19 above).

Formation of ordinals from cardinals (though fairly regular) is considered a process of word-formation here and therefore outside of the scope of this project.[25]

2.6.   Adverbs

Similar to adjectives, comparison of adverbs consists in adding the suffix/ending *–or* (and variants *–ur* and *–ar*) and the superlative in adding the suffix/ending *–ost* (or its variants *–ust, –ast* and *–st*).

Adverbs *wel* ("well"), *yfele* ("evilly"), *micle* ("much") and *lytel* ("little") create comparatives and superlatives with suppletive forms: *bet* or *sél* (comp.), *betst* or *sélest* (sup.); *wiers* (comp.), *wierrest* (sup.); *má* (comp.), *mǽst* (sup.); *lǽs* (comp.), *lǽst* (sup.), respectively.

---

[24] Again, the grammars and *BT* maintain that *óþer* only follows the strong adjectival inflection, but based on the corpus evidence, we also include it in the weak inflection.

[25] Common ordinals are usually included as independent entries in OE dictionaries.

Formation of adverbs from adjectives as well as formation of adverbs of motion from adverbs of place (though all fairly regular) are considered to be processes of word-formation here and therefore out of the scope of this project.

2.7.   Verbs

While not necessarily the most prolific in terms of the sheer number of functional slots and therefore function-forms (see Table 27 below), the verbal conjugation is undoubtedly the most complex of the OE inflectional systems.

The complexity is mostly due to the variety of affixes and their possible combinations (there is generally much less grammatical homography than in the case of nouns or adjectives), or, in more usual terms, in the number of conjugational paradigms. This number is large not only due to the complexity of the system as such, but also because of the impossibility to reduce it for the purposes of this project into a smaller number of paradigms and a few simple replacement rules, as it was possible in the case of the nominal system (see chapter 2.1. above).

This is due not only to the greater diversity of the verbal system, but also due to the technical difficulties caused by verbal infixation and the prolific changes on the stem–ending morphological boundary. In other words, verbal structures do not yield as easily to simple string operations such as final stem vowel deletion or stem-final syncopation.

Old English verb inflects for mood, tense, person and number. In addition, it has four non-finite forms: the infinitive, inflected infinitive, present and past participles.[26]

---

[26] Both the *–ing* and the *–ung* gerunds are not considered inflected verbal forms, but derived nouns and therefore out of the scope of the project.

The finite forms of both the indicative present tense and the preterit inflect for three persons in the singular and a common plural, but the first and the third persons of the preterit are not formally differentiated. Subjunctive forms inflect for singular and plural in both the present and preterit. Imperative inflects for singular and plural in the present tense only.

Traditionally, OE verbs are divided into four major types, depending on how they form their present and preterit forms.

- Strong verbs form the preterit by vowel gradation historically based on the Indo-European ablaut. They are further subdivided into seven or more classes depending on the specific series of vowels employed in the gradation: different vowels distinguish 1. & 3. p. sg. ind. from past. part. and from all the other preterit forms. The second and third person present tense forms of strong verbs also exhibit introflection of the root vowel, but that is historically a product of i-mutation and, consequently, no indication of the said class membership. In total, there are up to five different root vowels in a particular paradigm of a strong verb.
- Weak verbs form the preterit by a so-called dental ending, though the consonants forming the affix (*–d–* or *–t–*) are usually interpreted to have been alveolar in OE pronunciation. While weak verbs are not affected by the ablaut gradation, several weak paradigms have a different root vowel in present and preterit as a consequence of the i-mutation (which affected the present forms only).

- Preterit-present verbs form their preterits like weak verbs, but their present forms are similar to preterits of strong verbs.[27]

- Anomalous are all the verbs deviating from the structures of the first three types, e.g. verbs with suppletive forms like *béon* ("to be"), but traditionally also *willan* ("to want"), *dón* ("to do") and *gán* ("to go") – though the group may be larger due to derived forms like *nyllan* ("to not want").

We divide verbs into 132 paradigms for the purposes of the project and a complete overview of the paradigms and their principal parts is provided in Appendix 12.2. Here we offer a simplified overview of the endings of the strong and weak conjugation as well as a chart of the ablaut gradation of strong verbs (Baker, Verbs, 2012):

	Strong	Weak
Non-finite		
Infinitive	-an	-(i)an
Infl. Inf.	-anne	-(i)anne
Pres. part.	-ende	-(i)ende
Past. part.	-en	-ed / -od
Present		
1. sg.	-e	-(i)e
2. sg.	-st	-st
3. sg.	-þ	-þ
Plural	-aþ	-(i)aþ
Subj. sg.	-e	-(i)e
Subj. pl.	-en	-(i)en
Imp. sg.	-0	-0
Imp. pl.	-aþ	-aþ
Preterit		
1.&3. sg.	-0	-(o/e)de
2. sg.	-e	-(o/e)dest
Plural	-on	-(o/e)don
Subj. sg.	-e	-(o/e)de
Subj.pl.	-en	-(o/e)den

Table 27 Overview of strong and weak verbal endings.

---

[27] Historically, strong preterit forms of these verbs acquired present tense functions and a new set of preterits was formed on the basis of weak conjugation.

	1.&3. sg. pret.	Plural pret.	Past. part.
1. class	á	i	i
2. class	éa	u	o
3. class	a / ea	u	u / o
4. class	æ / a	ǽ / ó	o /u
5. class	æ	ǽ	e
6. class	ó	ó	a
7. class	é / éo	é / éo	á / ó

Table 28 The ablaut gradation of strong verbs.

The inflected forms of strong verbs can be derived from four principal parts (inf., 1. p. sg. pret. ind., pret. ind. pl. and past. part.), by the application of i-mutation and, in addition, with the knowledge of the above mentioned endings.

For weak verbs, the knowledge of infinitive should suffice, but due to the widespread variation in the stem, other forms are often necessary (especially 2. p. sg. ind. pres. and 1. p. sg. ind. pret.). Inflected forms of the preterit-present verbs and anomalous verbs are difficult to derive from a subset of the forms and may need to be supplied manually.

In other words, to achieve what we have set ourselves to do at the beginning of this chapter, that is – "to define the paradigms so that with the dictionary forms, with a set of endings and with as few replacement rules as possible", is rather difficult in the case of verbs. Namely, we will need more than a base form and some replacement rules to inflect a verb. We will need the principal parts and we will need them segmented, so that we can replace the changeable parts of the verb with the segments based on the example paradigm (e.g. its root vowel).

See the chapter 6.2.1. below on the verbal paradigms for information on how the paradigms were collected and segmented as well as the aforementioned list of the paradigms, principal parts and their segments in the Appendix 12.2.[28]

---

[28] See also the implementation in chapter 7.2. below for the details on how the forms are actually derived.

3. A Few Remarks on Variation and Spelling in Old English

"Old English" is a standard cover term for a linguistic continuum stretching from ca. 700 AD to 1100 AD, from Kent and Wessex to Northumbria, spoken by people in a number of largely independent tribes and kingdoms. This alone should explain why "Standard Old English" is mostly a modern invention practiced out of necessity by modern grammarians and lexicographers. This is of course not to say that there were no local standards of Old English (or perhaps more appropriately "scribal schools or traditions"), but their relation to the standard OE as recorded in the modern reference books is not direct.

Any project attempting to process an authentic non-standardized OE text computationally has to deal with this problem and it is important to understand the nature of the variation of the vernacular on the one hand and the ways modern reference books choose to standardize it on the other.[29]

The choices of the OE scribe, made when he decided to record the spoken language, were of different nature, but most importantly, the scribe was not bound by the idea behind the modern language standards that one word should be spelled in one way and one way only. Let us now consider the major factors of variation in written OE vernacular and discuss briefly how to treat them with respect to our goals of morphological analysis and lemmatisation based on modern secondary material (i.e. chiefly dictionaries and grammars).

---

[29] Especially if we base our project on the modern reference books, and not only on the vernacular of the day.

3.1.   Allographs

The first factor we should consider has to do with the introduction of writing the Latin alphabet to the speakers of OE. Since this was not a unified process, already at this stage spelling variant forms were introduced that do not have any special linguistic significance. Most importantly, such is the variation of *eth* (<ð>) and *thorn* (<þ>), or <k> and <c>, but also the variation of *wynn* (<p>) and <w>, variants of <s>, of <r> or the insular and Carolingian miniscule <a>. Only the first two pairs of allographs need concern us here, since *wynn* is almost always standardized into *w* in OE corpora and electronic editions, as are the letters *s* or *r*, and the difference between the insular and Carolingian *a* is standardized automatically in computer fonts (i.e. the two are considered to be two possible glyphs of the same character). Therefore, all except the first two pairs of allographs are already merged into single graphs. For our purposes, the remaining allographs can be safely merged into one standard form (*thorn* and *c* respectively).

3.2.   Non-Contrastive Grapheme Variation

The second factor of variation in OE written texts is the variation of single graphemes or digraphs (i.e. transcriptions of single phonemes). This may emerge either in the form of different renderings of the same sound (i.e. scribal variation, e.g. *wæter* vs. *uæter* – "water"), or as transcriptions of synchronically different sounds that however (in the context of OE) developed from the same sound and do not signal, in the form under scrutiny, a change of function (e.g. *beran* vs. *beoran* – "to bear"). It may not be of direct consequence to us which type of variation we are dealing with, whether it is a product of a local standard of transcription, of dialectal pronunciation or of a diachronic change inside one dialect. What is important for us is whether a pair or a group of variant

graphemes ever occurs in a minimal pair and, therefore, if their merger across the whole system could create potential homophones. A good example is the group of common variants *i, y* and *ie* – by the end of the OE period, there is only one phoneme[30] they could represent: the front closed unrounded vowel /i/. While they may have originally represented three distinct sounds, they never form minimal pairs and can therefore be merged into one grapheme for the whole of OE.[31] This type of variation is, in its effects on the project, rather similar to allography discussed above.

3.3.   Contrastive Grapheme Variation

The most problematic type of variation is the one concerning graphemes that may form minimal pairs and cannot be safely merged into one grapheme. Unfortunately, this is a very common type of variation in OE and the analyser will have to deal with it. The main principle of the solution has to be to capture as many variant forms under one standardized form and at the same time minimize the risk of merging forms of different lexical items under one standard form. For this reason, the rules governing variation need to be as restrictive as possible. The restrictions may be lexical (when the variation involves a single lexical item, e.g. the variation of *o* and *u* in the standard *weorold* vs. the variant *woruld*), morphological (such as when superlative ending varies between *–ast, –ost, –ust,* and *–st*) or phonological (such as the variation of *o* and *a* before *n*). And even if a variation is deemed systemic and unconditioned, its definition should be specific, involving as few variant graphemes as possible (a pair, or a meaningful set, like "back vowels", rather than "any vowels").

---

[30] Or, more precisely two, since each vowel in OE has both short and long counterpart.
[31] We will not try to come up with a definitive answer to the question about the status of such graphs – they may be considered allographs or graphemes, but the theoretical distinction is not important for us here.

3.4.   Spelling Variation and Morphological Analogy

The third important factor reflects the results of the operation of analogy in OE inflectional system. Even in the highly standardized modern language, certain words are inflected according to more than one paradigm and thus create "grammatical variation", which usually entails one or all of the variants being stylistically, geographically, socially or otherwise marked (consider pa. ptc. *got* and *gotten* and pret. *dived* or *dove*). Again, such types of markedness are not our concern here, but we need to make sure that all the variant forms come under the same lemma and carry the same grammatical functions. Therefore, this type of variation will have to be dealt with on the level of single lexemes or single inflectional paradigms.

3.5.   Standardisation in Modern Secondary Sources

All the problems with variation(s) are only exacerbated by the fact that while modern editors, grammarians and lexicographers naturally need some level of standardization to achieve their specific goals, each author may (and often does) decide for or formulates a different standard. While these standards cannot be completely authentic, they are usually based on some of the OE local standards, most usually on the West Saxon standard. Sometimes (mostly in case of later works), the reliance is on the late West Saxon standard, which influences the largest portion of the remaining OE texts but its inflection which manifests a much eroded state of morphological inflection. Sometimes (especially in the case of earlier works) the reliance is on the early West Saxon standard, in which there are very few texts surviving but in the opinion of modern scholars represents a useful example of well-preserved inflectional system.[32] It is important to pay close attention to

---

[32] The choice of a minor dialect in terms of the surviving evidence may seem odd and it certainly is unfortunate for our purposes, but it is a logical one in the light of the often didactic aims the authors have

the choice of the standard by the authors of the reference materials the project requires, since differences between the choices of the different authors may mean incompatibility of the forms derived from the choices (e.g. many forms in Wright do not match corresponding forms in *BT* simply because Wright spells e.g. conservatively *ie* while *BT* prefers a more progressive spelling, namely *i*)[33]. It is equally important to pay close attention to the differences between these modern standards and the OE vernacular, or the analyser may not be able to analyse real OE usage (as far as the *i/ie* variation is concerned, most surviving OE forms spell with *i* rather than *ie*, probably because by the period most of our textual evidence comes from, *ie* was already a spelling at odds with the actual pronunciation).

3.6.   Phonemic Length

Both consonantal and vowel length are phonemic in OE, as may be demonstrated by minimal pairs such as *fremme* (1. p. sg. ind. pres. "I advance") vs. *freme* (imp. sg. "advance!") or *ác* ("oak") vs. *ac* ("but"). While consonantal length exhibits some variation (probably both in pronunciation and writing), its indication by doubling the long consonant can be considered standard. The problem of vowel length is in transcription – in OE, length was hardly ever signalled and if so, then by doubling the vowels. The diacritics marking long vowels (be their macrons as in Latin, or accents as in several modern European languages) are added by modern editors, grammarians and

---

in mind. Consider what Wright says on the matter in the Preface to his Old English Grammar: "There can hardly be any doubt that all practical teachers of the subject will agree that it is far better and easier for the student to take early West Saxon as the standard for Old English, and to group around it the chief deviations of the other dialects, than to start with a grammar which treats all the dialects as being of equal importance. For us to have treated the subject in the latter manner would have defeated the very object with which the grammar was written." (Wright, p. vii)

[33] Although in the early parts of the main volume, *BT* also spells *ie*, which is probably due to the fact that those parts of the *Dictionary* were prepared for edition by Bosworth alone.

lexicographers to help modern readers.[34] This may mean that we should forget about the actual transcription in the manuscripts and work with vowel length as if it were signalled from the beginning. However, two facts prevent us from doing this. First, modern scholars and authors do not always agree on the particular cases of vowel length (cf. e.g. Wright *duru*, but *BT dúru*). Second, OE corpora are transcribed without vowel length and probably quite sensibly so, since the minimal pairs like *ác* and *ac* were actual homographs. They could have been a source of ambiguity which should be left to the reader (or editor, translator etc.) to disambiguate them. Yet, this projects would be less user-friendly if it disregarded all vowel length devised by the modern scholars: we will therefore not consider vowel length during the analysis proper,[35] but will try to present the results of the analysis in a traditional way, with the vowel length marked.

---

[34] Though there are accent marks in some of the OE manuscripts, they seem to indicate mostly stress patterns and in any case, their usage is highly irregular and they have no bearing on the modern practice of marking OE vowel length.

[35] We tried to keep the vowel length throughout the implementation and only remove it in the very last stages of the processing (see below) so that if need be, e.g. if a large portion of a text with vowel length systematically marked were to be analysed, we may make use of the diacritics to disambiguate forms by only a minor change in the implementation.

4. Theoretical Framework

The basic idea of the morphological analyser is simple: the analysis is nothing more than string matching of individual words of the analysed texts against a list of all grammatical forms in a dictionary of forms. These forms are also associated with all their possible grammatical functions and all possible lemmata. Therefore, when a form from the text matches against a form in the dictionary of forms, the user is presented with the data concerning the matched form and the analysis is complete.

4.1.   Morphological Analysers of Modern Czech

The selected theoretical framework has been inspired by the morphological analysers of Modern Czech as proposed and implemented by Osolsobě (1996), Sedláček (1999)and Sedláček & Smrž (2001). The principal idea was that Modern Czech is a highly inflected language (typologically similar to OE) and its automatic morphological analysis is well established and described in detail.[36]

The morphological analysers of Modern Czech (as noted above) work with dictionaries composed not of full forms, but of stems and endings and the matching process therefore combines segmentation of the analysed words and the matching proper.

However, we have decided not to follow this procedure due to two major differences between Modern Czech and Old English. Modern Czech is a highly standardized inflectional language in which introflection is relatively rare. Old English is a highly varied language (regional, diachronic and scribal variation is abundant) and though it is usually also characterized as inflectional, introflection plays a core role in the paradigms of verbs,

---

[36] While the methodology is in many ways similar to the projects described earlier in chapter 1.3., those project lack the detailed description provided by the Czech projects.

nouns and adjectives. Both of these differences and especially their combination preclude an easy segmentation of OE forms "on the go", that is, during the analysis itself.

Modern Czech words tend to have a single uniform stem, a changeable "intersegment" (defined by Osolsobě as a segment between the uniform stem and the ending proper) and a comparatively large number of endings (large number per each paradigm, but more importantly, large in sum of all paradigms). Old English words tend to have a larger number of stems (e.g. five standardized forms per strong verb, but multiples of that number due to variation), a highly varied intersegment (later in this paper called boundary) and comparatively few endings (not only is the number of functions and functional slots generally smaller, but the grammatical homography inside the paradigm as well as across paradigms and even across word-classes is very common).

This means that a) segmentation of forms during the analysis would be difficult and b) to prepare a dictionary of stems and endings would require a complete survey of OE morphology including a survey of dialectal forms, scribal practices and diachronic trends in orthography. The results of such an exhaustive survey would than need to be transformed into an unprecedented dictionary of morphemes (both lexical and grammatical).

4.2. Basic Plan

Since following the approach chosen for the analyser of Modern Czech too closely seems both problematic as well as going far beyond the scope of this project, we propose a more experimental method.

There are three restrictions that we need to keep in mind:

a) the analyser should be fast enough to allow its users an almost instantaneous analysis of a word (so that it can be used to look up words in a dictionary) or a fast analysis of a short text online (to create a glossary or annotate a text);

b) we want to avoid building a complete lexicon from scratch, carrying out a comprehensive survey of OE morphology or manually annotating a corpus;

c) the final product should be free and open-source, which means that all its direct sources should be freely available too.

In view of these restrictions, we propose to build a comprehensive dictionary of inflected forms that could be directly used to match words in any OE texts without any systematic segmentation of the words.

The database will be automatically generated by a script. On the most basic level, the script will be based on standard descriptions of OE inflectional morphology and a lexical dictionary.

The analyser's operation can thus be divided into three stages:

1. Preparation of the **input data** (the lexical data and the information about the lexical items necessary for stage 2).

2. Creation of a dictionary of forms from the data prepared in stage 1 by an automatic **generator** .

3. Matching items from an OE text with items in the dictionary of forms by the **analyser** proper.

For stage 1, we need to find or create a source of the lexical data. We need to identify information that is necessary for generating inflected forms from the selected lexical

source and either select a source that already contains them or fill in the information from (an)other source(s). The sources obviously need to be in a digital format.

For stage 2, we have to formulate as exhaustive rules of OE inflectional morphology as practicable. The rules need to be expressed in a programming/scripting language that will allow us processing the data prepared in stage 1. We also have to consider such variation that is restricted in its applications (i.e. that does not affect the whole of the system) and therefore produces variants of only a limited number of lexical items.[37] This also needs to be integrated in the formulation of the inflectional rules. The database of forms will thus include a certain number of variant inflected forms.

For stage 3, we have to formulate an efficient mechanism of handling the users' input and matching it with the dictionary of forms generated in stage 2. We also need to take care of systemic variation that affects (almost) the whole of the lexical inventory.[38] This will not be implemented during the previous stage, simply because it would lead to generation of an impracticably large number of forms. Instead, the variation at this stage will be defined as "variation filters" that will allow certain varied forms match successfully against standard forms.

At the end of the process, the data need to be presented in a form tailored to the users' requirements – as a lemma, a list of grammatical features, a complete paradigm, a glossary or an annotated text.

Let's go through the stages in a greater detail before we turn to the technological aspects of the project. We will focus on the problems that need to be more precisely

---

[37] E.g. variant endings of the 2. and 3. p. sg. pres. tense verb forms.
[38] E.g. regional or diachronic variation in vowels or their orthography.

formulated before a technology is chosen and an implementation is attempted. We will also focus on points that may potentially cause greatest troubles.

4.3. Stage 1: Input Data

4.3.1. Lexical Data

Since we do not want to compile our own dictionary of lexical forms, an already existing source will have to be selected. This poses one major problem as dictionaries (or at least all the existing dictionaries of OE) present the lexical items in their base forms, rather than as stems or sequenced morphs ready for generation of inflected forms. Therefore, while we may avoid segmentation during the analysis proper, we will need to do some segmentation prior to form-generation.

However, as the overview of the OE morphology above seems to suggest, creating stems ready to receive endings is not problematic in the case of the base forms of nouns or adjectives. The dictionary form of nouns, adjectives, adverbs and numerals is the nominative singular, which for most OE paradigms has either no ending, or a simple vowel. Creating a stem is thus very simple. The introflection in these classes (due to the effects of Anglo-Frisian Brightening and i-mutation) is also well defined and may be generated similarly to variants by a simple replacement pattern. Indefinite and interrogative pronouns follow paradigms of nouns and adjectives, while demonstrative and personal pronouns (including possessives) are probably too complicated to generate anyway (but also very limited in membership and therefore making little trouble in handling manually).

It is only verbs, then, that pose a significant challenge: the verbal base form – the infinitive – always receives a non-zero ending (usually *–an* or *–ian*, but sometimes *–n*).

More importantly, the verbal introflection is much more complex and depends on a particular function and a particular paradigm. Verbs also feature complex changes on the morphological boundary between the stem and the ending.

For that reason, if the verbs come in the infinitive forms as input to the generator, they will need to be segmented before the verbal forms can be generated. This holds good both for the lexical item to be generated as well as for the paradigm example according to which it is to be generated.

To be able to segment verbs and to attach correct endings in all inflected word-classes as well as to generate linguistically adequate variant forms, we need some basic grammatical information about each lexical item that is to be processed. Specifically, we need some grammatical information to correctly assign inflectional patterns, or in other words, paradigms.

4.3.2. Non-Lexical Data

Based on the overview of OE morphology above, it seems that in the case of adjectives, their morphological build-up (i.e. number and weight of stem syllables, quality of root vowels, stem-final sounds or derivational suffixes) should be enough to determine their inflectional patterns.

For nouns, the major factors in determining the inflectional paradigms are gender (masculine, feminine or neuter) and the morphological structure of the word (stem syllable weight, quality of root vowels, specific endings of the base form or derivational suffixes). Even these, however, may not, in some cases, be enough to correctly determine the inflectional pattern of a noun. Based on the gender and the stem weight, *cyn* may be

expected to inflect like *hof* (both neuter, short stem), but in fact it behaves more like *stán* (masculine, long stem).[39]

In the case of verbs, the basic type of a verb (strong, weak, preterit-present or anomalous), the membership in one of the (historically) seven ablaut-grade classes of strong verbs and the morphological structure (root vowel, ending of the infinitive) are all important factors to determine their inflectional paradigm. For many verbs, however, the paradigm is impossible to determine based on these properties alone.

With the exception of numerals (where we may need to distinguish between cardinals and ordinals), the inflectional patterns of the rest of the inflected word-classes are determined solely on the basis of their word-class membership.

To sum up: in addition to the lexical items as such, we need to know their word-class, gender (with nouns) and class (with verbs). In addition, however, we need some other principle on which to determine the inflectional behaviour of nouns and verbs.

Information on nominal gender is a standard feature of every OE dictionary and the only complication may be its retrieval. Verbal classes are marked by some of the OE dictionaries (like Clark Hall), but not by others.

The dictionaries like *BT* may also list several of the inflected forms for a given lexical item, but such additional information on the inflection of individual lexical items is only sporadic and inconsistent in form (i.e. difficult to systematically retrieve). The only resources to provide such information are consistently given are OE grammars.

It seems clear, then, that the generator must be based on two sources:

---

[39] With the exception of gemination in some cases, see 7.5.1.

- a digitized OE dictionary with a consistent and machine-readable information on nominal gender and verb-class membership;

- and an OE grammar with consistent and machine-readable information on inflectional patterns of as many lexical items as possible.

Moreover, the information on paradigm membership retrieved from the OE grammar will have to be associated with individual lexical items in the OE dictionary – i.e. the lexical items in both resources need to be paired.

4.4.   Stage 2: Generator

The first part in the generation process will have to be an unambiguous assignment of inflectional patterns to all lexical items (of inflected word-classes) based on the information from stage 1, so that the appropriate rules for the generation itself can be correctly applied.

The rules necessary to process the data from stage 1 will be based mostly on the overview of OE morphology above. The upshot of the overview is that for all the word-classes where the generation is feasible (with the exception of verbs) the inflected forms are mostly predictable from the base form alone. The rules will then consist mostly of concatenations of the base forms from the dictionary and appropriate endings with some modification to the stem. The point of the modifications being mainly the creation of variant forms. The information on endings may come as an external input in the form of a paradigmatic dictionary or may be supplied together with the rules directly in the code of the generator.

Verbs, as previously mentioned, will need a specific approach. To predict inflected verbal forms, at least three forms need to be known in the case of weak verbs (the

infinitive, pres. ind. 2. p. sg. and pret. ind. 1. p. sg.) and four principal parts in the case of strong verbs (the infinitive, pret. ind. 1. p. sg., pret. ind. pl. and past part.). The rest of the forms can be derived partly by a simple concatenation of endings, partly by replacing the stem vowel (or diphthong).

For this reason, each verbal paradigm is not characterized only by its endings and specific variations, but also, in the case of all strong and some weak verbs, by a set of the three or four vowels. As already noted above, not only do we need to store segmented information for each paradigm, but the individual verbs will need to be segmented during the generation process as well.

Consider, for example, the strong verb *gieldan* ("to yield"), which conjugates like the paradigm example *helpan* ("to help"). For us to generate the respective inflected forms of *helpan*, we record the following set of four vowels: *e, ea, u, o*. The last three of them need to replace the stem vowel of *gieldan* in the particular inflected forms.[40] This could be implemented as a simple replacement, similar to those in variants of nouns or adjectives.

Unfortunately, there is another catch – the changeable boundary preceding the ending. This is either a part of the stem (and then we call it "boundary"), or the dental part of the ending signalling past tense of weak verbs (and then we call it "dental"). This means that, with some verbs, we need to take the base form as a string and replace up to four sub-strings with the information recorded for the particular paradigm.[41] Replacing the correct parts of the string may prove complicated. The information on the paradigm should

---

[40] The first vowel is the stem vowel of the infinitive and strictly speaking does not have to bee stored, since we are working from the infinitive form. Moreover, the vowel of the infinitive can differ between the members of the same paradigm. We should store the information, however, because the difference between the infinitive vowel in a member verb of the paradigm and the paradigm example (which usually offers a better representation of the majority of the paradigm membership) signals a potential operation of analogy and a variant infinitive (containing the vowel from the paradigm example) should also be generated – in this case, the variant infinitive would be *geldan*.

[41] Or, in fact, five sub-strings, since past participles in some paradigms may require a specific prefix.

therefore be stored segmented so that during the generation process, the program does not need to segment both the lexical item as well as the paradigm example during the generation process.

To sum up, a dictionary of segmented verbal paradigms will be constructed in stage 1 recording at least five segments of at least three or four paradigm parts, as described in the preceding paragraphs. The verbal paradigmatic dictionary will be another of the generator's inputs.

All the generated forms and their particular grammatical functions will be stored in a single uniform table to function as an input for the analyser proper in stage 3. Apart from the grammatical functions usual for each word-class, the generator may also add all the information collected about each form so far: that is word-class and sub-class membership, paradigm membership, references to the source OE dictionary and grammar, standard/variant status, comments on provenance or specificities of particularly variant forms and, perhaps, also some information on the "probability" of a given form. While this project does not encompass the construction of a disambiguator that will ultimately have to work also on other than morphological level, it is clear that some information from the morphological level may be helpful for later disambiguation. The probability may thus be an experimental variable that need only to be better defined, once the operation of the disambiguator is properly established. At this stage, variant forms (e.g. some analogous forms) already deemed less probable may be assigned specific probability values, nothing more.

4.5. Stage 3: Analyser

Stages 1 and 2 only need to be carried out once, or whenever the input data change. The processes of stage 3, on the other hand, will run on users' request. As much pre-processing of the input data as can feasibly speed up the processes of stage 3 should therefore be carried out in the preceding stage. After all, one of the ideas behind this design is that a simple lookup in a large number of forms is technologically simple and can be fast even on very large data – unlike segmentation and complicated search & replace operations on much smaller data.

Stage 3 is, therefore, theoretically rather simple. It has to to provide the user interface that can accept OE text from the user, match it with the pre-generated data and present the results in a purposeful way.

The only real complication of this stage are the systemic variations. First, these may be computationally costly and cancel the advantages of the abovementioned design. Second, they may create a lot of "noise", or unexpected and unwanted matches, since it is difficult to predict e.g. what homonyms a truly system-wide change of a vowel may bring about. The formulation of these system-wide variations therefore needs to limit their effects as much as possible. That can be achieved 1) by only applying them when there is no match without them applied; and 2) by restricting their application only to lexical items of a certain type (e.g. only to a particular word-class) , or to only some parts of these items (e.g. only to the end of the string).

The results may be further enriched with data from other sources, depending on the purposes of the analysis. If the purpose is to improve the search capabilities of an online dictionary, the results are adequate as they are (i.e. a matched lemma or a link to the

dictionary item). If an automated glossary is to be generated, definitions or equivalents can be acquired from *BT* and online reference resources like grammars can be linked to the data. In the case of automatic annotation or lemmatisation of corpora, disambiguation may be necessary, be it through a future project of an automatic disambiguator, or through a user interface allowing some semi-automatic processing of the data.

5. Technology

Three factors contributed to the choice of the technology:

1. The project is built on free and open-source technology. Since all the input data are to be public-domain, all results of the project can (and will be) made publicly and freely available.

2. The technology used for the project should be widely used allowing for future modification of the project by others, as well as multiplatform so that it can reach as wide an audience as possible.

3. The tools should be adequate for the project goals they are to accomplish. That means they should be easy to use to achieve these goals both by the authors and by the users. This especially entails their efficiency in dealing with the functions and data in question.

All the input and output data of the generator are stored in tab delimited text-files in unsigned UTF-8 encoding that can represent every character of the Unicode[42] set. This should ensure that the special characters used to transcribe OE need not be transformed to conform to some more limited encoding, as was and often still is the case in projects concerning OE. This in turn aims at improving the longevity of the data and limiting errors from unnecessary data conversion.

The generator script is programmed in PERL 5.16. PERL is a high-level programming language well-known for its qualities in text manipulation and commonly used by linguists. It is a natural choice for a script whose major functions are loading, transforming and storing

---

[42] Unicode is the current industry standard for handling electronic text in most of the world's past and present writing systems.

strings of texts. PERL's strong implementation of regular expressions used here to match and replace strings is one of the important factors for its choice as well as its support of the Unicode standard.

To promote the online use of the project, the exported data are also stored in a MySQL database. MySQL is a relational database allowing data to be queried more efficiently than in the case of a simple text file. The tables of the database run on both the MyISAM and the InnoDB engines. The first one is used where full-text searching is concerned while the second one comes to play where performance for other operations is important. The reason for this division is that InnoDB has only acquired the full-text indexing (and, therefore, fast and efficient full-text search as well) only very recently and the university servers hosting the project are not yet updated to the latest version.

The analyser script is programmed in PHP – a server side scripting language widely used for dynamic web content and for processing database queries, often in concert with the MySQL databases. It is in many ways similar to (and partly derived from) PERL and can thus make use of some existing procedures already used in the generator. Also, it should be relatively easy to transform the script into PERL if an offline analyser should ever be needed (e.g. for annotating large corpora).

The user interface is in HTML with some JavaScript used for client-side scripting (e.g. for inserting special symbols by mouse or displaying reference sources on mouse-click).

Last but not least, MySQL, PHP and JavaScript are also the technologies behind the Anglo-Saxon Dictionary Online (online *BT*), which will allow a seamless integration of the two projects.

6. Processing the Input Data

The actual implementation logically follows from the chosen theoretical framework and the selected technology. There are four basic stages of the implementation discussed below: 1. collection and processing of the lexical dictionary; 2. definition of the morphological rules and their association with the lexical items; 3. generation of word-forms based on the lexicon and rules; 4. construction of the interface for textual analysis.

6.1.   Lexicon

The lexical dictionary should cover as much of the OE lexis in as many variants as possible. Since the project will not cover word-formation in any comprehensive way, as many derived lexical items as possible should be included in the dictionary, otherwise the derived items would be unanalysable. The lexicon should also include as much grammatical information about the lexical items as possible. This makes for the two main requirements for the source of the lexical information: size and machine-readability of the entries (to decode the grammatical information) that govern the choice of the source of the lexical dictionary.

6.1.1. Choice of the Dictionary

The obvious choice of the lexical dictionary is one of the specialised dictionaries of Old English, since all other sources lack significantly in one of the two requirements – including the well known the *Thesaurus of Old English* (Kay, Edmonds, Roberts, & Wotherspoon, 2005). While the corpora lack the grammatical information, the OE forms in the *OED* represent only the subset of the OE lexicon that survived to ME.

There are only three dictionaries of Old English more generally in use today: Bosworth-Toller's *An Anglo-Saxon Dictionary* (Bosworth, 1921), Clark Hall's *A Concise Anglo-Saxon Dictionary* (Hall, 1894), and the *Dictionary of Old English* (*DOE*) (Healey, 2014). While the *DOE* is not only fully digitalised but also the most recent and comprehensive of the three, it is also unfinished (by 2014, letters A-G were published) and will probably remain so for several years, if not decades. Clark Hall, on the other hand, is the least comprehensive one (some 432 pages including the later enlargement and supplement), but since it is intended as a student dictionary, it presents the information in a clear and consistent way and is therefore easy to parse. BT, with over 2,000 pages, remains the largest complete OE dictionary even after more than a century after its inception. And its age has its advantages too – with the exception of Alistair Campbell's "Enlarged addenda and corrigenda" from 1972, the main bulk of the text is safely in public domain. All this qualifies *BT* as the natural choice for the source of the lexical data, one of its major drawbacks has been apparent from the beginning. Its very name "*An Anglo-Saxon dictionary, based on the manuscript collections of the late Joseph Bosworth*" suggests the source of its inconsistencies – an unfinished project brought to completion by several authors and editors. In fact, *BT* was never published as intended. It has always existed as the main body of the dictionary largely based on Bosworth's manuscripts and on almost as much text in the supplements by Toller and Campbell containing both new entries, as well as additions, corrections and even deletions related to the entries of the main volume. No wonder that the inconsistencies trouble the microstructure of the entries at least as much as its macrostructure. Together, these make any automatic retrieval of the necessary data much more challenging.

6.1.2. Processing the Lexicon

The digitization of the two paper dictionaries of OE (Clark Hall & *BT*) was initiated in 2001 by Sean Crist and the *Germanic Lexicon Project*. We joined the initiative later on[43] mostly to support it by providing the hosting and by organizing manual corrections of the OCRed data.[44] In 2010, with the future use of the digitized data in mind, the *Anglo-Saxon Dictionary Online* website at www.bosworthtoller.com was set up. The site runs not only as a free dictionary service for the Anglo-Saxonists and the general public, but also as an evolving lexical depository. Since 2010, a large number of corrections have been submitted to the site, the typographical tagging has been transformed to unambiguously reflect the microstructure of the dictionary and, as a result, it has been made possible to store some of the grammatical information about the lexical items in a machine-readable format.

The last two steps are crucial for the design of the analyser. It is the machine-readability (or unreadability) of the original dictionary that dictates what types of information are available without a disproportionate amount of manual work required for its extraction – and, therefore, what kind of information the analyser will be able to use. Spelling variants and parts of the inflectional paradigm provide examples of the information we would deem most useful for the analyser's operation. To retrieve it, however proved too difficult at the time. Due to the inconsistency in the dictionary's microstructure and the pervasive ambiguity of its typography, there is no clear marking of this kind of information and an attempt to retrieve it through comparison with the headwords also failed because many

---

[43] See (Tichý, 2007)diploma thesis
[44] A Jan Hus Educational Foundation (JHEF) grant was awarded for this purpose for the duration of the project in 2001-2007.

of the variants are quoted in an incomplete form (sometime the variants are spelled in full, sometimes only the differing elements are provided).

Those parts of the microstructure that were typographically marked or characterised by its position and context have been semi-automatically transformed (using a large number of successive regular expression search & replace patterns) into a machine-readable unambiguous XML format. Some errors in the transformation have been, however, unavoidable and remain in the online dictionary to be manually corrected in the future.

The kinds of information (in addition to the headwords) we were able to retrieve from the parsed dictionary and intend to use for the analyser are: a) word-class; b) gender of nouns and c) type of verbs. All other information has to be supplied from other sources.

6.1.2.1.   Linking to Wright's *Old English Grammar*

Wright's *OE Grammar* (Wright & Wright, 1914) has been used to improve the quality of the grammatical information in the online dictionary (in order to supplement items lacking the grammatical information in the dictionary itself and to double-check the rest). For this reason, it is partially present in the data exported from the *Dictionary*. Namely, each headword identifiable in Wright's index has been associated with the corresponding paragraphs of the grammar book. This has been done semi-automatically by a string comparison of *BT* headwords and entries in Wright's index. While we did not consider possible miss-association due to homonymy an important problem at the time, tests of the analyser's precision have shown that this may have been underestimated and the association may need to be re-evaluated by hand at a later stage.[45]

---

[45] Errors in Wright's index are also not uncommon.

The choice of the grammar was more dependent on its function in the online *Dictionary*, than of its function as a source for the analyser alone.[46] It is now linked to entries of the online *Dictionary* and makes more detailed grammatical information concerning the headword readily available to users. While Campbell's OE Grammar is both more comprehensive and recent than the one by the Wrights, it is also still under copyright and therefore not available freely online. Wright is both in public domain, generally respected and also more or less contemporary with *BT*, which makes the two resources theoretically more consistent.

6.1.2.2. Loading the Lexicon into the Analyser

In addition, information about the type of entries was converted into machine-readable format, so that the editorial entries which consist only of changes to already existing entries of the same headword could be filtered out.

The lexical dictionary that is used as the input of the analyser is exported from the MySQL database into a tab-delimited text file in UTF8 encoding. Each line of its 23 columns represents one lexical item. The first column contains the ID of the lexical item that identifies it with the dictionary headword and can be used to access the respective entry online by prepending the dictionary URL. For example the lexical item *abbad* ("abbot") has the *BT* ID 000028[47] and the respective entry is thus available at http://bosworthtoller.com/000028. The second column contains the string with the lemma of the lexical item (the headword). The third column contains the numbers of paragraphs from Wright (divided by semi-colons) concerning the

---

particular lexical item. The rest of the columns contain binary values, one for each piece of grammatical information. See Appendix 12.5. for a sample of the dictionary file.

The generator script loads the dictionary file into an array of named arrays (or hashes) and carries out the following preprocessing on the loaded data. [48]

To make the string comparisons in the generator itself and later in the text analysis smoother, the generator normalizes the OE strings by converting everything to lowercase, by changing all *eth* (ð) graphemes to *thorns* (þ), all occurrences of *k* into *c*,[49] moving all accents in diphthongs to the first letter[50] and splitting the word-class information of participles into present and past participles based on the ending.[51] Namely, if the lemma string ends in *nde*, the item is considered a present participle. All other participial items are identified as past participles.

6.2.   Paradigms and Rules

When speaking about an inflectional paradigm in general, we mean a set of inflected forms where the variable (inflected) parts of the forms are shared by all words following that paradigm. The paradigms may be broader (allowing for more exceptions or sub-paradigms) or more specific (allowing fewer or no exceptions). Some paradigms carry

---

[48] The array has several more fields in addition to the dictionary columns. The additional fields store extra information about lexical items that is derived later in the process (e.g. the information on inflectional paradigms).

[49] The two pairs of graphemes are for all our intents and purposes functionally equivalent and may be compared to the modern lowercase grapheme variants "**a**" and "**a**". See also chapter 3.1. above.

[50] All accentuation in OE reference books is secondary and editorial. It may differ, in some cases, from editor to editor (see chapter 3.6. above). The function of the accents is to mark vowel length and there has been some academic debate over the pronunciation of OE diphthongs. One of the consequences of this debate is that some, especially earlier, OE reference books and editions mark the vowel length on the second element of the diphthong (e.g. *BT*), while later editors (e.g. W) put the accent over the first element. Again, there is no functional difference for our purposes but standardisation of the forms is essential.

[51] *BT* treats many participles as individual lexical items, probably both due to their frequency and their formal as well as functional similarity to adjectives (see chapter 6.2.2.6.). We follow suit and treat participles as a distinct word-class.

traditional names, usually due to their historical significance (e.g. nominal i-stems, 7th class of strong verbs), but since we often define our own paradigms, we will mostly refer to them by the example words.

A paradigm example is a word whose set of inflected forms exemplifies the structure of a paradigm – such as *nerian* ("to save") for certain weak verbs or *stán* ("a stone") for many strong nouns. The choice of the actual words usually follows Wright.

Parts of the paradigm are then the individual forms that constitute each paradigm.

The inflectional paradigms are based mainly on Wright (see the above chapter for the reasons behind that choice), but forms from Campbell or *DOE* often supplement the basic paradigm with dialectal forms and *BT* sometimes supplies spelling variants not mentioned by Wright. *DOEC* was consulted mostly in case of doubt: when information in Wright, Campbell and *BT* differ or when the reference books suggest but do not specify an inflectional pattern.[52] While a comprehensive re-evaluation of the description of OE inflection on the basis of corpus data (i.e. actual usage) is beyond both the scope and the aim of this work, the errors in the ensuing automatic analysis may prompt a suitable direction towards such re-evaluation.

6.2.1.1.   Prefixes and Stems

Since prefixes (in prefixed words) and non-final free morphemes (in compounds) should not affect the paradigm affiliation, we separate them to make the string comparison of the rest of the word during the process of paradigm assignment simpler. Therefore, all the dictionary items are searched for a hyphen or a space and all the

---

[52] E.g. for the interrogative pronouns *hwilc* ("which") and *hwæðer* (whether), both Wright and Campbell agree on strong adjectival inflection, but do not specify a particular paradigm; examples from DOEC and *BT* however point to weak inflection in a number of cases.

material that precedes the last one of these is stored separately and will be here called a "prefix".[53] The rest of the word is again stored separately and will be called here a "stem".[54] It is worth noting that the "stem" is derived from the base form, and if the particular base form features a non-zero ending, it will be part of the "stem". All the rest of the hyphens or spaces is then removed from the prefixes.

6.2.1.2. Syllable Count

The number of syllables is used while assigning paradigms in several word-classes. It is therefore carried out for the whole lexicon and stored separately with each dictionary item. Thanks to the highly phonetic nature of OE spelling, the actual process of counting syllables is relatively simple (compared e.g. to a much more complex spelling and therefore also syllable counting of PDE):

  a. the number of vowel + consonant combinations in the "stem" is counted,

  b. if the "stem" ends in a vowel, the count is increased by one.

To exemplify: *þeóden* ("king") contains two vowel + consonant combinations (*eó* + *d* and *e* + *n*) and no final vowel – it has two syllables; *wine* ("friend") contains one vowel + consonant combination (*i* + *n*) and a final vowel (*e*) – it also has two syllables.[55]

---

[53] This is our own terminology: the "prefix" may be a free morpheme(s) like *wíf* in the case of a compound *wíf-mann* ("woman") or a root like *cniht* in *cniht-hád* ("boyhood") – in case the derivational suffix was separated by a hyphen in *BT* (or even a morphologically complex form). In most cases, however, it really is a prefix in the general linguistic sense, such as *un-*.

[54] Again, this is our own terminology. In more general linguistic terms, the form can be a stem plus a base form ending – e.g. *sun* ("son") plus *u* – or a derivational suffix plus a base form ending – e.g. *-scip* ("ship") plus *e*. Often, the base form ending is zero and the form is a stem even in purely linguistic terms – e.g. *stán* ("stone").

[55] Note that we work with vowel + consonant combinations (in this order) and the initial consonant is thus ignored.

6.2.1. Verbal Paradigms

All the inflections except the verbal ones are formally trivial in that they seldom change anything in their form but the ending (with the notable exception of changes to the root vowel in a small number of paradigms). Therefore, only verbs required a more complex approach concerning their paradigms and a separate dictionary of verbal paradigms was manually collected into an external input file.[56]

The dictionary contains principal parts[57] of all paradigm examples necessary to generate all the OE verbal forms, i.e. all the paradigm examples used by Wright as well as all those that differ from them (mostly sub-paradigms and the paradigms of anomalous verbs). Strong verbs have a minimum of four principal parts, weak verbs only three, but all variant parts of the paradigm (not regularly derivable from the principal parts) and all the paradigm parts of the preterit-present and anomalous verbs are also included. Each of these parts forms a single row of 17 columns. The information in the columns is as follows:

1. **Paradigm ID**: connecting all the parts of a particular paradigm example;

2. **Lemma**: the base form of the paradigm example;

3. **Verb Type**: strong, weak, preterit-present or anomalous;[58]

4. **Wright**: if other verbs are listed as following the same paradigm, the paragraph that contains the list is recorded;

5. **Class1**: 1-7th class of strong verbs or 1-3rd class of weak verbs;

6. **Class2**: subclass of weak verbs (according to Wright)

---

[56] The material was collected from Wright, Campbell, *BT*, Clark Hall, DOE and OED with a substantial help of Michaela Hejná.

[57] For a discussion of verbal principal parts, see chapter 2.7.

[58] The "contracted" type of verbs is also recorded in *BT*, but there is no need to classify this type separately as every contracted verb also belongs to one of the other types.

7. **Class3**: further sub-classification according to Wright (e.g. *hyngrian* "to hunger" is weak, class 1, B, c)

8. **Variant**: if more forms carry the same function, they are numbered 0–n (so that the whole paradigm may be consistently displayed to the user – as the standard paradigm and its variants – and so that it may also be possible to calculate probability[59]);

9. **Function**: the grammatical function(s) carried by the part of the paradigm – each grammatical function is represented by a unique double letter code, so that the whole is both machine and human readable function tag[60] (e.g. *hyngrest* is tagged as "PsInSg2", or Present Indicative Singular 2nd person);

10. **Prefix**: while prefixed verbs are not used as paradigm examples, *ge-* prefix is recorded if it is a usual part of the past participle, likewise, *be-* is also recorded in certain present forms of *nugan* (verb attested in prefixed forms *genugan* "to suffice" and *benugan* "to need"). The prefix relates to individual paradigm parts, but is universal for all members of the paradigm (for each verb following the paradigm);[61]

11. **Pre-Vowel**: everything in the root of the verb preceding the root vowel; uniform for all paradigm parts, but specific for each member of the paradigm; e.g. *sm* in the verb *smúgan* ("to crawl");

12. **Root-Vowel**: root vowel (or diphthong); specific for each principal part, but uniform for all members of the paradigm; e.g. *ú* in the inf. of *smúgan*, or *éa* and *o* in its preterit singular (*sméah*) and past part. (*smogen*) respectively;

---

[59] Probability is discussed in chapter 4.4.
[60] See appendix 12.1. for the table of function tags.
[61] Notice that apart from this paradigm-derived "grammatical prefix", each verb may also have its "lexical prefix" (as discussed above). In such cases, the grammatical prefix would come second, the lexical first as in e.g. *un-ge-lǽred* ("ignorant", negated past. part. of *lǽran* – "to teach").

13. **Post-Vowel**: everything following the root vowel up to the material covered by 14 and 15 below; uniform for all paradigm parts (although it may be deleted in other forms than infinitive), but specific for each member of the paradigm; ; e.g. *g* in the inf. of *smúgan*, but *0* in the variant form of preterit indicative singular 1st person *sméah*;

14. **Boundary**: changeable material on the morphological boundary between the stem and the ending; specific for each principal part, but uniform for all members of the paradigm; e.g. *h* in *sméah*;[62]

15. **Dental**: the dental affix of past forms of weak verbs: *–ed*, *–od*, *–d* or *–t*; specific for each principal part, but uniform for all members of the paradigm;

16. **Ending**: the ending; specific for each principal part, but uniform for all members of the paradigm;

17. **Comments**: comments about dialectal or diachronic variants were recorded here, but since the dialectal or diachronic description is not one of our aims and this type of information was not recorded for other word-classes, the field is currently not displayed to end-users and remains for future use;

---

[62] The difference and significance of the Post-Vowel and the Boundary is discussed below in chapter 7.2.

ID	Lemma	Type	C1	C2	C3	Wright	Variant	Function	Pref	PreV	V	PoV	Bnd	Dent	End
73	nerian	w	1	A	a	525	0	Inf	-	n	e	r	-	-	ian
73	nerian	w	1	A	a	525	0	PsInSg2	-	n	e	r	-	-	est
73	nerian	w	1	A	a	525	0	PaInSg1	-	n	e	r	-	ed	e
73	nerian	w	1	A	a	525	1	PaInSg1	-	n	e	r	-	od	e
73	nerian	w	1	A	a	525	1	Inf	-	n	e	r	ig	-	an
73	nerian	w	1	A	a	525	2	Inf	-	n	e	r	g	-	an
73	nerian	w	1	A	a	525	3	Inf	-	n	e	r	ige	-	an
73	nerian	w	1	A	a	525	4	Inf	-	n	e	r	-	-	an
73	nerian	w	1	A	a	525	1	PsInSg2	-	n	e	r	-	-	es
5	béodan	s	2	-	a	493	0	Inf	-	b	éo	d	-	-	an
5	béodan	s	2	-	a	493	0	PaInSg1	-	b	éa	d	-	-	-
5	béodan	s	2	-	a	493	0	PaInPl	-	b	u	d	-	-	on
5	béodan	s	2	-	a	493	0	Pp	-	b	o	d	-	-	en

Table 29 Example of verbal paradigms input. Verbs *nerian* ("to save") and *béodan* ("to bid").[63]

The dictionary is loaded into a multi-dimensional array of hashes with each paradigm being represented by one element of the array, its properties (1–7 of the above list) being the elements of the hash linked to the array element. The "variant" element of the hash links to another array of hashes, where each element of the array represents one variant of the paradigm and links to a hash containing the individual forms and functions (8–17 of the above list). The whole array may be represented by the following table:

	ID	Lemma	Type	C1	C2	C3	Wright	variant	Function	Pref	PreV	V	PoV	Bnd	Dent	End
73 {	73	nerian	w	1	A	a	525		Inf	-	n	e	r	-	-	ian
								0	PsInSg2	-	n	e	r	-	-	est
									PaInSg1	-	n	e	r	-	ed	e
									Inf	-	n	e	r	ig	-	an
								1	PsInSg2	-	n	e	r	-	-	es
									PaInSg1	-	n	e	r	-	od	e
								2 {	Inf	-	n	e	r	g	-	an
								3 {	Inf	-	n	e	r	ige	-	an
								4 {	Inf	-	n	e	r	-	-	an
5 {	5	béodan	s	2	-	a	493		Inf	-	b	éo	d	-	-	an
								0	PaInSg1	-	b	éa	d	-	-	-
									PaInPl	-	b	u	d	-	-	on
									Pp	-	b	o	d	-	-	en

Table 30 Multi-dimensional array of hashes holding the verbal paradigms.

---

[63] See appendix 12.2. for the complete list of verbal paradigms.

Since the keys of the first array correspond to paradigm IDs and the keys in the second array correspond to the numbering of variants, it is easy to access the values of the paradigmatic dictionary (`$vparadigms[73]{lemma}` returns the lemma od the paradigm 73: "*nerian*", `vparadigms[73]{variant}[1]{PaInSg1}{Dent}` returns the dental ending of the preterit indicative singular 1ˢᵗ person of the first variant of the paradigm 73: "*od*"). Each lexical item has an element `vb_paradigm` in the lexical dictionary array (see 6.1.2.2 above). This again is linked to an array (each verb can follow multiple paradigms) and elements of this last array are linked to the top array of the paradigmatic dictionary during the assignment of paradigms. In other words, each verb is assigned one or more paradigms together with all the accompanying information on that paradigm including the individual parts of the inflected forms, their functions etc.

6.2.2. Assignment of Paradigms

Paradigm examples and inflectional material for adjectives and nouns is supplied directly in the source code of the analyser.

In technical terms elements of arrays linked to the elements `adj_paradigm` and `noun_paradigm` of the lexical dictionary do not link to dictionaries of paradigms, as is the case of `vb_paradigm`, but contain simple strings directly identifying the adjectival or nominal paradigm examples.

This proved beneficial during the development stages, as it was easy to spot and correct errors, but separation of this type of data into external input files similar to the verbal paradigms might be advisable before the application is more widely distributed, so that both lexical and grammatical material may be added or modified without the necessity to

touch the source code itself (and because of the general advisability of code/data separation).

The assignment of paradigm examples (or the paradigms they represent) to individual lexical items is carried out in several stages, separately for each word-class. The order of the stages is given mainly by the reliability of information upon which the decisions about the paradigm affiliation can be made.

It is important to remember that, with OE, the paradigm membership of lexical items is more ambiguous and also more conjectural than is the case with highly standardized and well described varieties of modern languages. Due to the diachronic and dialectal variety of OE, words may exhibit conflicting paradigm membership; because of the scarcity of evidence, there may often be no ground on which to securely base an inflectional pattern of a particular lexical item.

Moreover, the information that the analyser has to do with is incomplete. The analyser has to depend on:

1. the lemma as a string of characters;

2. word-class (each item may belong to any number of word-classes);

3. gender affiliation of nominal items (each noun may belong to any combination of the three genders);

4. strong/weak/preterit present/anomalous class membership of verbal items (any verb may belong to any combination of the classes).

While we may perceive 1. and 4. as highly reliable, 2. and 3. have been semi-automatically derived from *BT* and may be prone to some errors.

The stages of the paradigm assignment are as follows:

1. With verbs the "stems" of the lexical items are string-compared to the lemmata of the paradigm examples and the verb type of the lexical items is string-compared to the verb type of the paradigm examples.[64] If both match, the paradigm is assigned. Let us consider the example of the verb *á-déman* ("to judge"), its "stem" is *déman* and its verb type is weak. When compared to the weak paradigm example *déman* ("to deem"), both conditions are fulfilled and the verb is assigned (i.e. *a-déman* conjugates like *déman*).

2. The first step is improved on by going through all the unassigned verbs and string-comparing the beginnings of their "stems" with items in the list of verbal prefixes (derived from *BT*). If there is a match and the rest of the "stem" is then successfully compared to a paradigm example lemma, the verb is assigned. This step is useful, since *BT* is not quite consistent in marking prefixes and suffixes with hyphens, especially if there is more than one prefix. A case in point can be the verb *ge-in-settan* ("to institute"), which is assigned to the paradigm example *settan* ("to set") (*BT* marks its first prefix only).[65]

3. When stem comparison can yield no more results, the unassigned lexical items with links to Wright's paragraphs are assigned paradigms corresponding to those paragraphs.

Since the other inflected word-classes beside verbs do not have an external set of paradigms, this is the first step in which their lexical items are assigned. The comparison is therefore not with paradigm examples loaded from a file, but

---

[64] This prevents some homonyms to be mistakenly assigned.

[65] Even though detecting the unmarked „prefixes" for all lexical items while loading and processing the lexicon may seem like a good idea, that could lead to unwanted results. Matching a prefix with the beginning of a lemma is not a safe way to detect a real prefix in the lemma: as far the the example of *a-déman* is concerned, there are many lexical items starting with *a-* that have nothing to do with the *a-* prefix. Therefore, it is crucial to compare the rest of the lemma with the other existing lexical items and only if there is a match there too, the prefix can be safely detected.

each lexical item is searched for one (or more) paragraph number(s) and assigned a corresponding paradigm directly in the source code.

4. With a group of newly assigned items, the stem comparison from step 2 is repeated. In this way, even items derived from those mentioned in Wright, but not mentioned themselves, can be assigned.

5. At this point, the more reliable information based on Wright and stem similarity is exhausted and the stems themselves are analysed for phonological or morphological clues as to their paradigm affiliation. This obviously differs for each word class and will be described in the following sub-chapters. Due to the nature of the algorithm, all adjectives should be now assigned.

6. With each group of newly assigned items, the stem comparison from step 2 and 4 is repeated, but this time the differences between *i*, y and *ie* are ignored.[66] In this way, even items derived from those that were stem-analysed in the previous step, can be assigned.

7. The rest of the unassigned items is assigned according to the following rules:

   a. All the unassigned strong verbs are assigned to *helpan*; the rest of the verbs is then assigned to *déman*.

   b. All the unassigned masculine nouns and all the unassigned nouns of uncertain gender are assigned to *stán*, all the feminine nouns are assigned

---

to *ár* and all the neuter nouns according to their stem length.[67] Neuter long stems follow *stán*, short stems follow *hof*.

6.2.2.1. Assignment of Nouns

The algorithm for paradigm assignment of nouns on the grounds of the phonological and morphological analysis of the "stem" is largely based on Wright (chapters on Accidence and Word-formation). The decision process relies mostly on the ending of the lemma, derivational suffixes, quantity and quality of the root vowel and the syllabic structure of the "stem". Where the process relies on similar elements in more than one step, these need to follow from the more to the less specific ones (i.e. order matters), since the steps are mutually exclusive (the `if, elsif` statement is used by the program).

1) Stem ends in *–a*[68] and
   a) its root vowel is short: assigned to *guma*;
   b) its root vowel is long: assigned to *fréa*.
2) Stem ends in *–e* and
   a) is of feminine gender: assigned to *tunge*;[69]
   b) is of masculine gender: assigned to *stán*;[70]
   c) is of neuter gender: assigned to *hof*.[71]
3) Stem ends in *–nd* and is of masc. gender: assigned to *wígend*.
4) Stem ends in *–els* or *–scipe*: assigned to *stán*.
5) Stem ends in *–incel*: assigned to *hof*.
6) Stem ends in *–ness, -niss* or *-ung*: assigned to *ár*.
7) Monosyllabic stem of masc. gender and the root vowel is *–æ–* or *–ǽ–*: assigned to *dæg*.
8) Monosyllabic stem of neut. gender and the root vowel is *–æ–* or *–ǽ–*: assigned to *fæt*

---

[67] The length of a stem is based on the length (sometimes weight) of its first syllable. The length is calculated on demand as follows:
  1. If the first vowel/ diphthong (from the left) is long, the stem is long.
  2. If the first vowel/diphthong is followed by more than one consonant, the stem is long.
  3. If the stem is monosyllabic (see chapter 6.2.1.2) and ends in a consonant, the stem is long.
  4. Otherwise, the stem is short.

[68] All nouns ending in *–a* are weak.

[69] Note that all the other weak nouns follow *béo* or *éagan* and are already assigned based on the links to Wright's paragraphs.

[70] Traditionally já/i stems following *wine* and *ende*.

[71] Traditionally já/i stems following *spere* and *wíte*.

Of the total of 20,408 nouns, 1,096 are assigned according to Wright, 6,024 by stem analysis, 5,274 by stem comparison and 8,014 in the last step according to gender and (in case of neuters) their stem length.

The number of nouns assigned to each paradigm is summarized in the following table:

paradigm	nouns	paradigm	nouns	paradigm	nouns	paradigm	nouns
ár	6045	fæt	382	béo	12	*mǽd*	*2*
stán	5588	feld	150	duru	9	*sweostor*	*2*
word	2352	cynn	115	beadu	5	*dóhtor*	*1*
guma	1602	bealu	79	*lamb*	*5*	ealu	*1*
hof	1144	sunu	65	éage	4	*fæder*	*1*
tunge	1123	bearu	38	*fót*	*4*	hæleþ	*1*
wígend	562	hand	30	*fréond*	*3*	*mægþ*	*1*
dæg	419	strengu	23	*hnutu*	*3*	módor	*1*
fréa	401	*bóc*	*16*	*bróþor*	*2*	mónaþ	*1*

Table 31 Number of nouns following each paradigm (paradigms in italics were introduced by manually inflected forms).

6.2.2.2. Assignment of Adjectives

The algorithm for paradigm assignment of strong adjectives and interrogative & indefinite pronouns[72] on the grounds of the phonological and morphological analysis of the "stem" is largely based on Wright (chapters on Accidence and Word-formation). The decision process relies mostly on the ending of the lemma, derivational suffixes, quality of the root vowel, the syllabic structure of the "stem" and its length. Where the process relies on similar elements in more than one step, these need to follow from the more to the less specific ones (i.e. order matters) since the steps are mutually exclusive (the `if`, `elsif` statement is used by the program).

1) Stem ends in *þweorh*: assigned to *þweorh.*

2) Stem ends in *–sum*, *–isc* or *–lic*: assigned to *til.*

3) Stem ends in *–cund*, *–feald*, *-fæst*, *–léas*, *–full* or *–iht*: assigned to *blind.*

4) Stem ends in *–ihte*, *–bǽre*, *–ede* or *–wende*: assigned to *wilde.*

---

[72] And multiplicative numerals ending in *–feald.*

5) Monosyllabic and

    a) short stem with

        i) –æ– as the root vowel: assigned to *glæd;*

        ii) a root vowel different from *æ*: assigned to *til.*

    b) long stem ending in

        i) a vowel + *h*: assigned to *héah;*

        ii) a consonant + *h*: assigned to *þweorh;*

        iii) anything other than –*h*: assigned to *blind.*

6) Polysyllabic stem ending in

    a) –*u* or –*o*: assigned to *gearu;*

    b) –*e*: assigned to *wilde;*

    c) anything other than –*e* , –*u* or –*o;*

        i) and short stem: assigned to *hálig;*

        ii) and long stem: assigned to *manig.*

Of the total of 6,238 adjectives (including some pronouns and numerals),[73] 307 are assigned according to Wright, 4,338 by stem analysis and 1,593 by stem comparison.

The number of adjectives assigned to each strong paradigm is summarized in the following table.

paradigm	adjectives	paradigm	adjectives	paradigm	adjectives
blind	1695	hálig	1066	gearu	38
til	1409	manig	666	glæd	11
wilde	1264	héah	59	þweorh	7

Table 32 Number of adjectives following each paradigm.

---

[73] See below under pronouns and numerals.

6.2.2.3. Assignment of Pronouns

Most pronouns are inflected manually, but interrogative and indefinite pronouns are inflected automatically and follow adjectives in the assignment of their paradigm as well as in the generation of forms. Accordingly, they are marked as belonging to both word-classes, pronouns as well as adjectives, already in the *BT*.

6.2.2.4. Assignment of Verbs

The algorithm for paradigm assignment of verbs on the grounds of the phonological and morphological analysis of the "stem" is partly based on Mitchell and Robinson (pp. 52–53) and partly on Wright (Accidence). The decision process relies mostly on the type of the verb (strong/weak), quality and the quantity of the root vowel as well as on the surrounding consonants and the derivational suffixes. Where the process relies on similar elements in more than one step, these need to follow from the more to the less specific ones (order matters), since the steps are mutually exclusive (the `if, elsif` statement is used by the program).

Strong verbs are assigned as follows:

1) Root vowel is *–í–* or *–ý–* followed by

   a) *–þ–*: assigned to *sníþan;*

   b) *–s–*: assigned to *rísan;*

   c) any other consonant beside *–þ–* or *–s–*: assigned to *bídan.*

2) Root diphthong is *–éo–* followed by

   a) *–s–*: assigned to *céosan;*

   b) any other consonant beside *–s–*: assigned to *béodan.*

3) Root vowel is *–ú–* followed by any consonant: assigned to *brúcan.*

4) Root vowel is –*e*– followed by

   a) –*s*– and one more consonant: assigned to *helpan;*

   b) any other consonant beside –*l*– and one more consonant: assigned to *bregdan.*

   c) –*l*– or –*r*–: assigned to *beran;*

   d) –*c*–, –*d*–, –*f*–, –*g*–, –*p*–, –*s*–, –*t*– or –*þ*–: assigned to *metan.*

5) Root vowel is –*a*– followed by a consonant: assigned to *faran.*

6) Root diphthong is –*eo*– followed by –*h*– or –*r*– and one more consonant: assigned to *weorpan.*

7) Stem begins with *gie*– followed by two consonants: assigned to *gieldan.*

8) Root vowel/diphthong is –*i*–, –*ie*– or –*y*– followed by a nasal and one more consonant: assigned to *bindan.*

9) Root vowel is –*á*– : assigned to *hátan.*

10) Root vowel is –*ó*– : assigned to *blótan.*

   Weak verbs are assigned as follows:

   1. Stem ends in –*rian*: assigned to *nerian.*

   2. Stem ends in –*ian*: assigned to *sealfian.*

   3. Root vowel is short and followed by a doubled consonant: assigned to *fremman.*

   Of the total of 6,960 verbs, 733 are assigned according to Wright, 2,541 by stem analysis, 2,436 by stem comparison and 1,250 in the last step according to gender and (in case of neuters) their stem length.

   The number of verbs assigned to each paradigm is summarized in the following table.

paradigm	verbs	paradigm	vbs.	paradigm	vbs.	paradigm	vbs.	paradigm	vbs.
sealfian	2342	hyngrian	23	hycgan	10	spanan	6	pæcan	3
déman	1758	sníþan	22	hebban	9	þerscan	6	streccan	3
nerian	387	bregdan	21	rædan	9	witan	6	unnan	3
bídan	226	dón	20	rísan	9	béatan	5	ágan	2
bindan	200	sléan	20	sécan	9	lácan	5	bréoþan	2
fremman	176	gangan	19	wríþan	9	munan	5	cunnan	2
drencan	123	cuman	18	libban	8	tellan	5	dreccan	2
faran	115	cweþan	18	scéadan	8	þeccan	5	dugan	2
béodan	112	twéogan	18	wegan	8	wesan	5	durran	2
helpan	108	hléapan	15	willan	8	gilpan	4	dwellan	2
metan	95	wyrcan	15	brengan	7	hnígan	4	fricgan	2
weorpan	81	gierwan	14	reccan	7	rǽcan	4	frignan	2
beran	73	hátan	14	séoþan	7	slǽpan	4	gellan	2
weorþan	57	lǽtan	14	stellan	7	smúgan	4	mígan	2
fealdan	55	giefan	13	stígan	7	tǽcan	4	mótan	2
brúcan	53	gieldan	13	stregdan	7	þyncan	4	nugan	2
téon	52	þencan	13	þryccan	7	wǽcan	4	sceþþan	2
blótan	46	búgan	12	béon	6	weccan	4	scippan	2
bláwan	40	niman	12	bycgan	6	blandan	3	stæppan	2
settan	39	swerian	12	féolan	6	cweccan	3	þurfan	2
gán	30	berstan	11	findan	6	cwellan	3	wleccan	2
céosan	28	gitan	11	ícan	6	drædan	3	birnan	1
séon	27	habban	11	leccan	6	hlehhan	3	magan	1
lǽcan	26	licgan	11	sellan	6	míþan	3	nágan	1
biddan	23	secgan	11	sícan	6	murnan	3	nyllan	1
fón	23	bannan	10	sígan	6	nytan	3	sculan	1
								spornan	1

Table 33 Number of verbs following each paradigm.

6.2.2.5.　Assignment of Numerals

Numerals higher than 3[74] are assigned the weak adjectival paradigm, as well as the nominal paradigms during the generation phase.[75] The only decision in the process is based on whether the numeral is cardinal or ordinal. Cardinal numerals are recognised by also having a noun word-class assigned already from *BT* in addition to the adjective and numeral word-classes that are common to all numerals.

---

[74] See chapter 7.6. for numerals 1–3.
[75] Technically this means that there is no paradigm assigned to each lexical item in the relevant array, but the paradigm example is inserted directly into the array of forms (see chapter 7.6. below).

All numerals are declined as weak adjectives and their paradigm is therefore *blinda*. Cardinal numerals are, in addition, declined like nominal i-stems, each in all three genders, but each gender as a different paradigm: masculine *wine* ("friend"), feminine *cwéne* ("queen") and neuter *spere* ("spear"). Note that these are not the nominal paradigms listed in chapter 2.1. on nominal morphology as our working paradigm examples. Since they are assigned purely for end-users' benefit (e.g. as a reference to grammars) and do not otherwise function in the form-generation process, we have the luxury to pick more specific and descriptive paradigm examples as mentioned above.

Multiplicative numerals ending in –*feald* are a special case and are treated together with adjectives (see 6.2.2.2.).

6.2.2.6. Assignment of Participles

Similar to numerals, participles are only assigned paradigms during the generation of adjectives.[76] All participles are declined as weak adjectives and are therefore assigned the paradigm *blinda*. Present participles are also declined strong like *wilde*; past participles are declined strong as either *manig*, if their stem is short, or *hálig*, if their stem is long.

---

[76] Technically this means that there is no paradigm assigned to each lexical item in the relevant array, but the paradigm example is inserted directly into the array of forms (see chapter 7.3.5. below).

7. Generating Forms

Form generation is again carried out class by class. Generation of verbal forms differs in that most of the inflectional material (like endings) is derived from the paradigmatic dictionary, while the material (endings) for the generation of forms of other word-classes is supplied directly in the source code.[77]

Generated forms are all stored in an array of hashes that can be described as a table with the following columns:[78]

1. *BT* **ID**;
2. **lemma**;
3. "**stem**";
4. **form** – generated inflected form;
5. **formParts** – generated inflected forms with some of the morphemes divided by hyphens to illustrate the build-up of the form;
6. **var** – number of the variant ("0" means the standard form; it also helps to keep track of variants);
7. **probability** – probability of the form, higher value means lower probability and "0" is the most probable;[79]
8. **function**;
9. **Wright**;
10. **paradigm**: paradigm example;
11. **paraID**: paradigm example ID (for verbal forms only);
12. **word-class;**

---

[77] The reason for this is explained in chapter 6.2.2. above.
[78] Most of the columns were already described in 6.2. and 6.2.1. above.
[79] Probability is discussed in chapter 4.4.

13. **class1;**

14. **class2;**

15. **class3;**

16. **comment.**

7.1.   I-mutation

I-mutation is a historical vowel change affecting all classes of generated words. It is implemented here as a function that is called on demand with the original vowel as a parameter and an array of vowels on return – note that both the original (unmutated) vowel as well as the mutated vowels are returned (variation of mutated and un-mutated vowels is common due to morphological analogy). Not also that it is impossible to predict the mutated vowels from a synchronic infinitive with complete accuracy – historical forms would be necessary to accomplish that.[80]

The following table summarizes the algorithm.

unmutated	mutated
a	æ, e
æ	e
e	i
o	e
u	y
ea	ie, i
eo	ie, i
io	ie
á	ǽ
ó	é
ú	ý
éa	íe, í
éo	íe, í
ío	íe, í

Table 34 The i-mutation.

---

[80] Historically, i-mutation is a conditioned regressive distant sound change operating on the principle of vowel harmony. In effect, a vowel was raised during the pre-OE stage if the following syllable contained *i/j*.

7.2.   Verbs

Generation of verbal forms is mostly based on the lexical dictionary and the dictionary of verbal paradigms. Material derivable from the two sources (e.g. mutated forms) or material that is not specific for individual paradigms (e.g. some of the endings) are derived or supplied directly during the generation process.

The process loops through all the verbs the in the lexical dictionary. For every verb it loops through all its paradigms. For every paradigm it loops through all its variants. For every variant it loops through all the functions (functional slots defined in the paradigmatic dictionary). At this level of the loop, individual values from the paradigmatic dictionary and the "stem" of the lexical items are used to derive most of the variables necessary to generate the verbal forms. Each major verbal type is generated separately, but some of the variables are derived for all the types at the same time. From left to right – as they potentially appear in the generated forms – the segments are:

1. **lexical prefix**: the "prefix" as specified in the lexical dictionary for the current lexical item;[81] it is all the material that precedes the last hyphen or space in the lemma;[82]

2. **grammatical prefix**: the "grammatical" prefix as specified in the paradigmatic dictionary for the current function; it is usually only specified for past participles and in most cases it is the *ge–* prefix; if the grammatical prefix is the same as the lexical prefix, only one is included in the form;

---

[81] More forms can potentially be analysable if we generate each verbal form with and without the "prefix", since *BT* lists some lexical stems only in prefixed variants. This would, however, dramatically increase the number of generated forms (mainly due to derived inflected participles) and from a cursory analysis of such items in *BT* it seems that most of them in fact do not appear in the DOEC without the prefix.

[82] See chapter 6.2.1.1. above for a discussion of the "prefix" and "stem".

3. **pre-vowel**: the pre-vowel is derived from the actual lexical item by a simple regular expression – it is all the material of the "stem" that comes before the first vowel from the left;

4. **vowel**: see below, the root vowel is derived differently for different classes;

5. **post-vowel**: the post-vowel is also derived from the current lexical item using a regular expression, but the form of the expression depends on whether the paradigmatic dictionary has a non-empty boundary for the infinitive form[83];

    a. if the inf. boundary is non-empty, the post-vowel is all the material between the first vowel/diphthong and this boundary;

    b. if the inf. boundary is not defined, the post-vowel is all the material between the first vowel/diphthong and one or more vowels followed by *n* (all verbal infinitives end in *n*);

6. **boundary**: the boundary is derived from the paradigmatic dictionary for the current inflected form;

7. **dental**: the dental affix is derived from the paradigmatic dictionary for the preterit forms of weak verbs;

8. **ending**: the ending is derived from the paradigmatic dictionary (for forms specified there) or it is specified directly in the code.

7.2.1. Strong Verbs

Strong verbs are characterised by the ablaut gradation of their root vowel. Therefore their **root vowel** is derived from the paradigmatic dictionary for the current part of the paradigm.

---

[83] Note that the boundary of the infinitive has to be used, not the boundary of the actual inflected form of the paradigm example, since it is used to match against the actual lexical „stem", which is the infinitive (and the boundary can change for different parts of the paradigm).

Ideally, all verbs following the same paradigm should have the same vowel gradation. Synchronically, this is not true about the historical 7 classes of the strong verbs (due to diachronic change), and neither is it quite true about our own paradigms – due synchronic variation and inconsistency of the standards used by Wright and *BT*. In some cases, *BT* supplies variants with a root vowel in the lemma different from Wright that should still follow the paradigm suggested by Wright.

Therefore, we first test whether the root vowel of the infinitive in the paradigmatic dictionary matches the root vowel of the actual lexical item.[84] In case of the match, the vowel from the paradigmatic dictionary is used. In case of no match, both vowels are used as variants for the infinitive and all the forms derived from it. We do not attempt to capture deviations in the root vowel of individual lexical items in the other parts of the paradigm (compared to those of their paradigm example), since we have no material to base this upon.

Forms of those parts of the paradigm defined in the paradigmatic dictionary[85] are assembled as suggested above by concatenating pre-prefix + prefix + pre-vowel + vowel + post-vowel + boundary + ending and added to the forms without any further changes (e.g. *0-a-bl-á-w-0-an* "to blow").[86]

All the other parts are generated as follows:

The inflected infinitive, pres. part., pres. ind. sg. 1. p., pres. ind. pl., pres. subj. sg., pres. subj. pl., imp. sg and imp. pl. are all derived from the **infinitive**.

---

[84] Derived by regular expression from the "stem" of the infinitive as the first vowel/diphthong from the left.

[85] For all strong verbs these are: inf., pres. ind. sg. 2.p., pret. ind. sg. 1.p. and past part., but in case of variation in the paradigm, other parts may be defined as well.

[86] Again, as suggested above, the lexical prefix, the vowel, the boundary and the ending are derived from the lexical item itself, while the rest is derived from the paradigmatic dictionary.

The infinitive and the **inflected infinitive** are distinguished by specific endings. If the paradigm's inf. ending is *–an*, the form is generated with two variant endings: *–anne* (e.g. *0-a-bl-á-w-0-anne*) and *–enne* (*0-a-bl-á-w-0-enne*). Otherwise (e.g. in cases like *téon* "to draw", which gives *0-0-t-éo-0-0-nne*), only one variant is generated – with an ending *–nne*.

Present participle behaves similarly, the ending for verbs in *–an* is *–ende*, for the rest of the verbs it is *–nde*.

The **present indicative singular 1. person** ending for verbs in *–an* is generated in four variants, the first being assigned the highest probability: *–e*, *–u*, *–o* and *–æ*. For the rest of the verbs, only one variant is generated: *–0*.

The **present indicative plural** ending for verbs in *–an* is generated in four variants, the first being assigned the highest probability: *–aþ*, *–eþ*, *–as*, *–es* and *–þ*. For the rest of the verbs, only one variant is generated: *– þ*.

The **present subjunctive singular** ending for verbs in *–an* is *–e*, for the rest of the verbs it is *–0*.

The **present subjunctive plural** ending for verbs in *–an* is *–n*, for the rest of the verbs it is *–n*.

The **imperative singular** ending is *–0*.

The **imperative plural** ending for verbs in *–an* is *– aþ*, for the rest of the verbs it is *–þ*.

The forms of the **present indicative singular 2. person** and 3. person have a different root vowel compared to the infinitive, due to the i-mutation discussed above. The i-mutation function is therefore called with the root vowel as the parameter and the generation process loops through the resulting array of vowels.

The ending is generated in five basic variants, the first being assigned the highest probability: *–st*, *–stu*, *–est*, *–ist* and *–s*. A number of successive and cumulative replacement patterns is then applied creating variants with each step of the basic forms to deal with the assimilations and reductions on the morphological boundary between the stem and the ending or those of the ending itself. We replace the ends of the forms resulting from the basic generation to create variants, so that –dst > –tst, –þst >–tst, –tst > –st, –ngst > –ncst, –ncst > –nst, –gst > –hst, –hst > –xst

The form of the **present indicative singular 3. person** has a mutated root vowel (see above).

The ending is generated in three basic variants, the first being assigned the highest probability: *–þ* *–eþ*, and *–iþ*. A number of successive and cumulative replacement patterns is then applied creating variants at each step of the basic forms to deal with the assimilations and reductions on the morphological boundary between the stem and the ending or those of the ending itself. We replace the ends of the forms resulting from the basic generation to create variants, so that –tþ/–dþ > –tt, –tt/–dt > –t, –þþ > –þ, –þ > –t, –sþ > –st, –ngþ > –ncþ, –gþ > –hþ.

It is easy to imagine that verbs with *–a–* for the root vowel and stem ending in a dental or velar might produce a large number of forms for the two preceding paradigm parts. Let us consider the forms of the pres. ind. sg. 2. p. of *standan* ("to stand") as an example:

form	probability	frequency	form	probability	frequency
standest	2	1	**stændstu**	1	0
standist	2	0	**stænst**	1	0
stands	2	0	**stæntst**	1	0
standst	1	0	**stendest**	3	0
standstu	2	0	**stendist**	3	0
stanst	2	0	**stends**	3	0
stantst	2	0	**stendst**	2	1
stændest	1	0	**stendstu**	3	0
stændist	1	0	**stenst**	3	2
stænds	1	0	**stentst**	3	2
stændst	0	0			

Table 35 Forms generated for pres. ind. sg. 2. p. of *standan.*

Notice that the frequency (taken from *DOEC*[87]) confirms that the form exists in both mutated and unmutated forms (probably as a result of analogy). Likewise, it confirms the operation of assimilation and reduction on the morphological boundary, as expected. The comparison of the frequency and the calculated probability (0 is most probable) is much less encouraging.

The form of the **preterit indicative singular 3. person** is derived from the **1. person** with *–0* ending.

The form of the **preterit indicative singular 2. person, past subjunctive singular** and **plural** are derived from **preterit indicative plural** with the endings *–e*, *–e* and *–en* respectively.

7.2.2. Weak verbs

The weak verbs are characterised by marking the past tense forms with a dental affix rather than with a change of the root vowel. Therefore their **root vowel** can be derived

---

[87] See chapter 8.1. on how the frequency is derived.

by regular expression from the actual lexical item. It is the first vowel/diphthong from the left in the "stem" of the infinitive.

The forms are assembled by concatenation as suggested in the previous chapter except for the dental "ending", which is derived from the paradigmatic dictionary and inserted between the boundary and the ending of past forms.

The parts of the paradigm defined in the paradigmatic dictionary are generated first. For weak verbs, these are: inf., pres. ind. sg 2. p. and pret. ind. sg. 1. p., all forms of class 3 weak verbs (*habban* "to have", *secgan* "to say", *licgan* "to lie" and *hycgan* "to think") and all variant forms of the paradigms.

All parts derived from and including the **infinitive** generate a variant with a syncopated vowel of the ending in case the stem of the actual item ends in a vowel. The ending of the infinitive as derived from the paradigmatic dictionary is also checked for an initial *–i–* (paradigms *sealfian, nerian* or *hyngrian*); if there is a match, all parts derived from the infinitive add *–i–* between its stem and its ending.[88] The parts derived from the infinitive are:

The inflected infinitive with the endings: *–anne*, *–enne* and the syncopated *–nne*.

Present indicative singular 1. person with the endings: *–e*, *–u*, *–o, –æ* and the syncopated *–0*.

---

[88] This might seem like a perfect candidate to be added into the boundary variable, since it is an element between the stem and the ending (or an initial part of the ending) and it is common for all verbs following the paradigm. *BT* however provides several variants of the *–ian* verbs as separate entries, but without the *– i–*, yet still following its *–ian* paradigms. A corpus sample confirms that this agrees with the actual usage. This solution allows us to generate forms with and without the *–i–* in the same paradigm as suggested by *BT.*

Present indicative plural with the endings: *–aþ*, *–eþ*, *–as* and *–es* and the syncopated ending *–þ*.

Present subjunctive singular with the ending: *–e* and the syncopated *–0* ending.

Present subjunctive plural with the ending: *–en* and the syncopated ending *–n*.

Imperative plural with the ending: *–aþ* and the syncopated ending *–þ*.

Present participle with the ending: *–ende* and the syncopated ending *–nde*.

All parts derived from and including the **present indicative singular 2.person** and preterit indicative singular 1. person simplify the geminate consonants in the post-vowel.

In addition to the ending from the paradigmatic dictionary (mostly *–est*), variant endings: *–ist*, *–s* and *–st* are also generated for pres. ind. sg. 2. person.[89]

Other parts derived from the pres. ind. sg. 2. person are:

Present indicative singular 3.person with the endings *–eþ*, *–iþ* and *–þ*.[90] In addition, the forms of both this and the previous part of the paradigm also undergo the assimilations and reductions suggested for the corresponding parts of strong verbs (see above 7.2.1).

Imperative singular with the endings *–e* and *–0*.[91]

Although weak verbs are characterized by the dental preterit as opposed to the vowel gradation of strong verbs, there are some weak verbs like *bycgan* ("to buy") that do change

---

[89] Note that verbs following *sealfian* are also generated with variant endings *–ast* and *–as* as given in the paradigmatic dictionary.

[90] Note that verbs following *sealfian* are also generated with variant ending *–aþ* as given in the paradigmatic dictionary.

[91] Note that verbs following *sealfian* are also generated with variant ending *–a* as given in the paradigmatic dictionary.

their root vowel in addition to marking past forms with a dental (e.g. past plural *bohton*).[92] To generate such forms, the program checks whether the vowel of the infinitive matches the vowel of the preterit in the paradigmatic dictionary. If not, the vowel of the preterit from the paradigmatic dictionary is used to generate past forms rather than the vowel of the actual lexical item.

Parts of the paradigm derived from the **preterit indicative singular 1. person** are:

Preterit indicative singular 2. person with the endings *–est* and *–es*.

Preterit indicative singular 3. person with the ending *–e*.

Preterit indicative plural with the ending *–on*.

Past subjunctive singular with the ending *–e*.

Past subjunctive plural with the ending *–en*.

Past participle with the *–0* ending.

7.2.3. Preterit-present and Anomalous Verbs

All forms of preterit-present and anomalous verbs are manually defined in the paradigmatic dictionary from which they are thus directly derived – except for endings of five parts of the paradigm of the preterit-presents, namely:

The **inflected infinitive** is derived from the infinitive and receives the ending *–anne*.

Preterit indicative singular 1. person is the base for:

Preterit indicative singular 2. person with the endings *–est* and *–es*.

---

[92] See Wright $534 for an explanation of the phenomenon.

Preterit indicative singular 3. person with the ending *–e.*

Past subjunctive singular with the ending *–e.*

Past subjunctive plural with the ending *–en.*

7.2.4. Participles (base forms)[93]

Old English inflects its participles like adjectives. While *BT* contains a number of the more frequent participial forms as headwords,[94] any verb can potentially form both present and past participles. Whenever a participial form is generated, it is added into an array of adjectives with a word-class status of the participle and the tense of the participle as its class1 parameter. Since participles are treated as a word-class, it makes sense to also use the participial form as a lemma of the new item. However, to keep it linked it to its verbal base, the *BT* ID stays the same.

Generation of the inflected participles is treated together with the generation of adjectival forms below.

7.3.   Adjectives

Although the adjectival inflection produces a great number of forms, it is relatively uniform and uncomplicated. The description will only focus on the basic principles and some less obvious irregularities. The individual forms distinguished mostly by their endings alone can be gleaned from the description of adjectival morphology in chapter 2.2.

---

[93] See 7.3.5. for inflected forms of the participles.
[94] Since many of them are both functionally (as modifiers) as well as formally (by their syntactic role and in their morphological inflections) equivalent to the adjectives.

The process generating adjectival forms loops through all the adjectives and participles. The forms are generated paradigm by paradigm starting with positives of strong forms, followed by weak forms, comparatives and strong & weak superlatives. The forms are concatenated from the prefix and stem of the actual lexical item and the ending that is supplied directly in the source code.

7.3.1. Strong Positive

In the case of adjectives following the paradigm of *glæd*, the root vowel *–æ–* or *–ea–* in the genitive, dative and instrumental masc. & neut. as well as nom. and acc. fem. in singular and nom./acc. and dat. of all three genders in plural has been replaced by *–a–* (i.e. in the strong forms).

In the case of adjectives following the paradigms *héah* or *þweorh*, most of the oblique forms[95] lose the final *–h* of their stem and those following *þweorh* compensate by lengthening the root diphthong (e.g. gen. masc. *þwéores*).

In the case of adjectives following the paradigm *hálig*, forms of gen., dat., instr. masc. & neut. and acc. fem. sg. as well as nom./acc. & dat. pl. of all genders (with the exception of nom./acc. neut.) syncopate the medial vowel (e.g. *–i–* in *hálig*, undergoes syncopation in the gen. masc. sg. *hálges*). This is achieved by a regular expression replacing the last vowel of the "stem".

In the case of adjectives following the paradigm *wilde* or *gearu*, the final vowel of the lemma is replaced before endings are attached (which accounts for differences such as e.g. nom. masc. sg. *wilde*, acc. *wildne*).

---

[95] All except acc. neut., but including nom. fem.

7.3.2. Weak Positive

The weak adjectival inflection is subsumed under one paradigm (*blinda*), but there is some variation that requires application of similar regular expression replacement patterns to create alternative[96] "stems" as in the case of the strong declension of paradigms *héah*, *þweorh, hálig, gearu* and *wilde* (hence e.g. the nom. masc. sg. *gearu* and gen. *gearwes*).

7.3.3. Comparative

The same replacement patterns are used to create variant stems for the comparative forms. In addition, all the comparative forms are created with mutated and un-mutated variants.[97]

The adjectives *gód*, *yfel, micel and lytel* form comparative stems by suppletive forms, which are supplied manually in the source code together with their variants: *beter* (or the variants *betr, bettr, sélr, selr*), *wiers, már* and *lǽss* respectively.

7.3.4. Superlative

Generation of stems for superlatives of both the strong and weak declension is similar. The suppletive stems for the abovementioned adjectives are: *betst* (or *betest, best, sélest*), *wierrest, mǽst* and *lǽst*.

---

[96] Stems are created as variants due to the strong analogy at work in adjectival paradigms: for example while *hálgum* is the regular form of dat. pl., the non-syncopated form *háligum* is also frequent, according to the DOEC.

[97] See chapter 7.1. on i-mutation.

7.3.5. Participles (inflected forms)

Past participles with short stem are processed together with adjectives following *manig*; those with long stems together with adjectives following *hálig*. Present participles are processed together with adjectives following *wilde*. All the participles are also declined according to the uniform declension of weak adjectivesl they also form comparative and superlative forms in the same manner. Notice that since several participial base forms (present, past and their variants) are created for every verb, together with the full adjectival declension, inflected participles generate a great amount of forms.[98]

7.4.   Adverbs

Generation of adverbial comparative and superlative forms is very straightforward: the final vowel is stripped of the stem and three variant endings for the comparative (*–or*, *–ur* and *–ar*), as well as four for the superlative (*–ost*, *–ust*, *–ast* and *–st*) are added. The adverbs *wel*, *yfele*, *micle* and *lytel* require suppletive stems similar to the adjectival forms: *bet* (or *sél*), *wiers, má* and *lǽs* in the comparative, and *betst* (or *best*, *sélest*), *wierrest* (or *wyrst*), *mǽst* and *lǽst* for the superlative, respectively.

A few irregular adverbs are also added manually.

7.5.   Nouns

Generation of nominal forms generation is similar to that of the adjectival forms. All nominal lexical items are looped through, and all paradigms assigned to each item are used to generate singular and plural forms in succession. The forms are assembled by

---

[98] See the results of the generator in chapter 9 below.

concatenation of the "prefix", the "stem" and the ending, which is supplied directly in the source code.

Since grammatical gender is an inherent property of every OE noun, the generator should include this type of information along with the generated form. Two complications preclude a straightforward use of the gender of the particular lexical item as derived from *BT*. First, nouns can be of more than one gender. Second, the example paradigm may be of a different gender than that of the lexical item declined following its inflection. The second case should be theoretically impossible, since gender plays an essential role in the organization of the inflectional paradigms. However, as explained in chapter 6.2.2.1 above, nouns are assigned paradigms here on several principles (such as stem comparison, stem analysis), gender being only one of them. Moreover, we have simplified the system of traditional paradigms by subsuming similar paradigms into larger groups. In three cases, we are thus including a paradigm of one gender into a group of similar paradigms of a different gender: *cynn* ("kin", neuter) is subsumed under *stán* (masc.); *hand* ("hand", fem.), is subsumed under *feld* (masc.); *duru* ("door", fem.), is subsumed under *sunu* ("son", masc.). For these reasons, when forms are generated, the gender that is generated for each form is the gender of the paradigm example. It is a simple solution to both problems and although not a perfect one, we believe it to be both linguistically sound[99] and user-friendly, as displaying more genders for one particular form could be misleading.

---

[99] After all, it is the inflected forms beside the grammatical concord that manifest the gender affiliation of a particular form – and there is ample evidence in the DOEC that nouns of a declension similar to nouns of a different gender were often used in both genders – e.g. "þonne cynn" (Healey, Holland, McDougall, & Mielke, 2000, p. T25220056800), or "þonne hand" (Healey, Holland, McDougall, & Mielke, 2000, p. T00050012300).

Since the paradigms defined above allow for some variation, it is mostly the replacement patterns necessitated by such variation that are discussed in this chapter. In other words, the objective here is the formation of a stem (or an array of stem variants) which can take an ending (or an array of variant endings) specific for each part of the paradigm without any need of further changes. The division of stem and ending is here purely practical: while in the paradigm of *stán*, the final *–e* is dropped in some cases only to accommodate both nouns like *ende* (nom. sg. *ende-0*, dat. pl. *end-um*) as well as *stán* (nom. sg. *stán-0*, dat. pl. *stán-um*), in the paradigm of *bearu*, the final *–u* or *–w* can be dropped in all cases, since the paradigm is more uniform so that the endings can be more specific and include the vowel or approximant (e.g. nom. sg. *bear-u* and dat. pl. *bear-wum*).[100]

We proceed by paradigms assigned in the previous steps.

7.5.1. Strong Nouns

Stán (masc.):

- Stem-final geminate consonants are simplified in the nom. & acc. of nouns traditionally following *cyn* (but remain in a variant stem): e.g. nom. sg. *cyn* or *cynn*, but gen. sg. *cynnes*. While *cyn* is not used as a separate paradigm for generation, it was assigned previously to allow the nouns following it to be assigned neuter gender in an otherwise masculine paradigm of *stán.*

- Stem-final *–e* is dropped in all cases (except nom. & acc. sg.) of nouns traditionally following *ende* ("end") and *wine*: e.g. dat. pl. *endum*.

---

[100] Traditionally, the structure of nom. sg. of *ende* and dat. pl. of *bearu* would be analysed as *end-e* and *bearw-um*. Since displaying a linguistically more adequate structure of the inflected forms is not our goal, we avoid a more complex paradigmatic system this would entail.

- Stem-final –*h* is dropped in all cases (except nom. & acc. sg.) of nouns traditionally following *mearh* ("horse"): e.g. gen. sg. *méares.*

- Vowel in a stem-final vowel + consonant combination is syncopated in all cases (except nom. & acc. sg.) of polysyllabic nouns traditionally following *engel* ("angel"), but may remain in a variant stem: e.g. gen. sg. *engles*, or *engeles.*

- Stem-final –*o*– + consonant combination changed into –*e*– + consonant in plural of nouns traditionally following *heofon* ("heaven"), but remains in a variant stem: e.g. gen. pl. *heofenas*, or *heofonas.*

Word (neuter):

- Vowel in a stem-final vowel + consonant combination is syncopated in all cases (except nom. & acc.) of polysyllabic nouns traditionally following *tungol* ("star"), but remains in a variant stem: e.g. gen. sg. *tungles*, or *tungeles.*

Hof (neuter):

- Vowel in a stem-final vowel + consonant combination is syncopated in all cases (except nom. & acc.) of polysyllabic nouns traditionally following *héafod* ("head"), but remains in a variant stem: e.g. gen. sg. *héafdes*, or *héafodes.*

- Stem-final –*e* is dropped in all cases (except nom. & acc. sg.) of nouns traditionally following *wíte* ("torment") and *spere* ("spear"): e.g. dat. pl. *sperum.*

- Stem-final geminate consonants are simplified in nom. & acc. of nouns traditionally following *wésten* ("desert"), but remain in a variant stem: e.g. nom. sg. *wésten* or *wéstenn*, but gen. sg. *wéstennes.*

Ár (feminine):

- Stem-final *–u* or *–o* is dropped in all cases (except nom. & acc. sg.) of nouns traditionally following *giefu* ("gift") and *strengþu* ("strength"): e.g. dat. sg. *giefe*.

- Vowel in a stem-final vowel + consonant combination is syncopated in all cases (except nom. & acc. sg.) of polysyllabic nouns traditionally following *sáwol* ("soul"), but remains in a variant stem: e.g. gen. sg. *sáwle*, or *sáwole*.

- Stem-final geminate consonants are simplified in nom. & acc. of nouns traditionally following *hen* ("hen") and *byrþen* ("burden"), but remain in a variant stem: e.g. nom. sg. *hen* or *henn*, but gen. sg. *henne*.

Feld (masculine):

- While *hand* is not used as a separate paradigm for generation, it was assigned previously to allow the nouns following it to be recognized and assigned feminine gender in an otherwise masculine paradigm of *feld.*

Sunu (masculine):

- While *duru* is not used as a separate paradigm for generation, it was assigned previously to allow the nouns following it to be recognized and assigned neuter gender in an otherwise masculine paradigm of *sunu.*

- Stem-final *–u* is dropped in all cases: e.g. gen. pl. *suna*.

Dæg (masculine) and *fæt* (neuter):

- Plural stems replace *–æ–* or *–ǽ–* with *–a–* or *–á–* respectively: e.g. nom. pl. *daga*.[101]

Strengu (feminine):

---

[101] However, due to the operation of analogy, there are 20 occurrences of levelled forms like *dægum* in the DOEC. Therefore, we generate plural stems with both alternate and the basic vowel.

- Stem-final *–u* is dropped in all cases: e.g. gen. pl. *strenga*.

Bearu (masculine) and *bealu* (neuter):

- Stem-final *–u* or *–w* is dropped in all cases: e.g. gen. pl. *strenga*.

7.5.2. Weak Nouns

Guma (masculine):

- Stem-final *–a* is dropped in all cases: e.g. gen. pl. *gumena*.

Tunge (feminine):

- Stem-final *–e* is dropped in all cases: e.g. acc. sg. *tungan*, but also *tungean*.[102]

Éage (neuter)

- Stem-final *–e* is dropped in all cases: e.g. acc. sg. *éagan*, but also *éagean*.[103]

Paradigms not mentioned in this chapter generate their forms regularly by simply adding the endings listed in chapter 2.1. to an unchanged "stem" derived from the lemma.

7.6. Numerals

Numerals from one to three are defined manually, due to the number of variants and irregularity of their forms. The rest of the cardinal numerals decline like nouns[104] as well as weak adjectives, while the ordinal numerals above three decline like weak adjectives only. Mostly for reasons of transparency, however, the forms of numerals are generated separately. Only plural forms are generated.

---

[102] The variant probably formed by analogy.
[103] The variant probably formed by analogy.
[104] See chapter 7.5. The fact that the inflected forms are infrequent compared to the non-inflected forms in actual usage plays no role here.

Cardinal numerals are recognized by being marked both as numerals as well as nouns in the *BT*. They follow three nominal paradigms used specifically for numerals. Each paradigm is used for one gender and the forms are generated for each gender of each numeral. There is therefore no need for previous paradigm assignment, but, for the benefit of the users, the three paradigm examples are assigned during the generation process.

Masculine forms follow the paradigm of *wine*, feminine *cwéne* and neuter *spere*. Stem-final *–e* or *–a* is dropped in all cases of all three paradigms, although it remains as a variant in nom. and acc.

Cardinals and ordinals declined like weak adjectives follow the paradigm of *blinda* exactly as described in chapter 7.3.2.

7.7.   Pronouns

Personal (including possessive), demonstrative and interrogative pronominal forms are inflected manually. Indefinite pronouns are declined together with weak positive forms of adjectives (like *blinda*).[105]

7.8.   Other Word-classes

All other word-classes are considered uninflected. The lexical items of these word-classes are loaded and processed similarly to the inflected items (see chapter 6.1.2.2.), but do not generate any new forms and are outputted along with the other forms in the following chapter.

---

[105] See chapter 7.3.2.

7.9.   Generator Output

All forms are loaded into an array during the generation process (or during the loading of the input files in the case of the uninflected forms and the manually inflected forms). The whole array is then processed and outputted into an external UTF-8 encoded tab-delimited file of 19 columns with one form per line. Two of these columns are created during this process: `counter` and `form_i`.

`Counter` is an identification number unique for each form starting with 0. It can be used to trace an individual form from the user feedback back to its source in the database. In particular, this can be done during the de-bugging phase. The order of the items is not significant.

`Form_i` is a generated form stripped of diacritics and having all the occurrences of *i, y* or *ie* standardized into *i* all the occurrences *k* standardized into *c.* Also, if there are double consonants in any form, a variant form on a separate line is generated with a single consonant and a lower `probability`. This is done to facilitate the matching process during the analysis of an actual text as well as to decrease the computational load on the database.[106]

The other columns were already described at the beginning of this chapter.

---

[106] See the chapter 8.2.1. for more details.

ID	form_i	BT	lemma	stem	form	formParts	var	prob.	function	wright	pEx	pID	wclass	c1	c2	c3	com.
240508	ætswummon	1014	æt-swymman	swymman	ætswummon	0-æt-sw-u-mm-0-on	0		painpl	Null	bindan	10	verb	s	3	a	
240509	ætswumon	1014	æt-swymman	swymman	ætswummon	0-æt-sw-u-mm-0-on	0	1	painpl	null	bindan	10	verb	s	3	a	
240510	ætswumme	1014	æt-swymman	swymman	ætswumme	æt-sw-u-mm-0-e	0		PaInSg2	null	bindan	10	verb	s	3	a	
240511	ætswume	1014	æt-swymman	swymman	ætswumme	æt-sw-u-mm-0-e	0	1	PaInSg2	null	bindan	10	verb	s	3	a	
240512	ætswumme	1014	æt-swymman	swymman	ætswumme	æt-sw-u-mm-0-e	0		PaSuSg	null	bindan	10	verb	s	3	a	
240513	ætswume	1014	æt-swymman	swymman	ætswumme	æt-sw-u-mm-0-e	0	1	PaSuSg	null	bindan	10	verb	s	3	a	
240514	ætswummen	1014	æt-swymman	swymman	ætswummen	æt-sw-u-mm-0-en	0		PaSuPl	null	bindan	10	verb	s	3	a	
240515	ætswumen	1014	æt-swymman	swymman	ætswummen	æt-sw-u-mm-0-en	0	1	PaSuPl	null	bindan	10	verb	s	3	a	
240516	ætswamm	1014	æt-swymman	swymman	ætswamm	0-æt-sw-a-mm-0-0	0		painsg1	null	bindan	10	verb	s	3	a	
240517	ætswam	1014	æt-swymman	swymman	ætswamm	0-æt-sw-a-mm-0-0	0	1	painsg1	null	bindan	10	verb	s	3	a	
240518	ætswamm	1014	æt-swymman	swymman	ætswamm	æt-sw-a-mm-0-0	0		PaInSg3	null	bindan	10	verb	s	3	a	
240519	ætswam	1014	æt-swymman	swymman	ætswamm	æt-sw-a-mm-0-0	0	1	PaInSg3	null	bindan	10	verb	s	3	a	
240520	ætswummen	1014	æt-swymman	swymman	ætswummen	0-æt-sw-u-mm-0-en	0		papt	null	bindan	10	verb	s	3	a	
240521	ætswumen	1014	æt-swymman	swymman	ætswummen	0-æt-sw-u-mm-0-en	0	1	papt	null	bindan	10	verb	s	3	a	
240522	ætswimman	1014	æt-swymman	swymman	ætswimman	0-æt-sw-i-mm-0-an	0		if	null	bindan	10	verb	s	3	a	
240523	ætswiman	1014	æt-swymman	swymman	ætswimman	0-æt-sw-i-mm-0-an	0	1	if	null	bindan	10	verb	s	3	a	

Table 36 Sample of generated forms for the verb *æt-swymman*.

A sample of the conjugation of the verb *æt-swymman* ("to swim out") is included here. Notice the last two lines in the `form_i` column: *y > i* and *mm > m* in the last line.

In total, the output of the generator is 10,796,162 forms.[107] To be precise, we only generate some 1,928,374 unique forms, but there are many identical forms that carry a different grammatical function (grammatical homographs), or belong to a different word-class or lemma (lexical homonyms). When we speak about "forms", we refer to this "form-function pair"; otherwise, we speak of "unique forms".

The following table breaks down the number of both forms and unique forms by word-classes, genders of nouns, types of verbs and participles. The ratio shows how many forms are generated per base form (e.g. on average there are 13.1 forms generated for each noun). The unique ratio shows how many unique forms are generated per base form (on average only 7.1 unique forms are generated for each noun).

---

[107] In fact, the output is higher – 13,828,441 forms are generated-, but there are many duplicates which are filtered out during the process described in the following chapter.

word-class	words	forms	ratio	unique forms	un. ratio
adjective	6 235	1 882 110	301,9	294 349	47,2
adverb	1 863	16 640	8,9	15 652	8,4
conjunction	54	61	1,1	55	1,0
interjection	25	26	1,0	26	1,0
noun	19 835	259 496	13,1	140 837	7,1
masculine noun	8 668	93 874	10,8	59 251	6,8
feminine noun	7 306	122 508	16,8	57 133	7,8
neuter noun	4 067	43 114	10,6	27 477	6,8
numeral	110	3 344	30,4	773	7,0
preposition	148	160	1,1	153	1,0
pronoun	43	4 996	116,2	1 010	23,5
verb	6 959	561 736	80,7	395 506	56,8
weak verb	5 094	461 239	90,5	325 768	64,0
strong verb	1 786	103 313	57,8	74 128	41,5
preterit-present vb.	32	1 245	38,9	813	25,4
anomalous verb	59	2 481	42,1	1 569	26,6
participle	35 642	8 715 027	244,5	1 247 483	35,0
past participle	12 959	2 930 381	226,1	433 783	33,5
present participle	22 683	5 784 646	255,0	814 928	35,9
TOTAL (avg. ratio)	70 914	11 443 596	161,4	2 095 844	29,6

Table 37 An overview of generated forms.

A few comments are necessary to clarify the numbers:

- The numbers do not total up as expected because lexical items can belong to multiple word-classes, or have more than a single gender or a single verb-class membership.

- While a particular class of class nouns and verbs or a particular gender of a noun can also be "uncertain" in *BT*, the category of the "uncertain" does not appear here – in other words, even members of this category are assigned a paradigm and are thus subsumed under one of the classes in Table 36.

- Uninflected word-classes have more forms than bases because forms with geminate consonants generate variants with simple consonants.

- Even though the number of functions is the same for every item in a given word-class, the number of forms in sub-classes can differ. This can be due to differences

in the number of base forms (participles), or the fact that the paradigms relevant to a particular class are more varied than those relevant to another class (e.g. feminine vs. masculine or neutral paradigms of nouns).

- The number of the base forms of present participles is almost twice the number of the base forms of past participles. The reason this lies again in the variation, though here it is the difference in the variation of different parts of the same verbal paradigms. For example, the *–ian* verbs (paradigm *nerian*, *hyngrian* and *sealfian*) have five different variants of the infinitive (e.g. *nerian, nerigan, nerigean, nergan* and *neran*), but only two variants of the preterit (e.g. *nerode* and *nerede*). Present participles are generated from the infinitive forms, while past participles from the preterit forms – which means there will be 2.5 as many present participle base forms and generated forms as past participle base and generated forms.

- The difference in the number of base forms between strong and weak verbs is easy to understand, but as the difference in the ratio shows, on average, weak verbs also generate ca. 1.5 times more forms than strong verbs. The reason for this difference is harder to explain and perhaps points to a weakness in the input data, namely in the dictionary of verbal paradigms. Almost half of the weak verbs belong to the *–ian* verbs which feature a relatively high number of variants in the dictionary of verbal paradigms. On the other hand, the strong paradigms that are followed by a large number of verbs (e.g. *bídan*) are recorded as almost uniform. This, however, may be due to a bias of the descriptive grammars like Wright's and only a careful probe in the *DOEC* could show whether it holds true in the actual data.

- The number of forms generated in each word-class is determined by the number of "functional slots" for a given word-class and the degree of variation. While intuitively one might think of verbs as being the most inflected word-class in OE,

adjectives generate more than three time as many forms as OE verbs. That is mostly due to the fact that we include comparative and superlative forms into inflection, rather than into the domain of word-formation.

- On the other hand, verbs have a higher ratio of unique forms than adjectives, because adjectives have a high amount of grammatical homography (there are 24 functional slots for the positives of weak adjectives, but only ca. 5 unique forms to carry these functions). The difference between the ratio and unique ratio (together with its height) is a good indicator of the level of disintegration of the inflectional system of a given word-class – if the difference between the number of functional slots and the number of forms is too high, the system will probably lose the functional distinctions altogether. This corresponds to the traditional descriptions of the development of English quite well. The adjectival inflection collapses very soon, the nominal inflection has more or less disappeared in PDE as well, but the verbal and, to a smaller degree, also the pronominal inflections (more precisely the inflection of personal pronouns)[108] are still functional in PDE.

- The surprisingly high number of forms overall is mainly due to participles. Formation of participles from verbs and the formation of participial comparatives and superlatives are processes bordering on word-formation. In other words, together with adjectival comparison, inflected participles show a high degree of agglutination. To illustrate this point, let us consider the form *ner-i-end-ost-um*[109] (based on *nerian* – "to save"): the *–end–* signals present participle, *–ost–* signals superlative, *–um* signals dative plural (while *-i-* signals a derivation from nominal

---

[108] The relatively high ratio difference for pronouns is given by the high grammatical homography of the interrogative and indefinite pronouns that decline like weak adjectives.

[109] While this example is not attested, formally similar inflected participles are, e.g. *ge-lǽr-d-est-an* from *lǽran* ("to teach").

ner – "refuge", but word-formation proper is not our topic here). It is precisely such concatenation of monofunctional affixes, so typical of agglutination, that is manifested by the multiplication of the generated forms: each verb generates at least two participles, while each participle inflects both weak and strong for two numbers, four cases and three genders and creates comparatives and superlatives besides the positive forms, or in quantitative terms: 2 x 2 x 2 x 4 x 3 x 3 = 288. If we take into account the fact that comparatives only inflect according to the weak declension, we arrive at 240 participial forms generated for each verb. If there is a variation in the verbal forms of the participles (e.g. a geminate consonant), the number is doubled. Any other variation multiplies the number again – verbs spawning over 2,000 forms are not uncommon. It is easy to see, then, why the proposed system may not work well for highly agglutinating languages.[110]

- The number of generated forms per word-class has no significance in terms of the actual frequency of forms in OE texts. Although there are ca. 7 times more adjectival forms generated than there are nominal forms, the actual ratio in the corpus is ca. 1:3 in favour of the nominal forms.[111]

---

[110] To put it differently: if a language were also prone to variation requiring complex queries against the database in addition to the large number of forms it would naturally generate, such system would necessitate faster hardware.

[111] According to the YCOE markup, there are 73,742 adjectival, 253,980 nominal and 176,557 verbal items.

8. Analysing Text

The generated data have many potential uses (see chapter 1.1.), but we will focus here on using and evaluating the data for morphological analysis of OE texts by end-users and for automatic creation of glossaries based on such analysis. The generated glossaries are easy to compare to hand-made glossaries and will allow us to evaluate the data and the whole process in the following chapters.

8.1.   The Database

Since text analysis is most likely to be carried out by end-users (as opposed to the generation of forms), it should be available as an online service, rather than as a command-line run Perl script. Therefore, the analysis itself requires technology more suitable for web services.[112]

After the data are exported into a text file by the generator, they are imported into a MySQL database. Two tables are used to store that data. Two tables are used mostly to speed up the text analysis, but in principle, the same thing could be accomplished with one table only.

Table 1 uses the InnoDB engine and is structured identically to the generator output file (see the Table 35 above), except that a second `ID2` is added, linking it to Table 2. The duplicities mentioned in the previous chapter are removed on import due to a unique key set over *BT*, `form_i`, `function` and `wordclass` columns. *BT*, both `ID` columns and `probability` column are indexed to improve analysis speed.

---

[112] See the description of the technology in chapter 5.

Table 2 uses the MyISAM engine to allow full-text indexing even on pre-5.6 installations of MySQL. The structure consists of 4 columns: 1. `ID` (connecting it to Table1), 2. `form_i` (containing the generated forms from the `form_i` column of the generator ouput, see above), 3. `freq` and 4. `texts`. The two latter columns contain information on the frequency of a given form in the *DOEC* corpus (see below). All columns are indexed and the `form_i` column has a full-text index, since it is the column used for looking up forms during the analysis. `form_i` also allows only unique values, which is the main feature speeding up form look-ups. Instead of querying Table 1 with ca. 12 million records, we query Table 2 with ca. 2 million records and then join the results with Table 1. This saves considerable processing time especially in the case of very costly regexp queries.

Since a single text analysis can generate hundreds of queries and since we want to make the application publicly available, efficiency is an important factor in the case of the online application (unlike in the case of the offline generator script). Even with the data divided into two tables, speed is a limiting factor in the development of the online interface of the analyser.

As noted above, Table 2 contains additional information to the data generated by the generator script. The additional information is derived from the *DOEC* and its purpose is to inform both the analyser (and later perhaps the disambiguator) as well as the end-users about the occurrence and frequency of the generated forms in the *DOEC*.

To accomplish this, the *DOEC* was processed by the WordSmith Tools in the following fashion:

- headers and tagged foreign words were omitted from the process[113]

- text was standardized to conform with the forms generated into the `form_i` column (i.e. *ie* and *y* > *i*, eth > thorn, all characters > lowercase)

- wordlist was generated with the frequency of each word (type) and a number of texts in which the word appears throughout the corpus

- resulting data were imported into a database and into Table 2 via string comparison of the forms in the `form_i` column and the words in the wordlist

Of the 172,800 *DOEC* types, 71,302 (ca. 41%) were identified with the generated forms by a direct string comparison.[114] That, however does not mean that the rest of the types would not be analysable, since no variation filters were applied. At some later stage, we may try inputting the whole wordlist into the analyser to apply all the variation filters and obtain a list of unanalysable items as well. Such a list, especially in connection with the actual frequency information, could be used with profit for future improvements and debugging.

8.2.   The Analyser Script

The aim of the PHP script (that is currently the only front-end of the analyser) is to:

1. take OE text from the end-user on input,

2. process it to derive a list of all individual types,[115]

---

[113] Many words of foreign, especially Latin and Greek, origin remain in the data, since not all foreign words are tagged as such in the DOEC. This may also be due to the contentious status of foreign words – do single Latin words in an otherwise fully OE context qualify as foreign, or are they borrowings? It is not our purpose to discuss the policies of the DOEC tagging, but it should be noted that the number of unmatched forms in the DOEC is adversely influenced by a large (but difficult to estimate with any precision) number of Latinate or Greek words. Proper nouns have a similar effect.

[114] Which in turn means that only ca. 0.04 % of the generated forms appear (without any variation) in the DOEC.

[115] By type we here mean a unique word-form, not a lemma.

3. match each type with a form in the database,

4. fetch all data about the form from the database,

5. and finally present the data about each form back to the end-user.

The text is submitted via a standard HTML form and its processing is relatively straightforward. All non-letter characters (except spaces) are discarded, the rest is standardized to conform with the forms generated into the `form_i` column (i.e. *ie* and *y* > *i*, *eth* > *thorn*, *c* > *k*, all characters > lowercase), each string divided from others by spaces is loaded into an array of tokens and the array is then sorted alphabetically and reduced to unique elements only rendering an array of types.

8.2.1. Matching Forms and Variation Filters

Each type is then matched against the forms in the `form_i` column of Table 2.[116] If there is no match, the query is repeated, but this time, variation filters are applied. The filters are implemented as cumulative modifications of the query. Therefore, if more than one of the filters is applicable to the query in question, all the combinations of the applicable filters will be used. This principle may lead to a large number of queries generated on account of just one original type, and so the efficiency is an important factor again.

Every filter first tests whether it is applicable to the query in question. Where possible, the modification is implemented as a change to the argument of the simple `LIKE` operator. If necessary, the two wildcard characters ("_" and "%") available in the syntax of the `LIKE` comparison are used. Only if there is no more efficient solution, the "costly" regular expression pattern of the `RLIKE` operator is used. To speed up the regular expression matching, the `RLIKE` operator is always preceded by a simple `LIKE` operator in the `WHERE`

---

[116] With the variation filter for superlative affixes already applied. See below.

condition which reduces the size of the temporary table to only those forms that may potentially match the RLIKE comparison by substituting the regular expression wildcards with the two aforementioned simple wildcards. For example, if the type/form to match against the database is *þiostru* ("darkness") and it is not found by a simple LIKE comparison, a filter may be applied which tries to match it taking the variation of *eo/io/ie* diphthongs into account. The regular expression pattern for this query would be "`^þ[iíeé]+[eo]stru$`",[117] but that would be preceded by a simple pattern "`þ__sttru`" as an argument of the LIKE operator effectively limiting the number of rows of Table 2, which need to be matched against the regular expression.

In short, the variation filters attempt to make the match possible by taking into account the most common scribal, diachronic and dialectal variations in OE texts. The variations are loosely based on (Baker, Common Spelling Variants, 2012) and Wright, but some were combined into one or simplified. The rules given below may in fact be implemented by several queries.

We may divide the variations into three basic groups:

Variation in unaccented syllables

- the final unstressed −*e* was dropped especially in the late and northern OE[118]

- −*n* can be lost in unaccented syllables (endings) −*an*

- *e* can be syncopated in final −*gen*

- the unaccented dative ending −*um* can be spelled as −*un* or −*on*

---

[117] Specifying the beginning and end of the string with the "^" and "$" special characters helps to speed up the RLIKE comparison.

[118] If a variation is specified by period or dialect, such specification is naturally ignored by the analyser, since it has no way to check for such conditions. It is included here to explain the motivation for the variation and the analyser can display that type of information for the benefit of the end-users.

- *e* in final or pre-ending –*er* combinations may be syncopated

Consonantal variations

- consonants can be doubled before *l* or *r*

- geminates may be shortened or spelled as single consonants[119]

- *fn* can stand for *mn* or *mm* in later texts

- final –*g* can stand for –*h*

- initial *geo*– can stand for *iu*–

- *g, ig, i* can be interchangeable after front vowels

- *r* can undergo a metathesis in a syllable ending in *n* or *s*[120]

- *sc* can be spelled as *x*

- *éagn* may be shortened to *éan* or *én*

Vowel variations[121]

---

[119] It is worth noting here that most of the variation filters work in two (opposite) "directions" although we do not add "and vice versa" to all the rules listed here. For example, if the diphthongs *eo* and *io* are said to be in variation, not only must every occurrence of *eo* in the text type be replaced by *io* in the query, but also every *io* has to be replaced by *eo*. The variation of geminate consonants, however, implements one direction of the variation already in the generator by outputting all double consonants with a single consonant variant (see 7.9. above). Therefore, if the user looks for the preposition *mid* ("with"), the form is correctly identified, even though the standard generated form is *midd*. This would be computationally costly to implement as a modification to the query (i.e. variation filter), since every consonant in the text type would be a potential geminate and the ensuing number of combinations could be very high (to be precise, 2^n where n is a number of consonants, so that in the case of a simple word like *dryhten*, this one variation would create 32 queries). To illustrate the procedure in the other direction: if the user enters e.g. *mann*, while the standard form in the database is *man,* we can easily modify a query by replacing each geminate with a single vowel. Since geminates are much rarer than single consonants, this solution will be efficient, whereas generating every form with any combination of its consonants doubled would lead to an enormous increase in the total number of forms.

[120] Syllable detection is not implemented in the online script and this variation filter is implemented as a simple replacement pattern: *Vr[n/s] > rV[n/s]* disregarding syllable boundaries. It is however to be expected that a *Vr[n/s]* combination would usually form a syllable boundary anyway.

[121] Vowel variations are mostly implemented using regular expressions. This makes them computationally costly and also very productive. On the other hand, the vowel variation is, in general, the most common type of variation in the texts. If too many forms are identified by mistake, this is a good area for a review based on the unidentified forms on the one hand and the "noise" in the actual results of the analyser on the other hand.

- the closed front vowels *i, y, e* and (rarely) *u* may be interchangeable

- the back vowels *a, o, u* may be interchangeable

- the closed front vowels *i, y, e* and (rarely) *u* may be interchangeable for the diphthongs *eo, io, ie* or they can stand for the diphthongs *eo, io*

- the vowels *e* or *a* may stand for the diphthong *ea*

- *æ* can be spelled as *ea*

- the diphthongs *eo, io, ie* may be interchangeable

- the diphthongs *eo, io* are interchangeable or can stand for the vowels *i, y, e, o, u*

- the vowel of the superlative suffix may be *e, o, a, u* or zero + –st[122]

<u>Negative prefix</u>[123]

- initial *ne–* can prefix a negated verb form

- stem-initial *h–* or *w–* often disappeared after *ne–* in certain negated verbal forms (e.g. *nabban* or *næron,* cf. Wright $267, $325)

---

[122] Although the predictable variants of affixes are usually supplied by the generator, generating five variants of the superlative affix would create a disproportionate amount of forms (see chapter 7.9. above), while the variation filter works quite effectively. Since this type of variation is very common and its application does not lead to false positives, the filter is applied already for the initial query, i.e. not only after the initial query fails.

[123] Filters taking care of negative prefixes do not represent variation in the linguistic sense, but allow identification of many negative verbal forms. Verbal forms are not regularly generated with the negative prefix because the number of the forms would be high and the process borders on word-formation. Moreover, we can do with only one direction of the "variation" here – if the user looks for a negative form, say *nelædde* ("do not lead"), we remove the negative prefix so that the query matches the correct form of the verb *lædan*, but we do not really need to add the negative prefix to positive queries.

8.2.2. Displaying the Results

All the matched items from the queries are stored in an array recording all the information on a particular form from Table 1, but also frequency data from the *DOEC* and a comment on variation filters applied (if any were applied) to match the form in question.

Currently, the output takes the form of a glossary, because it is a simple, yet versatile: It can be used for an analysis of a single word or a text of hundreds of words. It works well online, but can also be printed. There are, however, many other ways to present the data: the input text can be annotated using hypertext or hover effects, the information can be exported into an xml file to enable further processing by corpus clients etc.

The glossary is formatted as a bulleted list in five levels with each type/word from the analysed text starting its own hierarchy:

1. the word form (type) from the text;

2. the lemma (each word can be analysed as belonging to multiple lemmata; the lemma is clickable and linked to the online version of *BT* – in the future, definitions or equivalents can be included directly in the glossary);

3. the word-class (each lemma can belong to multiple word-classes, with participles as one of the word-classes);

4. the paradigm example word (each lemma can by inflected according to multiple paradigms in a given word-class; paradigm examples are clickable and linked to their full conjugation/declension);

5. the functions of the form (the function can be displayed abbreviated or in full; additional morphological information is included in the brackets of each form).

To illustrate this arrangement, here is an analysis of the word HÚS as displayed in the glossary format:[124]

HÚS

- ○ *<u>húsian</u>*
  - ▪ <u>verb</u>
    - ● infl. like **sealfian**
      - ○ *Sg. 2. p. Ind. Pres.* (w, 2, a, húss)
      - ○ *Sg. Imp.* (w, 2, a, hús)
- ○ *<u>hús</u>*
  - ▪ <u>noun</u>
    - ● infl. like **word**
      - ○ *Sg. N. Nom.* (strong, hús)
      - ○ *Sg. N. Acc.* (strong, hús)
      - ○ *Pl. N. Nom.* (strong, hús)
      - ○ *Pl. N. Acc.* (strong, hús)

The two-step approach to variations[125] is a compromise between a significant number of forms that would not be matched at all and a high level of "noise".[126] It is not a perfect compromise, however. If a variant form is homonymous with a standard form of a different word, the variant form will be mistakenly identified with this word and the variation filters that could identify it with a correct word are not used at all. Consider the word *déad* ("dead"). The *DOE* records several variants of the word, including dative sg.

---

[124] ID codes and the frequency information from the DOEC were omitted here to make the presentation clearer. See screenshots of the actual webpage in Appendix 12.3.

[125] First a query with only the superlative affix variation is applied and only if that produces no match, the rest of the filters is applied – see above.

[126] Linguistically wrong matches, technically these are classified as false positives.

dǽde. Unfortunately, *dǽde* is also a standard form of the dative sg. of *dǽd* ("deed"). Therefore, if the user tries to analyse the form *dǽde*, the analyser will return dat. sg. of *dǽd*, since that will be a clear match in the first step. Variation filters that would identify the form correctly as dat. sg. of *déad* will never get applied.

In view of this problem, the variation filters could perhaps be divided in terms of probability and applied in more steps, each following only if the previous returns no match.[127] This would, however, require a comprehensive statistical evaluation of OE variation that is beyond the scope of this dissertation.

More detailed evaluation and analysis of the results can be found in the following chapters.

---

[127] See chapter 10.1.2. below.

9. Discussion of the Results

9.1.   Testing the Analyser

The finished online tool was first tested on random samples from the *DOEC* and by probes directed to potentially problematic areas that helped to debug the tool on all levels: the lexical and paradigmatic input data, the philological algorithms of the generator as well as the programming of the generator and the analyser themselves.

In the following step, a series of sample texts was analysed and their analysis manually checked to measure the accuracy of the analyser across the variable linguistic continuum of OE as well as to accumulate material and formulate plans for future improvements.

9.1.1. Text Selection

The texts were selected to cover the diatopic (geographical), diachronic and generic variety of OE textual production. Since the results of the analyses (i.e. automatically generated glossaries) were checked manually against man-made glossaries, the length of the texts was a limiting factor, as well as the existence of man-made glossaries for these texts.[128] The selection is therefore not representative in the narrow sense of equal or comparable representation of all varieties of OE, as the predominance of West-Saxon prose of the tenth century is obvious. However, samples from all major dialects, different periods and genres are included and are therefore represented in the selection. Moreover, the predominance of West-Saxon reflects its predominance among the attested material.

---

[128] Automatic and man-made glossaries of texts 3–7 were collated by a skilled student assistant (Dominika Ortutayová) and glossaries for texts 8–10 were kindly created for this purpose by prof. Jan Čermák.

9.1.2. Testing Method

Ten texts of 2,554 tokens were selected, automatically analysed and the results were collated with their original glossaries. For each type we noted whether the lemma and the grammatical function suggested by the man-made glossary were among those automatically generated by the analyser. We also noted lemmata and grammatical functions that should not have been generated, although this practice is problematic (see the following). In addition, we marked missing or mistakenly generated items for word-class.

A token-by-token analysis of the results is exemplified in what follows using the first sentence from the sample text "House on the Rock" (see Appendix 12.4.1. below): "*Ǣlc þāra þe þās mīn word gehīerþ, and þā wyrcþ, biþ gelīc þǣm wīsan were, se his hūs ofer stān getimbrode.*" ("Each of those who hears my words and acts accordingly is like the wise man who built his house on stone.")[129] The output of the analyser for this sentence is followed by the analysis itself. For better readability and due to space limitations, links to declension/conjugation of the respective lemmata, IDs of forms, probability values and frequencies in *DOEC* were removed from the output. Otherwise, the output is printed here without any further modifications.[130]

---

[129] The translation consciously follows the OE text very closely, hence the resulting ungrammatical expression "those who hears".

[130] For a detailed description of the output, see chapter 8.2.2. above. For a screenshot of the HTML version, see appendix 12.3.

9.1.2.1.   Sample of Automatically Generated Glossary

ǼLC

o *ǽlc*
  ▪ pronoun
  • infl. like blind
    o Sg. M. Nom. Pos. (strong, ǽlc)
    o Sg. M. Acc. Pos. (strong, ǽlc)
    o Sg. M. Dat. Pos. (strong, ǽlc)
    o Sg. N. Nom. Pos. (strong, ǽlc)
    o Sg. N. Acc. Pos. (strong, ǽlc)
    o Sg. N. Dat. Pos. (strong, ǽlc)
    o Sg. F. Nom. Pos. (strong, ǽlc)
    o Sg. F. Gen. Pos. (strong, ǽlc)
    o Sg. F. Dat. Pos. (strong, ǽlc)
    o Pl. M. Nom. Pos. (strong, ǽlc)
    o Pl. M. Acc. Pos. (strong, ǽlc)
    o Pl. M. Gen. Pos. (strong, ǽlc)
    o Pl. N. Nom. Pos. (strong, ǽlc)
    o Pl. N. Acc. Pos. (strong, ǽlc)
    o Pl. N. Gen. Pos. (strong, ǽlc)
    o Pl. F. Gen. Pos. (strong, ǽlc)

AND

o *and*
  ▪ preposition
  • infl. like and
    o (and)
o *and*
  ▪ conjunction
  • infl. like and
    o (and)
o *andian*
  ▪ verb
  • infl. like sealfian
    o Sg. Imp. (w, 2, a, and)

BIÞ

o *béon*
  ▪ verb
  • infl. like béon
    o Sg. 3. p. Ind. Pres. (a, biþ)
o *búgan*
  ▪ verb
  • infl. like búgan
    o Sg. 3. p. Ind. Pres. (s, 2, f, býþ)
o *búgan*
  ▪ verb

  • infl. like búgan
    o Sg. 3. p. Ind. Pres. (s, 2, f, býþ)
o *bión*
  ▪ verb
  • infl. like béon
    o Sg. 3. p. Ind. Pres. (a, biþ)
o *búan*
  ▪ verb
  • infl. like helpan
    o Sg. 3. p. Ind. Pres. (s, 3, d, bieþ)

GEHÍERÞ

o *ge-híran*
  ▪ verb
  • infl. like déman
    o Sg. 3. p. Ind. Pres. (w, 1, a, gehírþ)
o *ge-hýran*
  ▪ verb
  • infl. like déman
    o Sg. 3. p. Ind. Pres. (w, 1, a, gehýrþ)
o *ge-hýran*
  ▪ verb
  • infl. like déman
    o Sg. 3. p. Ind. Pres. (w, 1, a, gehýrþ)

GELÍC

o *ge-lícian*
  ▪ verb
  • infl. like sealfian
    o Sg. Imp. (w, 2, a, gelíc)
o *ge-lícian*
  ▪ verb
  • infl. like sealfian
    o Sg. Imp. (w, 2, a, gelíc)
o *ge-liccian*
  ▪ verb
  • infl. like sealfian
    o Sg. Imp. (w, 2, a, gelic)
o *ge-lícan*
  ▪ verb
  • infl. like déman
    o Sg. Imp. (w, 1, a, gelíc)
o *ge-líc*
  ▪ adjective
  • infl. like til
    o Sg. M. Nom. Pos. (strong, gelíc)
    o Sg. M. Acc. Pos. (strong, gelíc)

- o Sg. M. Dat. Pos. (strong, gelíc)
- o Sg. N. Nom. Pos. (strong, gelíc)
- o Sg. N. Acc. Pos. (strong, gelíc)
- o Sg. N. Dat. Pos. (strong, gelíc)
- o Sg. F. Acc. Pos. (strong, gelíc)
- o Sg. F. Gen. Pos. (strong, gelíc)
- o Sg. F. Dat. Pos. (strong, gelíc)
- o Pl. M. Nom. Pos. (strong, gelíc)
- o Pl. M. Acc. Pos. (strong, gelíc)
- o Pl. M. Gen. Pos. (strong, gelíc)
- o Pl. N. Gen. Pos. (strong, gelíc)
- o Pl. F. Gen. Pos. (strong, gelíc)
- o ***ge-líc***
  - ▪ noun
    - • infl. like word
      - o Sg. N. Nom. (strong, gelíc)
      - o Sg. N. Acc. (strong, gelíc)
      - o Pl. N. Nom. (strong, gelíc)
      - o Pl. N. Acc. (strong, gelíc)

GETIMBRODE

- o ***ge-timbran***
  - ▪ verb
    - • infl. like hyngrian
      - o Sg. 1. p. Ind. Pret. (w, 1, c, getimbrode)
      - o Sg. 3. p. Ind. Pret. (w, 1, c, getimbrode)
      - o Sg. Subj. Pret. (w, 1, c, getimbrode)
- o ***ge-tymbrian***
  - ▪ verb
    - • infl. like nerian
      - o Sg. 1. p. Ind. Pret. (w, 1, a, getymbrode)
      - o Sg. 3. p. Ind. Pret. (w, 1, a, getymbrode)
      - o Sg. Subj. Pret. (w, 1, a, getymbrode)
- o ***getimbrod***
  - ▪ participle
    - • infl. like halig
      - o Sg. M. Ins. Pos. (strong, past, getimbrode)
      - o Sg. N. Ins. Pos. (strong, past, getimbrode)
      - o Sg. F. Acc. Pos. (strong, past, getimbrode)
      - o Pl. M. Nom. Pos. (strong, past, getimbrode)
      - o Pl. M. Acc. Pos. (strong, past, getimbrode)
      - o Pl. F. Nom. Pos. (strong, past, getimbrode)
      - o Pl. F. Acc. Pos. (strong, past, getimbrode)
    - • infl. like blinda
      - o Sg. N. Nom. Pos. (weak, past, getimbrode)
      - o Sg. N. Acc. Pos. (weak, past, getimbrode)
      - o Sg. F. Nom. Pos. (weak, past, getimbrode)
- o ***getymbrod***
  - ▪ participle
    - • infl. like halig
      - o Sg. M. Ins. Pos. (strong, past, getymbrode)
      - o Sg. N. Ins. Pos. (strong, past, getymbrode)
      - o Sg. F. Acc. Pos. (strong, past, getymbrode)
      - o Pl. M. Nom. Pos. (strong, past, getymbrode)
      - o Pl. M. Acc. Pos. (strong, past, getymbrode)
      - o Pl. F. Nom. Pos. (strong, past, getymbrode)
      - o Pl. F. Acc. Pos. (strong, past, getymbrode)
    - • infl. like blinda
      - o Sg. N. Nom. Pos. (weak, past, getymbrode)
      - o Sg. N. Acc. Pos. (weak, past, getymbrode)
      - o Sg. F. Nom. Pos. (weak, past, getymbrode)

HIS

- o ***he***
  - ▪ pronoun
    - • infl. like he
      - o Sg. M. Gen. 3. p. (personal, possessive, reflexive, his)
      - o Sg. N. Gen. 3. p. (personal, possessive, reflexive, his)
- o ***his***
  - ▪ pronoun
    - • infl. like his
      - o Sg. M. Nom. 3. p. (personal, possessive, his)
      - o Sg. N. Nom. 3. p. (personal, possessive, his)

HÚS

- o ***húsian***
  - ▪ verb
    - • infl. like sealfian
      - o Sg. 2. p. Ind. Pres. (w, 2, a, húss)
      - o Sg. Imp. (w, 2, a, hús)
- o ***hús***
  - ▪ noun
    - • infl. like word
      - o Sg. N. Nom. (strong, hús)
      - o Sg. N. Acc. (strong, hús)
      - o Pl. N. Nom. (strong, hús)
      - o Pl. N. Acc. (strong, hús)

MÍN

- o ***ic***
  - ▪ pronoun
    - • infl. like ic
      - o Sg. Gen. 1. p. (personal, possessive, mín)
- o ***mín***
  - ▪ pronoun
    - • infl. like mín
      - o Sg. M. Nom. 1. p. (personal, possessive, mín)
      - o Sg. F. Nom. 1. p. (personal, possessive, mín)
      - o Sg. N. Nom. 1. p. (personal, possessive, mín)
      - o Sg. N. Acc. 1. p. (personal, possessive, mín)
      - o Pl. N. Nom. 1. p. (personal, possessive, mín)
      - o Pl. N. Acc. 1. p. (personal, possessive, mín)
    - • infl. like blind
      - o Sg. M. Nom. Pos. (strong, mín)

- o Sg. M. Acc. Pos. (strong, mín)
- o Sg. M. Dat. Pos. (strong, mín)
- o Sg. N. Nom. Pos. (strong, mín)
- o Sg. N. Acc. Pos. (strong, mín)
- o Sg. N. Dat. Pos. (strong, mín)
- o Sg. F. Nom. Pos. (strong, mín)
- o Sg. F. Gen. Pos. (strong, mín)
- o Sg. F. Dat. Pos. (strong, mín)
- o Pl. M. Nom. Pos. (strong, mín)
- o Pl. M. Acc. Pos. (strong, mín)
- o Pl. M. Gen. Pos. (strong, mín)
- o Pl. N. Nom. Pos. (strong, mín)
- o Pl. N. Acc. Pos. (strong, mín)
- o Pl. N. Gen. Pos. (strong, mín)
- o Pl. F. Gen. Pos. (strong, mín)
- o *mynian*
  - ▪ verb
    - • infl. like sealfian
      - o Sg. Imp. (w, 2, a, myn)
- o *min*
  - ▪ adjective
    - • infl. like blind
      - o Sg. M. Nom. Pos. (strong, min)
      - o Sg. M. Acc. Pos. (strong, min)
      - o Sg. M. Dat. Pos. (strong, min)
      - o Sg. N. Nom. Pos. (strong, min)
      - o Sg. N. Acc. Pos. (strong, min)
      - o Sg. N. Dat. Pos. (strong, min)
      - o Sg. F. Nom. Pos. (strong, min)
      - o Sg. F. Gen. Pos. (strong, min)
      - o Sg. F. Dat. Pos. (strong, min)
      - o Pl. M. Nom. Pos. (strong, min)
      - o Pl. M. Acc. Pos. (strong, min)
      - o Pl. M. Gen. Pos. (strong, min)
      - o Pl. N. Nom. Pos. (strong, min)
      - o Pl. N. Acc. Pos. (strong, min)
      - o Pl. N. Gen. Pos. (strong, min)
      - o Pl. F. Gen. Pos. (strong, min)

OFER

- o *ofer*
  - ▪ preposition
    - • infl. like ofer
      - o (ofer)
- o *of-ferian*
  - ▪ verb
    - • infl. like nerian
      - o Sg. Imp. (w, 1, a, offer)
- o *oferian*
  - ▪ verb
    - • infl. like nerian
      - o Sg. Imp. (w, 1, a, ofer)
- o *of-féran*
  - ▪ verb

- • infl. like déman
  - o Sg. Imp. (w, 1, a, offér)
- o *ófer*
  - ▪ noun
    - • infl. like stán
      - o Sg. M. Nom. (strong, ófer)
      - o Sg. M. Acc. (strong, ófer)

SE

- o *se*
  - ▪ pronoun
    - • infl. like se
      - o Sg. M. Nom. (demonstrative, se)
  - ▪ conjunction
    - • infl. like se
      - o (se)
- o *se*
  - ▪ adverb
    - • infl. like se
      - o Pos. (se)

STÁN

- o *stán*
  - ▪ noun
    - • infl. like stán
      - o Sg. M. Nom. (strong, stán)
      - o Sg. M. Acc. (strong, stán)

ÞÁ

- o *se*
  - ▪ pronoun
    - • infl. like se
      - o Sg. F. Acc. (demonstrative, þá)
      - o Pl. Nom. (demonstrative, þá)
      - o Pl. Acc. (demonstrative, þá)
- o *þá*
  - ▪ adverb
    - • infl. like þá
      - o Pos. (þá)

ÞǼM

- o *se*
  - ▪ pronoun
    - • infl. like se
      - o Sg. M. Dat. (demonstrative, þǽm)
      - o Sg. N. Dat. (demonstrative, þǽm)
      - o Pl. Dat. (demonstrative, þǽm)

ÞÁRA

- *se*
  - pronoun
    - infl. like se
      - Pl. Gen. (demonstrative, þára)

ÞÁS

- *þes*
  - pronoun
    - infl. like þes
      - Sg. F. Acc. (demonstrative, þás)
      - Pl. Nom. (demonstrative, þás)
      - Pl. Acc. (demonstrative, þás)
- *þás*
  - pronoun
    - infl. like blind
      - Sg. M. Nom. Pos. (strong, þás)
      - Sg. M. Acc. Pos. (strong, þás)
      - Sg. M. Dat. Pos. (strong, þás)
      - Sg. N. Nom. Pos. (strong, þás)
      - Sg. N. Acc. Pos. (strong, þás)
      - Sg. N. Dat. Pos. (strong, þás)
      - Sg. F. Nom. Pos. (strong, þás)
      - Sg. F. Gen. Pos. (strong, þás)
      - Sg. F. Dat. Pos. (strong, þás)
      - Pl. M. Nom. Pos. (strong, þás)
      - Pl. M. Acc. Pos. (strong, þás)
      - Pl. M. Gen. Pos. (strong, þás)
      - Pl. N. Nom. Pos. (strong, þás)
      - Pl. N. Acc. Pos. (strong, þás)
      - Pl. N. Gen. Pos. (strong, þás)
      - Pl. F. Gen. Pos. (strong, þás)
  - noun
    - infl. like stán
      - Sg. M. Nom. (strong, þás)
      - Sg. M. Acc. (strong, þás)
    - infl. like ár
      - Sg. F. Nom. (strong, þás)

ÞE

- *þu*
  - pronoun
    - infl. like þu
      - Sg. Dat. 2. p. (personal, þe)
      - Sg. Acc. 2. p. (personal, þe)
- *se*
  - pronoun
    - infl. like se
      - Sg. M. Nom. (demonstrative, þe)
- *þe*
  - conjunction
    - infl. like þe
      - (þe)
  - adverb
    - infl. like þe
      - Pos. (þe)

WERE

- *werian*
  - verb
    - infl. like nerian
      - Sg. Imp. (w, 1, a, were)
      - Sg. 1. p. Ind. Pres. (w, 1, a, were)
      - Sg. Subj. Pres. (w, 1, a, were)
- *werian*
  - verb
    - infl. like nerian
      - Sg. Imp. (w, 1, a, were)
      - Sg. 1. p. Ind. Pres. (w, 1, a, were)
      - Sg. Subj. Pres. (w, 1, a, were)
- *werian*
  - verb
    - infl. like nerian
      - Sg. Imp. (w, 1, a, were)
      - Sg. 1. p. Ind. Pres. (w, 1, a, were)
      - Sg. Subj. Pres. (w, 1, a, were)
- *wær*
  - adjective
    - infl. like blinda
      - Sg. N. Nom. Comp. (weak, werre)
      - Sg. N. Acc. Comp. (weak, werre)
      - Sg. F. Nom. Comp. (weak, werre)
- *wóh*
  - adjective
    - infl. like blinda
      - Sg. N. Nom. Comp. (weak, wére)
      - Sg. N. Acc. Comp. (weak, wére)
      - Sg. F. Nom. Comp. (weak, wére)
- *wer*
  - noun
    - infl. like stán
      - Sg. M. Dat. (strong, were)
- *wer*
  - noun
    - infl. like stán
      - Sg. M. Dat. (strong, were)
- *wer*
  - noun
    - infl. like stán
      - Sg. M. Dat. (strong, were)
- *wer*
  - noun
    - infl. like stán
      - Sg. M. Dat. (strong, were)

WÍSAN

- *wísian*

- ▪ verb
  - • infl. like sealfian
    - o Inf. (w, 2, a, wísan)
- o *wissian*
  - ▪ verb
    - • infl. like sealfian
      - o Inf. (w, 2, a, wissan)
- o *wísan*
  - ▪ verb
    - • infl. like déman
      - o Inf. (w, 1, a, wísan)
- o *wís*
  - ▪ adjective
    - • infl. like blinda
      - o Sg. M. Acc. Pos. (weak, wísan)
      - o Sg. M. Gen. Pos. (weak, wísan)
      - o Sg. M. Dat. Pos. (weak, wísan)
      - o Sg. N. Gen. Pos. (weak, wísan)
      - o Sg. N. Dat. Pos. (weak, wísan)
      - o Sg. F. Acc. Pos. (weak, wísan)
      - o Sg. F. Gen. Pos. (weak, wísan)
      - o Sg. F. Dat. Pos. (weak, wísan)
      - o Pl. M. Nom. Pos. (weak, wísan)
      - o Pl. M. Acc. Pos. (weak, wísan)
      - o Pl. N. Nom. Pos. (weak, wísan)
      - o Pl. N. Acc. Pos. (weak, wísan)
      - o Pl. F. Nom. Pos. (weak, wísan)
      - o Pl. F. Acc. Pos. (weak, wísan)
- o *wiss*
  - ▪ adjective
    - • infl. like blinda
      - o Sg. M. Acc. Pos. (weak, wissan)
      - o Sg. M. Gen. Pos. (weak, wissan)
      - o Sg. M. Dat. Pos. (weak, wissan)
      - o Sg. N. Gen. Pos. (weak, wissan)
      - o Sg. N. Dat. Pos. (weak, wissan)
      - o Sg. F. Acc. Pos. (weak, wissan)
      - o Sg. F. Gen. Pos. (weak, wissan)
      - o Sg. F. Dat. Pos. (weak, wissan)
      - o Pl. M. Nom. Pos. (weak, wissan)
      - o Pl. M. Acc. Pos. (weak, wissan)
      - o Pl. N. Nom. Pos. (weak, wissan)
      - o Pl. N. Acc. Pos. (weak, wissan)
      - o Pl. F. Nom. Pos. (weak, wissan)
      - o Pl. F. Acc. Pos. (weak, wissan)

- o *wísa*
  - ▪ noun
    - • infl. like fréa
      - o Sg. M. Acc. (weak, wísan)
      - o Sg. M. Gen. (weak, wísan)
      - o Sg. M. Dat. (weak, wísan)
      - o Pl. M. Nom. (weak, wísan)
      - o Pl. M. Acc. (weak, wísan)
- o *wíse*
  - ▪ noun
    - • infl. like tunge
      - o Sg. F. Acc. (weak, wísan)
      - o Sg. F. Gen. (weak, wísan)
      - o Sg. F. Dat. (weak, wísan)
      - o Pl. F. Nom. (weak, wísan)
      - o Pl. F. Acc. (weak, wísan)
- o *wíse*
  - ▪ noun
    - • infl. like tunge
      - o Sg. F. Acc. (weak, wísan)
      - o Sg. F. Gen. (weak, wísan)
      - o Sg. F. Dat. (weak, wísan)
      - o Pl. F. Nom. (weak, wísan)
      - o Pl. F. Acc. (weak, wísan)

WORD

- o *wordian*
  - ▪ verb
    - • infl. like sealfian
      - o Sg. Imp. (w, 2, a, word)
- o *word*
  - ▪ noun
    - • infl. like word
      - o Sg. N. Nom. (strong, word)
      - o Sg. N. Acc. (strong, word)
      - o Pl. N. Nom. (strong, word)
      - o Pl. N. Acc. (strong, word)

WYRCÞ

- o *wyrcan*
  - ▪ verb
    - • infl. like wyrcan
      - o Sg. 3. p. Ind. Pres. (w, 1, wyrcþ)

9.1.2.2. Analysis of the Automatically Generated Sample

Let us now consider and briefly comment the results for each token. We will focus mostly on problems or philologically surprising results.

ÆLC ("each"): As an indefinite pronoun, *ælc* declines like the adjective *blind*, therefore such a form without any ending has a high number (16) of possible functions.

AND ("and"): Although the identification of this conjunction should pose no problems, there are three lemmata with which it is identified. Not only is *and* also a preposition in OE which *BT* lists under a separate entry, there is also a weak verb *andian* and from a purely morphological point of view, there is no reason why this should not be a sg. imp. of that verb, as this form receives no ending. Perhaps the disambiguator should take into account word-class and function frequency and prefer the closed classes like conjunctions to verbs, due to the higher frequency of their members, on the one hand, and, on the other hand, take some bias against relatively rare functions like imperatives.

BIÞ: Correctly identified as 3. p. sg. indicative present form of *béon* ("to be"), the form is also identified as a member of the paradigms of several other verbs: *búgan*, *bíon* and *búan*. *Búgan* is listed twice, since there are two homonyms in the *BT*: one meaning "to bow" and the other "to inhabit". The first one is a strong verb with an alternative reduced stem which can regularly produce a form like *byþ* (note that i=y for these purposes). The second, however, is a weak verb according to the *BT* and therefore wrongly conjugated (3. p. sg. ind. pres. should give a form like *bugþ*). This is due to the stem comparison being deemed more accurate for stem assignment than e.g. the verb type. In future revisions, stem comparison matches should be conditioned by other morphological properties (like verb type or gender). On the other hand, another of the identified lemmata is *búan*, which is an anomalous spelling form of the verb *búgan* meaning "to dwell", so the conjugation of

this second *búgan* may not be quite wrong linguistically, as it is wrong in principle. *Bíon* is a variant of *béon* that receives a separate entry in the *BT*, but such variant entries should be manually marked in the dictionary to be recognized for variants by the Analyser. Whether the form actually represents *béon*, *búgan* or any of their variants or homonym is impossible to decide on the basis of morphology alone.

GEHÍERÞ: Three lemmata were identified with the 3. p. sg. indicative present form. Two of them are variants of the same verb *ge-hýran* ("to hear") one is a homonym meaning "to hire".

GELÍC: A relatively transparent form is identified as verbal imperative sg., as adjective in several different functions and as a neuter noun ("likeness") in nom./acc. The four verbal lemmata similar in form represent three lexemes ("to please", "to liken" and "to lick") since two of them are just variant forms (though each correctly following a different conjugation). The identification with the adjective *ge-líc* ("like") is straightforward, but the limited formal differentiation of the adjectival paradigm causes the form to be identified with 14 possible functions across three genders, four cases and two numbers.

GETIMBRODE: The form is identified with two variants of the verb *ge-timbrian* ("to build") as 1. and 3. p. sg. ind. pret. as well as sg. subj. pret. Notice that while the form is not identified as participle under the verbal lemmata (since it does not match the uninflected participle form), it is nevertheless identified with the corresponding inflected participial forms that are listed separately (we treat inflected participles as an independent word class). A large number of functions identified under the participial lemmata reflects again on the low formal differentiation of the adjectival paradigm.

HIS: *His* is identified with both 3. p. sg. masc. and neut. gen. of the personal pronoun *he* as well as nom. of the possessive *his*. The latter, however, links back to the *BT* entry of *he*.

This should pose little problems for a disambiguator, but remains a question of suitable presentation to the end-user as well as a theoretical question reflecting on selection of pronominal base forms.

HÚS: Identified correctly as a nom./acc. of the noun "house", it also matches the sg. imp. form of the derived verb *húsian* (this is a systemic problem that has no solution on this level). Interestingly, the form also matches the 2. p. sg. ind. pres. of the same verb – since the stem ends in *–s*, the sibilant ending[131] is merged with it in a variant form by the generator.

MÍN: Similarly to *his*, the form is identified both with the personal pronoun as well as the possessive. Unlike *his*, it is also identified with a number of incorrect functions, because *mín* is listed separately in the *BT* as an adjectival pronoun which is then assigned pronominal as well as adjectival paradigms. In addition to the pronominal functions, the form is also identified as a sg. imp. of the verb *mynian* ("to intend") and as the adjective *min* ("mean").

OFER: Apart from being identified as a preposition and a nom./acc. of the noun *ófer* ("edge"), the form is also interpreted as sg. imp. of three formally similar verbs: *of-ferian* ("to bear off"), *oferian* ("to exalt") and *of-féran* ("to overtake"). The single *f* in the form matches the *ff* in the two of the verbs because of the variation filter for geminate consonants that functions implicitly in this direction (i.e. a form with one consonant matching generated forms with two). This filter functions implicitly since it is taken care of already on the level of the generator. To prevent the filter from being applied even when there are matches not requiring any filter, we would need to mark the forms generated

---

[131] *-s* is one of the variant endings for the 2. p. sg. ind. pres. – see 7.2.1. above.

with the consonant reductions in a new field, which would then be checked during the matching process.

SE: The pronoun and conjunction are listed under one lemma, since *BT* treats them as a single lexeme. The form is also identified as an adverb "so".

STÁN: The form is the base form of the most frequent nominal paradigm and is identified accordingly.

ÞÁ: Like *sé,* the form represents a pronoun, conjunction and an adverb. The identification as a conjunction is missing due to an error in the digitized markup of the *BT* that was duly corrected.

ÞǢM: Identified correctly as a dative of a demonstrative pronoun *se.*

ÞÁRA: Identified correctly as a genitive of a demonstrative pronoun *se.*

ÞÁS: Identified correctly as a nom./acc. of a demonstrative pronoun *þes.*

ÞE: Apart from being an adverb/conjunction (*BT* collectively classes the lexeme as particle, which category is not used here), the form is also recognized as an analogical variant of sg. masc. nom. of pronoun *se* and a acc./dat. of 2. p. sg. personal pronoun *þu.*

WERE: *Were* is a witness to the homonymy and homography of OE lexicon (and morphology) on the one hand and to the difficulties with establishing discrete categories in the structure of a dead language on the other. The form can be identified as sg. imp., sg. subj. pres. and 1. p. sg. ind. pres. of three homonymous verbs under *werian* (i.e. 9 function-form pairs): "to hinder", "to cloth", "to remain"; sg. masc. dat. of four homonymous nouns *wer*: "man", "wergild", "troop", "weir"; and, finally, also as positive and comparative forms of two adjectives respectively: *wær* ("aware") and *wóh* ("crooked"). Several of these lexemes are undoubtedly etymologically connected and it is to a large degree up to an

arbitrary decision of the lexicographer where he draws the line between separate lexemes (and homonymy). This kind of ambiguity may be very difficult to deal with for any automatic disambiguator.

WÍSAN: *Wísan* is similar to the previous item in that it is identified with multiple lexical items of a semantic (and to some degree also etymological) group. The two verbs *wísian* and *wísan* are variants of the verb meaning "to direct". The connection to the verb *wissian* of a similar meaning is apparent, though a discussion of their relation is beyond the scope of this commentary. Likewise, the same applies to the two adjectives *wís* ("wise") and *wiss* ("certain"). Clearer semantic distinctions can be made between the three nouns with which the form is also identified: *wísa* ("leader"), *wíse* ("manner") and *wíse* ("sprout").

WORD: A form with a distinct ending is characteristically identified both as a sg. imp. of a derived verb *wordian* ("to speak") as well as the noun *word* itself. The grammatical functions analyzed include nom./acc. sg. as well as pl. since *word* is a neuter a-stem (paradigm historically to give rise to the zero plurals, such as PDE *fish* etc.)

WYRCÞ: The fricative ending clearly marks the form as a 3. p. sg. ind. pres. of *wyrcan* ("to work").

9.1.2.3.   Conclusions to the Analysis

Apart from the specific comments to each form, few general observations can be made based even on such a small sample. The basic ambition of the analyser – that is identifying the lemma and the grammatical functions of a particular form – appears to have been fulfilled very well, i.e. no lemma or function actually represented by the forms in the sentence is missing in the results (in other words, recall is 100%). However, in a sentence of 21 tokens, there were identified 67 lemmata and 240 function-form pairs. That is over three times (in the case of the lemmata) and over ten times (in the case of the function-

form pairs) the number actually implied by the text. [132] This ambiguity caused by the homography of OE forms, is mostly apparent in forms with endings that are common in paradigms of multiple word-classes, most obviously *–an* (ubiquitous in both adjectival and nominal weak declensions and also serving as an ending of verbal infinitives) or the zero ending (characteristic of base forms of nouns and adjectives as well as imp. sg. of verbs, but also sometimes identified[133] with the *–e* ending prolific in all three systems).

9.1.3. Accuracy

The common measures of accuracy in the field of information retrieval (including corpus linguistics and lemmatisation) are precision and recall. While precision is a ratio of the number of relevant results and the number of retrieved results, recall is a ratio of the number of relevant results and the number of items under consideration. Low precision thus means there is a lot of "noise" in the results and low recall means a lot of relevant results have not been retrieved.

Both precision and recall are important in measuring the accuracy of an analyser, but if our present tool is by design to be followed by a disambiguator, recall seems to be more important. In other words, low precision may pose a greater challenge for the disambiguator (greater choice), but with low recall, there would be nothing to disambiguate. In addition, as a measure, precision suffers rather more from the problem of relevance described in the following paragraph.

The precise definition of function of the analyser is important to measure its accuracy. This is not only a banal statement. The different functions outlined in the introduction may each require, or benefit from, slightly different results. In analysing a specific text,

---

[132] The result is more than twice the overall average of the tested texts and is therefore not typical. See Table 38 below for data from a larger sample.

[133] Specifically by application of a variation filter.

only one analysis is usually correct and expected by the user. In analysing a token out of context (e.g. in searching a dictionary), many feasible analyses are possible and usually desirable.

This, in short, is the problem of relevance in our results. Is a result that misses the lemma and/or function suggested by the man-made glossary automatically irrelevant and, in contrast, the result that includes them automatically relevant? In the case of the analysis of the particular text probably yes. But in case of a context-independent morphological analysis, neither of the options is quite true. A single token may have several morphologically feasible analyses. If one particular analysis is included, some others may be missing. If one particular analysis is missing, several other feasible analyses may be missing.

The problem is only exacerbated by the fact that we are dealing with a "dead" language. To determine whether a particular analysis is in fact feasible or "correct" we cannot rely on a comprehensive descriptive grammar, large representative corpora or judgment of native speakers. Is a form based on structurally sound morphological rules relevant, even if unattested? And what if a morphologically equivalent form that exhibits phonological or scribal variation is attested? How do these factors influence precision of the analyser? On the one hand, the chance that the analyser would have to deal with currently unattested forms is relatively small, but it is far from impossible, because, after all, *DOEC* only includes some versions of the known texts and new texts continue being discovered, especially in late OE.

In order to gauge the accuracy in a consistent and replicable manner, we consider its function in connection with a future disambiguator and thus a) consider such analysis by the analyser relevant that is consistent with the man-made glossary for the particular text

(i.e. that includes the results of the manual analysis and is therefore very probably correct at least for that particular text) and b) consider recall to be the more illuminating measure (since we have no meaningful and practical way to measure precision).

The percentage of recall for individual texts as well as the total recall for all tests is summarised in Table 38 below. Currently, the recall rate is at 95%, but we believe it can be improved by further 3 or 4% (see below).

name	description	genre	date	dialect	tokens	recall	rec. 2[134]
House on the Rock (or the Parable of the Wise and the Foolish Builders)	West Saxon Gospels, Matthew 7:24–27	religious prose	990	West-Saxon	100	100%	100%
Wise and Foolish Virgins (or the Parable of the Ten Virgins)	West Saxon Gospels, Matthew 25:1–13	religious prose	990	West-Saxon	194	100%	100%
The Voyages of Ohthere and Wulfstan	from Historiarum Adversum Paganos Libri VII	prose narrative	890	West-Saxon	347	99,4%	99,4%
Cynewulf and Cyneheard	entry for 754 AD in the Anglo-Saxon Chronicle (Parker MS)	historical prose	890	West-Saxon	313	99%	99%
Bede's "The Sun and the Moon"	from Ælfric's translation of Bede's De Temporibus	scientific prose	1000	West-Saxon	146	98,6%	98,6%
Wulfstan's Sermo Lupi ad Anglos	from Wulfstan's sermon	sermon in prose	1015	West-Saxon (written in Northumbria)	320	97,1%	**97,8%**
Beowulf	lines 702-757	heroic poetry	ca. 1000	mixed	313	96,8%	**98,4%**
Rushworth Gospels	interlinear gloss to Latin text of Matthew 6:1-6	religious prose	late 10c	Mercian	167	94,6%	**97%**
Lindisfarne Gospels	interlinear gloss to Latin text of Matthew 6:1-6	religious prose	10c	Northumbrian	183	89,6%	**90,7%**
Oswulf's Charter	charter	legal prose	806	Kentish	471	84,7%	**94,2%**
Total					**2554**	**95%**	**97,3%**

Table 38 Test texts and their recall.

That the performance of the analyser differs for each text is to be expected, but what are the variables that influence the accuracy most? Based on the table above, it is the

---

[134] Recall 2 are the values of recall with a modified variation filters, see chapter 9.1.4.

dialect, date of composition and genre of the sample. The more the text diverges from the best established and most prestigious tradition of (narrative and homiletic) religious prose (i.e. mostly the West-Saxon dialect of the 9th and 10th centuries) and from prose in general, the lower the recall. This again is no surprise: the analyser performs best on texts that best conform to the standard described in modern reference books that in turn are the basis of the analyser.

9.1.4. Typical Problems

What follows is a classification of typical problems encountered during the analysis, with examples and suggested solutions:

1) Expected ("standard") form was not generated:

   a) *norþmest* ("northernmost") was not identified as superlative of *norþ,* since the historically "double" superlatives in –*mest* were not generated. Although some are listed as independent lexemes in the *BT*, they have been usually marked as variants and are therefore not used by the generator. All such superlative forms should be added manually (based on Wright §446), see also chapter 2.3 note 20.

   b) *gebún* was not identified as past. part. of *búan* ("to dwell"), since although the verb is marked as anomalous in *BT*, it does not have its own paradigm in the verbal paradigmatic dictionary. All anomalous verbs in *BT* should be double-checked to have their paradigms in the verbal dictionary.

   c) *flotmen* was not identified as plural of *flotman* ("sailor"), since the noun was assigned to the paradigm of *stán* instead of the correct *mann* (notice the double *n*). Once the variants (like *man* for *mann*) are consistently marked in *BT*, these should also be used to match lexemes with paradigm examples.

2) Basic lexeme missing in *BT*:[135]

   a) *utræsde* was not identified as a preterit of *utræsan* ("to rush out"), since *BT* only

   lists the underived base verb *ræsan* and some of its several derivations, but not this

   one. Either the analyser would have to check input forms for affixes (as a type of

   filter), but that could have impact on speed and may add many false positives, or

   the generator would need to tackle word-formation by combining existing forms

   with derivational affixes. That would, however, greatly increase the already large

   number of generated forms.

3) Word-final/initial variation inside a complex form:

   a) *sibbegedriht* was not identified as *sibgedryht* ("band"), since the final –*e* variation

   filter is not applied to *sibbe* when it forms the first part of a compound.

   b) Similarly, the syncope in *wæpngewrixl* was not detected, since it is not word-final

   and the form was thus not identified with *wæpengewrixl* ("fight"). As in 2, either

   we need to generate the variants or we need to detect complex forms during the

   analysis proper.

4) Variation filter matches and replaces all occurrences (substrings) in the form instead

   of replacing only some of them:

   a) *gesene* was not identified as *geseon* ("to see"), since the variation filter *e*>*eo* is

   either applied to every *e* in the form or to none at all (and neither *gesene* nor

   geoseoneo match any record in the database).

   b) Similarly, *huommum* was not identified as a dative plural of *hwamm* ("corner") ,

   since the variation filter *u*>w is either applied to every *u* in the form or to none at

   all. We may apply filters to each match in succession, but this would multiply the

---

[135] The overall rarity of this type of problem is a testimony to the comprehensiveness of the *BT*.

number of database queries (from dozens to hundreds per form) and may have a significant impact on processing speed.

5) If there is a match without any variation filter applied to the input form, the filters are not applied to it. While this is the intended behaviour, it happens, in some cases, that the matched lexeme/function does not correspond to the one suggested by the context. The application of variation filters would cause the expected match with the corresponding lexeme/function, as show in the following examples:

a) *ymbe* matches *ymbe* (a relatively rare noun meaning a "swarm of bees"), but not *ymb* (an ubiquitous preposition meaning "about, by, around).

b) Similarly, *singa* is identified as a regular imperative singular of *syngian* ("to sin") rather than a variant of the infinitive of *syngan* ("to sing"). This problem can probably be solved only in connection with a disambiguator. The disambiguator could request additional variants from the analyser if the contextually derived score goes below a certain threshold. Otherwise, having variations filters apply to all forms would probably lower the accuracy of the disambiguator rather than improve it.

6) Variants unexpected either by the *BT*, the generator rules or the variation filters:

a) *simle* should be identified with *symbel* ("feast"). While marking and using the variants listed by *BT* may help in this example, we could also add the homorganic merger of *mb>m* into the variation filters.

b) *heolster* should match *heolstor* ("darkness"). We may add final *–or > –er* into the variation filters or formulate a more general filter that would take into account weakening of vowels in final syllables. This would, however, require either a syllable count (since the filter should not apply to monosyllables) or a generation

of a large number of variants. The former is a feasible solution and should be tested in the future versions.

c) *before* should match *beforan* ("before"). Similarly to the previous example, the weakening of the final unstressed syllable is the cause of the variation here and is to be expected in northern and/or late OE texts. It is not difficult to cover such variation per se, but it is important to remember that by admitting such variation into the variation filters we almost completely abandon the idea of endings as structural elements useful for identification of grammatical function (including word-class). Currently, the only feasible solution seems to be a multi-layered system of variation filters with increasingly "relaxing" filters being applied if the previous ones did not bring any matches. This would, however, only exacerbate the problem noticed under 5.

d) *wala* should be identified as a variant of the interjection *weg la* ("wellaway"). If the variant is not recorded by the *BT* and is at the same time unsystematic, the only solution is to add it to the list of manually defined forms.

e) *doas* should be identified as the third person singular indicative of *dón* ("to do"). The northern *–s* ending can be either generated for all third person singulars as a variant, or it can be added as a variation filter *þ > s*. An adverse effect of this step would be an ambiguity with the second person singular that is already generated having a variant *–s* ending.

f) *gearðad* is not identified as a plural indicative present of *geweorðian* ("to value") since the initial consonant of the root is syncopated after the prefix. The frequency of this phenomenon would need to be investigated further, but it may prove challenging to implement as a general rule for a variation filter. It would require prefix identification during the analysis and such identification would probably

rely on identifying both the prefix and the rest of the form (if only the prefix is matched, many forms with their root onsets identical to an existing prefix may be incorrectly analysed). Since it is the root that is modified here, it would be necessary to apply the variation filter during the prefix identification, which may in its turn lead to significant performance loss as well as to unforeseen problems.

g) Although a large number of variant forms was manually collected for personal pronouns (40 for the second person alone,) *iuch* is not among them and should be added manually.

h) Unpredictable variation may also arise from editorial emendations in an analysed text that were not recorded in (or available to) *BT*. Thus *scynscaþa* cannot be identified with *synsceaþa* ("miscreant"), unless the form is added into the dictionary.

The largest number of unmatched forms is due to unexpected variation that is either:

i) Scribal in nature, eg.: *cg > g*, *d > þ/(ð)*, *uu > w*; or

ii) phonological (though the two may not be easy to distinguish). In this case, it is especially the dialectal and diachronic differences in vowels that cause greatest difficulties.

On the basis of the texts tested, we have established two groups of vowel variations that could be added into the variation filters.

The first group was found mostly in the northern texts: *a = æ, ae > e, ae > æ, æ = e, æ > ea*.

The second group is distinctly Kentish: *e = oe, ea = eo, eo > e, o > e, o > ea, o > eo*.

Experimentally, we have added the scribal variation and the "northern" vowel variation into the variation filters of the analyser and have re-run the analysis of the tested texts.

The new recall values are listed as "rec. 2" in the Table 38 above. Interestingly, the recall rose for the northern text (*Lindisfarne* and *Rushworth Gospels*), but by much smaller amount than that of the Kentish *Oswulf's Charter.* This is mostly due to the fact that the recall value for the Kentish text was the lowest to start with.[136] By manual analysis, we have also determined that although the Kentish variants would improve the recall for the Kentish text by few more percent, they would have no impact on the northern texts. At the same time, due to their nature, they would have greater impact on the number of returned matches. In other words the "northern" variants stretch over a relatively narrow phonological space (mid front to low back), while the "Kentish" variants also incorporate the mid back /o/. While combining the latter variation with those listed in chapter 8.2.1. above would improve recall values, the resulting vowel correspondences would stretch over most of the phonological space rendering vowel differences almost meaningless.

Subject to further tests, it seems that instead of adding more and more vowel variants, other, more robust methods may be applied to take care of the currently unmatchable forms (see chapter 10.1. below).

---

[136] This was not initially expected. We have predicted the late and/or northern texts to pose greatest difficulties in the analysis, due to their weakened endings. However, after examining the results of the analysis, we have noticed that a) *BT* does almost never list Kentish variants as variant lexemes and if the variant is quoted in example texts, it is standardised and not listed as a variant at the beginning of the entry (even if the Kentish variant is the only attested form: e.g. the entry "fugel-dæg" and the quoted form "fuguldaeg"). And b) Wright mostly ignores Kentish variants in the example paradigms. Only Campbell seems to provide for all dialectal variants with the same thoroughness.

10. Conclusion

To conclude the paper, we will now look at possibilities of further improving the application itself and in connection with the proposed disambiguator. We will finish by assessing how much of our aims and objectives have been accomplished and what remains to be done.

10.1. Suggestions of Future Development

Further improvements should be aimed at three basic areas: 1. the accuracy and comprehensiveness of the input data; 2. the quality of the generator and analyser and 3. the usability of the output and the interface of the application.

10.1.1. The Input Data

Apart from the individual improvements below, the limited, but also highly reliable data from the Nijmegen lemmatisation project currently being developed by Erwin R. Komen should be integrated into the programme.[137]

A great deal of time and effort was already invested into the improvement of the *BT* data and their accessibility for end-user. This effort is valuable since it concerns not only the current project, but the widely used online version of *BT*. [138]

The dictionary tagging should be manually revised with several specific features in mind:

    1. In the macrostructure:

---

[137] See 1.3. above for the details of the Nijmegen project.
[138] See (Tichý, 2007) for the details of the *BT* digitization project.

a. "Editorial" entries (i.e. entries from the Supplement and the Addenda and Corrigenda that add, change or delete entries of the main volume should be unambiguously linked to their "primary" entries. That way, material from the editorial entries may be used by the analyser. Now they are tagged as "editorial" and their connection to "primary" entries is through the lemma and therefore it is ambiguous.

Consider the preposition *geond* ("through"). The related editorial entry in the Supplement adds four spelling variants to those of the main entry. If these are to be made use of (see point 2. a. below), we need an unambiguous link between the two related entries. At the moment, there are two entries for *geond* in the Main volume and two in the Supplement: one pair for the preposition and one for the adverb meaning "yonder" and these four entries are all linked together through their formal similarity.

b. Lexemes that are currently tagged as "linked entries" (i.e. entries primarily belonging to a different entry) should be distinguished by the type of their connection. They can be either variants spelling (e.g. the spelling variants *déogol* or *dígul* linked to the main entry *dígol* meaning "secret") or parts of paradigms (e.g. *dǽdon*, past pl. ind. of *dón*, "to do"). Neither of these are noted under the appropriate main entry and can, therefore, be very useful to the analyser.

c. Word-class tagging especially of the closed word-classes should be double checked (pronouns, for example, are often tagged very inclusively, i.e. they are tagged as pronouns, but also adjectives and nouns)

2. In the microstructure:

a. Variant spellings and parts of paradigm (inflectional forms) listed at the beginning of the entry should be tagged. Now they are tagged very broadly as "grammatical information" and therefore cannot be used by the analyser.

   Consider the wealth of morphological information in the entry of *habban* ("to have"): "*tó habbanne, hæbbene*; pres. part. *hæbbende*; pres. indic. *ic hæbbe, hafa, ðú hæfst, hafast, he hæfþ, hafaþ*, pl. *habbaþ, hæbbaþ*; p. *hæfde*; subj. *hæbbe*, pl. *hæbben, habban*; imper. *hafa*, pl. *habbaþ*; pp. *hæfed*." Not all forms listed in this manner are necessarily generated or even recognised by the analyser.

b. Modern English and Latin equivalents should be double-checked. These are generally tagged, but due to the inconsistent microstructure are not tagged correctly in all entries. The equivalents may be useful if the analyser is to be used in the future for generation of automatic glossaries or certain types of automatic tagging.

c. Entries with mangled (i.e. either not clearly structured in the paper *Dictionary* or incorrectly tagged in the online version) microstructure should be corrected.

At the same time, paradigmatic information should be better supplemented from Campbell to improve the results for dialectal texts. Currently, only verbal paradigms (through the verbal paradigmatic dictionary) and closed word-classes (through manually defined forms) draw extensively on variants listed by Campbell for each paradigm. Information on variants of nominal and adjectival declensions should be improved in similar manner.

10.1.2.   Quality of the Generator and the Analyser

The improvements of the quality of the application itself should be based largely on further results of its automatic textual analysis.

The paradigm assignment should be improved by incorporating the grammatical information known already on input (gender, verb type) and cases of lexeme homonymy should be investigated in *BT* and resolved either by tagging them with additional disambiguating grammatical information, or manually.

The form generation should be supplemented with more variants of grammatical affixes based on Campbell (as already suggested above) and test results to tackle as many variants already at this stage.

Greatest effort should perhaps be exerted on improving the matching of the input forms against the generated forms. Firstly, a multi-stage process should replace the current two (or three) stage process. Each stage should apply less and less specific variation filters until the last stage, where all unmatched forms should be matched using some kind of string distance algorithm.

10.1.3.   Output and the Interface

The output as well as the public interface should be further developed depending on the future use of the tool. Generally, more information should be made available to the end-user about the process by which the results were arrived at. The end user needs a better way to gauge the reliability of the results than by delving into the underlying database or scripts (in fact, these will not be easily accessible for most users). In particularly information on why a specific paradigm was assigned, how a certain form was generated and what variation filters were applied and why should be presented in a clear

manner (the latter two types of information are currently available on the output, but not in a user-friendly format).

Apart from that, interfaces to connect the analyser with the planned disambiguator as well as with *BT* need to be prepared and a version of the analyser for public use should be tested and refined. There are numerous performance and security related issues that have not been satisfactorily tackled yet.

10.2. Disambiguation

The essential part of any automatic morphological analysis (even in formally highly unambiguous agglutinating languages like Turkish) is disambiguation. Morphological ambiguity is "caused" mainly by natural lexical homonymy and homography and, in the case of inflectional languages like OE especially by grammatical homography. Some ambiguity may be ultimately unresolvable (or even intentional on the part of the speaker/writer) and some morphological data can contribute to resolving the ambiguity, but, most importantly, the ambiguity is usually resolved from the context.

We should therefore make sure that the analyser as it is now delivers to the disambiguator as much information needed for disambiguation as possible. This has been partly attempted, but not satisfactorily resolved, by the introduction of "probability" value that is derived for each form during the generation process and depends mostly on the probability of the morphological processes generating the forms (e.g. some forms of grammatical affixes are, given our knowledge of actual texts, less likely to appear than their functional/structural variants). Similarly, the currently available data on attestation and frequency in *DOEC*, as well as more general frequency data derived e.g. from the word-class and function, may also contribute to the decision-making process. That is, frequency

of the form or class-function combination (e.g. higher frequency of nouns in the nominative compared to verbs in the imperative) may be a positive contributing factor.

The context-dependent disambiguation may, like the analyser itself, take a solution developed for a typologically similar language as an example. Specifically, Jan Hajič's *Disambiguation of Rich Inflection* (Hajič, 2004) offers a possible path of development.

While planning a future disambiguator is far beyond the scope of this paper, it should be noted that syntactic rule based disambiguator may be much less successful for OE compared to languages with a much stricter word-order (e.g. PDE). Therefore, stochastic disambiguation based on the results of the analyser supplemented by the frequency and probability information listed above (and possibly some simple rules based on OE word-order) seems the best solution at the moment.

10.3.   Assessment of Aims and Objectives

All the five objectives that were explicitly set in chapter 1.3. were fulfilled by the research project. In conclusion, we will review them one by one and assess the state or degree of their completion:

1. *Objective*: "Simplify the process of finding entries in reference books – specifically, the tool will be embedded into the online version of the *Anglo-Saxon Dictionary* (*BT*)."

   Results: The tool is ready to be embedded into the online version of *BT*. In fact, the testing version is already hosted on the same machine as *BT* and shares the same Content Management System installation including its style sheets and libraries. The entries returned by the tool are also linked directly to the respective *BT* entries. What remains to be done is to integrate the analyser into the search

strategy of the online dictionary. *BT* now tries to accommodate various types of queries through one search box by a cascade of interconnected searches: first it searches the dictionary headwords, if there are no results it standardizes the user input to cater for major spelling variants and performs the search again; if there are still no results, it searches the full-text of the dictionary and if the search is still unsuccessful, it redirects the user to the advanced search page. The decision is still to be made as to whether the analyser should be inserted before the full-text search or whether it should be initiated only on user request. The decision will largely depend on the recall rates and performance impact on the live server when it is deemed fully stable and secure for deployment.

2. *Objective*: "Automatically generate all inflectional forms for a given paradigm (for corpus search and educational purposes)."

   Results: The objective has been fulfilled, but a better output for corpus querying could be devised. Either formally unique forms only should be outputted (e.g. one form only for nom. and acc. sg. of masc. *n*-stems nouns), or a regular expression could be automatically constructed from the list of the resulting forms. While the second solution is more demanding, it may overcome query length limitations set by certain corpora or corpus managers.

3. *Objective*: "Automatically create template glossaries that would simplify the manual creation of glosses to OE texts."

   Results: Template glossaries are currently the default type of the analyser's output, but they lack one key aspect to make them generally useful: the Present-Day English equivalents of the resulting lemmata. Currently, the lemmata are linked to the *BT* entries, which is fine for online glossaries, but not for printed materials.

Equivalents will be supplied to the analyser by incorporating data from the YCOE lemmatisation project and by the results of the manual revisions of the *BT* tagging (see chapter 10.1.1.).

4. *Objective*: "Make a semi-automatic (or assisted) lemmatisation and morphological analysis of large amounts of texts feasible."

Results: This objective is the least fulfilled in the sense that practical lemmatisation will require disambiguation of the analyser output. However, only a different formatting and interface is needed to make assisted lemmatisation possible now. The output in XML[139] should be one of the future development goals in any case, and there are plenty of XML tools to provide the necessary interface for manual disambiguation.

5. *Objective*: "Attempt to implement and, if necessary, modify frameworks already employed in automatic morphological analysis of typologically similar languages."

Results: If we can consider all the other objectives of the project fulfilled, we may also consider this goal successfully completed, since our project is based on previous frameworks as noted in chapter 4.1.

Going back to the three research questions we asked at the beginning of the paper (chapter 1.2.), we believe that the framework for morphological analysis of modern standardised languages such as the one introduced by Osolsobě (Osolsobě, 1996) can be successfully applied to languages that have been long dead but exhibit some degree of variation. For this purpose, the framework has been supplemented to deal with dialectal, diachronic and scribal variants of Old English in a more satisfactory way, with the

---

[139] Extensible Markup Language – currently the best open standard for textual data markup, sharing and archiving.

resulting recall values of 95% (which we plan to further improve). The problems in analysis characterised especially in chapters 9.1.3. and 9.1.4. point indeed to dialectal and diachronic variation as the major cause of the analyser's deficiencies, while grammatical homography, as noted in chapters 2.1., 7.9. and 9.1.2.2., is expected to be the greatest problem for the planned disambiguator.

We hope the project will help students, teachers and researchers of Old English as well as contribute to the knowledge of the field in general.

11. Bibliography

Adams, J. (2007). Retrieved from A Morphological Analyzer for Old English Verbs: http://www.cs.cmu.edu/~jmadams/NLPLabProject.pdf

Baker, P. S. (2012). *Common Spelling Variants*. Retrieved from The Electronic Introduction to Old English: http://www.wmich.edu/medieval/resources/IOE/variants.html

Baker, P. S. (2012). *Pronunciation*. Retrieved from The Electronic Introduction to Old English: http://www.wmich.edu/medieval/resources/IOE/pronunciation.html

Baker, P. S. (2012). *Verbs*. Retrieved from The Electronic Introduction to Old English: http://www.wmich.edu/medieval/resources/IOE/inflverb.html

Bosworth, J. (1921). *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth.* (T. N. Toller, Ed.) Oxford: Clarendon Press.

Calle Martín, J., & Triviño Rodríguez, J. L. (1998). Algoritmos de derivación de palabras con ortografía irregular en el análisis morfológico automático del Inglés Antiguo. *Interlingüística*, 67-70.

Campbell, A. (1983). Old English Grammar. Oxford: Clarendon Press.

Čermák, J., & Znojemská, H. (2001). *Čítanka staroanglických, středoanglických a raně novoanglických textů.* Praha: Nakladatelství Karolinum.

Hajič, J. (2004). Disambiguation of Rich Inflection. Prague: The Karolinum Press.

Hall, J. R. (1894). *A Concise Anglo-Saxon Dictionary.* London: Swan Sonnenschein & Co.

Healey, A. d. (Ed.). (2014). *Dictionary of Old English A to G Online*. Retrieved from http://www.doe.utoronto.ca/

Healey, A. d., Holland, J., McDougall, I., & Mielke, P. (2000). *The Dictionary of Old English Corpus in Electronic Form.* Toronto: University of Toronto.

Hogg, R. M., & Fulk, R. D. (2011). A Grammar of Old English: Morphology. Wiley-Blackwell.

Kay, C., Edmonds, F., Roberts, J., & Wotherspoon, I. (2005). *A Thesaurus of Old English.* Amsterdam: Rodopi.

Kiernan, K. (2011). Electronic Beowulf. London: British Library.

Kleinman, S. (20012). *A demo version of the OE lemmatiser.* Retrieved from http://www.csun.edu/english/lemmatise/main.php

Lindberg, T. (2014, February 7). Retrieved from Verbix: http://www.verbix.com/

Martín Arista, J. (2014, February 7). Retrieved from Nerthus Project: http://www.nerthusproject.com/

McGillivray, M. (2014). *Cynewulf and Cyneheard*. Retrieved from Old English Texts: http://www.ucalgary.ca/UofC/eduweb/engl401/texts/ohthfram.htm

McGillivray, M., & Chevallier, G. (2014). *The Voyage of Ohthere*. Retrieved from Old English Texts: http://www.ucalgary.ca/UofC/eduweb/engl401/texts/ohthfram.htm

Miranda García, A., Calle Martín, J., Moreno Olalla, D., & Muñoz González, G. (2006). The Old English Apollonius of Tyre in the light of the Old English Concordancer. In A. Renouf, & A. Kehoe, *The Changing Face of Corpus Linguistics* (pp. 91-98). Rodopi.

Miranda García, A., Triviño Rodríguez, J. L., & Calle Martín, J. (2000). MAOET: Morphological Analyser of Old English Texts. *Proceedings of the 10th International Conference of SELIM* (pp. 127-145). Zaragoza: Institución Fernando el Católico.

Mitchell, B., & Robinson, C. F. (2001). *A Guide to Old English* (6th ed.). Oxford: Blackwell Publishing.

Osolsobě, K. (1996). *Algoritmický popis české formální morfologie a strojový slovník čeština (unpublished disertation).* Brno: Masarykova Univerzita.

Oxford University Press. (2014). *OED Online.* Retrieved from http://www.oed.com/

Quirk, R., & Wrenn, C. L. (1957). An Old English Grammar. London: Routledge.

Rissanen, M., Kytö, M., Kahlas-Tarkka, L., Kilpiö, M., Nevanlinna, S., Taavitsainen, I., . . . Raumolin-Brunberg, H. (1991). *The Helsinki Corpus of English Texts.* Helsinki: University of Helsinki.

Sedláček, R. (1999). *Morfologický analyzátor češtiny (unpublished diploma thesis).* Brno: Masarykova Univerzita.

Sedláček, R., & Smrž, P. (2001, June). Automatic Processing of Czech Inflectional and Derivative Morphology. *FI MU Report Series.*

Sweet, H. (1887). *A Second Anglo-Saxon Reader: Archaic and Dialectal.* Oxford: Calendon Press.

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). The York-Toronto-Helsinki Parsed Corpus of Old English Prose.

Taylor, A., Warner, A., Pintzuk, S., & Plug, L. (2001). The York-Helsinki Parsed Corpus of Old English Poetry.

Tichý, O. (2007). *Digitization of Old and Middle English Dictionaries (unpublished thesis).* Prague: Faculty of Arts, Charles University in Prague.

Wright, J., & Wright, E. M. (1914). *Old English Grammar* . London: H. Milford, Oxford

University Press.

12. Appendices

12.1. List of Codes Used in the Grammatical Function Tags

Code	Function
Sg	Singular
Pl	Plural
Ma	Masculine
Fe	Feminine
Ne	Neutral
No	Nominative
Ac	Accusative
Ge	Genitive
Da	Dative
Is	Instrumental
Po	Positive
Co	Comparative
Sp	Superlative
1	1. person
2	2. person
3	3. person
If	Infinitive
In	Indicative
Su	Subjunctive
Im	Imperative
Ps	Present
Pa	Preterite
Pt	Participle
Id	Inflected

12.2. Verbal Paradigms

#ID	verb ID	type	class	subd	subc	§ W	var	paraID	pref	PreV	V	PoV	bnd	dent	end
1	bídan	s	1	0	a	490	0	Inf		b	í	d			an
1	bídan	s	1	0	a	490	0	PaInSg1		b	á	d			
1	bídan	s	1	0	a	490	0	PaInPl		b	i	d			on
1	bídan	s	1	0	a	490	0	Pp		b	i	d			en
2	sníðan	s	1	0	c	491	0	Inf		sn	í		ð		an
2	sníðan	s	1	0	c	491	0	PaInSg1		sn	á		ð		
2	sníðan	s	1	0	c	491	0	PaInPl		sn	i		d		on
2	sníðan	s	1	0	c	491	0	Pp		sn	i		d		en
3	rísan	s	1	0	c		0	Inf		r	í	s			an
3	rísan	s	1	0	c		0	PaInSg1		r	á	s			
3	rísan	s	1	0	c		0	PaInPl		r	i	s			on
3	rísan	s	1	0	c		0	Pp		r	i	s			en
4	téon	s	1	0	d	492	0	Inf		t	éo				n
4	téon	s	1	0	d	492	0	PaInSg1		t	á		h		
4	téon	s	1	0	d	492	0	PaInPl		t	i		g		on
4	téon	s	1	0	d	492	0	Pp		t	i		g		en
4	téon	s	1	0	d	492	1	Inf		t	ío				n
4	téon	s	1	0	d	492	2	Inf		t	éo		h		n
5	béodan	s	2	0	a	493	0	Inf		b	éo	d			an
5	béodan	s	2	0	a	493	0	PaInSg1		b	éa	d			
5	béodan	s	2	0	a	493	0	PaInPl		b	u	d			on
5	béodan	s	2	0	a	493	0	Pp		b	o	d			en
6	céosan	s	2	0	b	494	0	Inf		c	éo		s		an
6	céosan	s	2	0	b	494	0	PaInSg1		c	éa		s		
6	céosan	s	2	0	b	494	0	PaInPl		c	u		r		on
6	céosan	s	2	0	b	494	0	Pp		c	o		r		en
60	bréoðan	s	2	0	d		0	Inf		br	éo	ð			an
60	bréoðan	s	2	0	d		0	PaInSg1		br	éa	ð			
60	bréoðan	s	2	0	d		0	PaInPl		br	u	ð			on
60	bréoðan	s	2	0	d		0	Pp		br	o	ð			en
7	séoðan	s	2	0	c		0	Inf		s	éo		ð		an
7	séoðan	s	2	0	c		0	PaInSg1		s	éa		ð		
7	séoðan	s	2	0	c		0	PaInPl		s	u		d		on
7	séoðan	s	2	0	c		0	Pp		s	o		d		en
8	téon	s	2	0	d	495	0	Inf		t	éo				n
8	téon	s	2	0	d	495	0	PaInSg1		t	éa		h		
8	téon	s	2	0	d	495	0	PaInPl		t	u		g		on
8	téon	s	2	0	d	495	0	Pp		t	o		g		en
8	téon	s	2	0	d	495	1	Inf		t	éo		h		n
9	brúcan	s	2	0	e	496	0	Inf		br	ú	c			an
9	brúcan	s	2	0	e	496	0	PaInSg1		br	éa	c			
9	brúcan	s	2	0	e	496	0	PaInPl		br	u	c			on

9	brúcan	s	2	0	e	496	0	Pp		br	o	c			en
61	búgan	s	2	0	f		0	Inf		b	ú	g			an
61	búgan	s	2	0	f		0	PaInSg1		b	éa	g			
61	búgan	s	2	0	f		1	PaInSg1		b	éa		h		
61	búgan	s	2	0	f		0	PaInPl		b	u	g			on
61	búgan	s	2	0	f		0	Pp		b	o	g			en
61	búgan	s	2	0	f		1	Inf		b	ú				an
61	búgan	s	2	0	f		2	Inf		b	ú		h		an
61	búgan	s	2	0	f		2	PaInSg1		b	éa				
61	búgan	s	2	0	f		2	PaInPl		b	u				on
61	búgan	s	2	0	f		2	Pp		b	o				en
62	smúgan	s	2	0	f		0	Inf		sm	ú	g			an
62	smúgan	s	2	0	f		0	PaInSg1		sm	éa	g			
62	smúgan	s	2	0	f		1	PaInSg1		sm	éa		h		
62	smúgan	s	2	0	f		0	PaInPl		sm	u	g			on
62	smúgan	s	2	0	f		0	Pp		sm	o	g			en
62	smúgan	s	2	0	f		1	Inf		sm	ú		h		an
10	bindan	s	3	0	a	498	0	Inf		b	i	nd			an
10	bindan	s	3	0	a	498	0	PaInSg1		b	a	nd			
10	bindan	s	3	0	a	498	0	PaInPl		b	u	nd			on
10	bindan	s	3	0	a	498	0	Pp		b	u	nd			en
10	bindan	s	3	0	a	498	1	PaInSg1		b	o	nd			
11	findan	s	3	0	b		0	Inf		f	i	nd			an
11	findan	s	3	0	b		0	PaInSg1		f	a	nd			
11	findan	s	3	0	b		0	PaInPl		f	u	nd			on
11	findan	s	3	0	b		0	Pp		f	u	nd			en
11	findan	s	3	0	b		1	PaInSg1		f	u	nd			
12	birnan	s	3	0	c		0	Inf		b	i	rn			an
12	birnan	s	3	0	c		0	PaInSg1		b	a	rn			
12	birnan	s	3	0	c		0	PaInPl		b	u	rn			on
12	birnan	s	3	0	c		0	Pp		b	u	rn			en
12	birnan	s	3	0	c		1	Inf		b	ie	rn			an
12	birnan	s	3	0	c		1	PaInSg1		b	o	rn			
12	birnan	s	3	0	c		2	PaInSg1		b	ea	rn			
13	helpan	s	3	0	d	499	0	Inf		h	e	lp			an
13	helpan	s	3	0	d	499	0	PaInSg1		h	ea	lp			
13	helpan	s	3	0	d	499	0	PaInPl		h	u	lp			on
13	helpan	s	3	0	d	499	0	Pp		h	o	lp			en
63	gieldan	s	3	0	e		0	Inf		g	ie	ld			an
63	gieldan	s	3	0	e		0	PaInSg1		g	ea	ld			
63	gieldan	s	3	0	e		0	PaInPl		g	u	ld			on
63	gieldan	s	3	0	e		0	Pp		g	o	ld			en
64	gellan	s	3	0	e		0	Inf		g	e	ll			an
64	gellan	s	3	0	e		0	PaInSg1		g	ea	ll			
64	gellan	s	3	0	e		0	PaInPl		g	u	ll			on
64	gellan	s	3	0	e		0	Pp		g	o	ll			en

64	gellan	s	3	0	e		1	Inf		g	ie	ll		an
50	gilpan	s	3	0	e		0	Inf		g	i	lp		an
50	gilpan	s	3	0	e		0	PaInSg1		g	ea	lp		
50	gilpan	s	3	0	e		0	PaInPl		g	u	lp		on
50	gilpan	s	3	0	e		0	Pp		g	o	lp		en
50	gilpan	s	3	0	e		1	Inf		g	ie	lp		an
14	weorpan	s	3	0	e	500	0	Inf		w	eo	rp		an
14	weorpan	s	3	0	e	500	0	PaInSg1		w	ea	rp		
14	weorpan	s	3	0	e	500	0	PaInPl		w	u	rp		on
14	weorpan	s	3	0	e	500	0	Pp		w	o	rp		en
14	weorpan	s	3	0	e	500	1	Inf		w	ea	rp		an
15	féolan	s	3	0	f	501	0	Inf		f	éo	l		an
15	féolan	s	3	0	f	501	0	PaInSg1		f	ea	l	h	
15	féolan	s	3	0	f	501	0	PaInPl		f	u	l	g	on
15	féolan	s	3	0	f	501	0	Pp		f	o	l	g	en
15	féolan	s	3	0	f	501	1	Inf		f	éo	l	h	an
15	féolan	s	3	0	f	501	1	PaInPl		f	ǽ	l		on
15	féolan	s	3	0	f	501	2	PaInPl		f	ú	l		on
15	féolan	s	3	0	f	501	1	Pp		f	ó	l		en
16	bregdan	s	3	0	g		0	Inf		br	e		gd	an
16	bregdan	s	3	0	g		0	PaInSg1		br	æ		gd	
16	bregdan	s	3	0	g		0	PaInPl		br	u		gd	on
16	bregdan	s	3	0	g		0	Pp		br	o		gd	en
16	bregdan	s	3	0	g		1	Inf		br	é		d	an
16	bregdan	s	3	0	g		1	PaInSg1		br	ǽ		d	
16	bregdan	s	3	0	g		1	PaInPl		br	ú		d	on
16	bregdan	s	3	0	g		1	Pp		br	ó		d	en
56	stregdan	s	3	0	g		0	Inf		str	e		gd	an
56	stregdan	s	3	0	g		0	PaInSg1		str	æ		gd	
56	stregdan	s	3	0	g		0	PaInPl		str	u		gd	on
56	stregdan	s	3	0	g		0	Pp		str	o		gd	en
56	stregdan	s	3	0	g		1	Inf		str	é		d	an
56	stregdan	s	3	0	g		1	PaInSg1		str	ǽ		d	
56	stregdan	s	3	0	g		1	PaInPl		str	ú		d	on
56	stregdan	s	3	0	g		1	Pp		str	ó		d	en
57	berstan	s	3	0	g		0	Inf		b	e	rst		an
57	berstan	s	3	0	g		0	PaInSg1		b	æ	rst		
57	berstan	s	3	0	g		0	PaInPl		b	u	rst		on
57	berstan	s	3	0	g		0	Pp		b	o	rst		en
59	ðerscan	s	3	0	g		0	Inf		ð	e	rsc		an
59	ðerscan	s	3	0	g		0	PaInSg1		ð	æ	rsc		
59	ðerscan	s	3	0	g		0	PaInPl		ð	u	rsc		on
59	ðerscan	s	3	0	g		0	Pp		ð	o	rsc		en
17	frignan	s	3	0	h		0	Inf		fr	i		gn	an
17	frignan	s	3	0	h		0	PaInSg1		fr	æ		gn	
17	frignan	s	3	0	h		0	PaInPl		fr	u		gn	on

17	frignan	s	3	0	h		0	Pp		fr	u		gn	en
17	frignan	s	3	0	h		1	Inf		fr	í		n	an
17	frignan	s	3	0	h		1	PaInSg1		fr	á		n	
18	spornan	s	3	0	i		0	Inf		sp	o	rn		an
18	spornan	s	3	0	i		0	PaInSg1		sp	ea	rn		
18	spornan	s	3	0	i		0	PaInPl		sp	u	rn		on
18	spornan	s	3	0	i		0	Pp		sp	o	rn		en
18	spornan	s	3	0	i		1	Inf		sp	u	rn		an
58	murnan	s	3	0	i		0	Inf		m	u	rn		an
58	murnan	s	3	0	i		0	PaInSg1		m	ea	rn		
58	murnan	s	3	0	i		0	PaInPl		m	u	rn		on
58	murnan	s	3	0	i		0	Pp		m	o	rn		en
58	murnan	s	3	0	i		1	Inf		m	o	rn		an
19	beran	s	4	0	a	503	0	Inf		b	e	r		an
19	beran	s	4	0	a	503	0	PaInSg1		b	æ	r		
19	beran	s	4	0	a	503	0	PaInPl		b	ǽ	r		on
19	beran	s	4	0	a	503	0	Pp		b	o	r		en
20	cuman	s	4	0	b		0	Inf		c	u	m		an
20	cuman	s	4	0	b		0	PaInSg1		c	ó	m		
20	cuman	s	4	0	b		0	PaInPl		c	ó	m		on
20	cuman	s	4	0	b		0	Pp		c	u	m		en
20	cuman	s	4	0	b		1	PaInSg1		cw	ó	m		
20	cuman	s	4	0	b		1	PaInPl		cw	ó	m		on
20	cuman	s	4	0	b		1	Pp		c	y	m		en
21	niman	s	4	0	c		0	Inf		n	i	m		an
21	niman	s	4	0	c		0	PaInSg1		n	ó	m		
21	niman	s	4	0	c		0	PaInPl		n	ó	m		on
21	niman	s	4	0	c		0	Pp		n	u	m		en
21	niman	s	4	0	c		1	PaInSg1		n	a	m		
21	niman	s	4	0	c		1	PaInPl		n	á	m		on
22	metan	s	5	0	a	505	0	Inf		m	e	t		an
22	metan	s	5	0	a	505	0	PaInSg1		m	æ	t		
22	metan	s	5	0	a	505	0	PaInPl		m	ǽ	t		on
22	metan	s	5	0	a	505	0	Pp		m	e	t		en
23	wegan	s	5	0	c		0	Inf		w	e	g		an
23	wegan	s	5	0	c		0	PaInSg1		w	æ	g		
23	wegan	s	5	0	c		0	PaInPl		w	ǽ	g		on
23	wegan	s	5	0	c		0	Pp		w	e	g		en
23	wegan	s	5	0	c		1	PaInPl		w	á	g		on
24	giefan	s	5	0	d		0	Inf		g	ie	f		an
24	giefan	s	5	0	d		0	PaInSg1		g	ea	f		
24	giefan	s	5	0	d		0	PaInPl		g	éa	f		on
24	giefan	s	5	0	d		0	Pp		g	ie	f		en
24	giefan	s	5	0	d		1	Inf		g	i	f		an
24	giefan	s	5	0	d		1	PaInSg1		g	e	f		
24	giefan	s	5	0	d		2	PaInSg1		g	a	f		

24	giefan	s	5	0	d		3	PaInSg1		g	æ	f			
24	giefan	s	5	0	d		1	Pp		g	i	f			en
65	gitan	s	5	0	d		0	Inf		g	ie	t			an
65	gitan	s	5	0	d		0	PaInSg1		g	ea	t			
65	gitan	s	5	0	d		0	PaInPl		g	éa	t			on
65	gitan	s	5	0	d		0	Pp		g	ie	t			en
65	gitan	s	5	0	d		1	Inf		g	i	t			an
65	gitan	s	5	0	d		1	PaInSg1		g	e	t			
65	gitan	s	5	0	d		2	PaInSg1		g	a	t			
65	gitan	s	5	0	d		3	PaInSg1		g	æ	t			
65	gitan	s	5	0	d		1	Pp		g	i	t			en
25	cweðan	s	5	0	e		0	Inf		cw	e		ð		an
25	cweðan	s	5	0	e		0	PaInSg1		cw	æ		ð		
25	cweðan	s	5	0	e		0	PaInPl		cw	ǽ			d	on
25	cweðan	s	5	0	e		0	Pp		cw	e			d	en
26	wesan	s	5	0	f		0	Inf		w	e			s	an
26	wesan	s	5	0	f		0	PaInPl		w	ǽ			r	on
27	séon	s	5	0	g	506	0	Inf		s	éo				n
27	séon	s	5	0	g	506	0	PaInSg1		s	ea			h	
27	séon	s	5	0	g	506	0	PaInPl		s	á			w	on
27	séon	s	5	0	g	506	0	Pp		s	e			w	en
27	séon	s	5	0	g	506	1	PaInPl		s	ǽ			g	on
27	séon	s	5	0	g	506	1	Pp		s	a			w	en
27	séon	s	5	0	g	506	1	Inf		s	éo			h	n
27	séon	s	5	0	g	506	2	Pp	ge	s	e			g	en
28	biddan	s	5	0	h	507	0	Inf		b	i	d			an
28	biddan	s	5	0	h	507	0	PaInSg1		b	æ	d			
28	biddan	s	5	0	h	507	0	PaInPl		b	ǽ	d			on
28	biddan	s	5	0	h	507	0	Pp		b	e	d			en
29	licgan	s	5	0	i		0	Inf		l	i		cg		an
29	licgan	s	5	0	i		0	PaInPl		l	á		g		on
29	licgan	s	5	0	i		1	PaInPl		l	ǽ		g		on
30	fricgan	s	5	0	j		0	Inf		fr	i		cg		an
30	fricgan	s	5	0	j		0	Pp	ge	fr	i			g	en
30	fricgan	s	5	0	j		1	Pp	ge	fr	u			g	en
31	faran	s	6	0	a	508	0	Inf		f	a	r			an
31	faran	s	6	0	a	508	0	PaInSg1		f	ó	r			
31	faran	s	6	0	a	508	0	PaInPl		f	ó	r			on
31	faran	s	6	0	a	508	0	Pp		f	a	r			en
31	faran	s	6	0	a	508	1	Pp		f	æ	r			en
32	spanan	s	6	0	b		0	Inf		sp	a	n			an
32	spanan	s	6	0	b		0	PaInSg1		sp	ó	n			
32	spanan	s	6	0	b		0	PaInPl		sp	ó	n			on
32	spanan	s	6	0	b		0	Pp		sp	a	n			en
32	spanan	s	6	0	b		1	PaInSg1		sp	éo	n			
32	spanan	s	6	0	b		1	Pp		sp	æ	n			en

33	sléan	s	6	0	c	509	0	Inf		sl	éa				n
33	sléan	s	6	0	c	509	0	PaInSg1		sl	ó			g	
33	sléan	s	6	0	c	509	0	PaInPl		sl	ó			g	on
33	sléan	s	6	0	c	509	0	Pp		sl	a			g	en
33	sléan	s	6	0	c	509	1	PaInSg1		sl	ó			h	
33	sléan	s	6	0	c	509	1	Pp		sl	æ			g	en
33	sléan	s	6	0	c	509	2	Pp		sl	e			g	en
33	sléan	s	6	0	c	509	1	Inf		sl	éa			h	n
34	hebban	s	6	0	d		0	Inf		h	e			bb	an
34	hebban	s	6	0	d		0	PaInSg1		h	ó			f	
34	hebban	s	6	0	d		0	PaInPl		h	ó			f	on
34	hebban	s	6	0	d		0	Pp		h	a			f	en
34	hebban	s	6	0	d		1	Pp		h	æ			f	en
34	hebban	s	6	0	d		2	Pp		h	e			f	en
34	hebban	s	6	0	d		1	Inf		h	e			bb	an
35	hlehhan	s	6	0	e		0	Inf		hl	e			bb	an
35	hlehhan	s	6	0	e		0	PaInSg1		hl	ó			h	
35	hlehhan	s	6	0	e		0	PaInPl		hl	ó			g	on
35	hlehhan	s	6	0	e		0	Pp		hl	a			g	en
35	hlehhan	s	6	0	e		1	PaInSg1		hl	ó			g	
35	hlehhan	s	6	0	e		1	Pp		hl	æ			g	en
35	hlehhan	s	6	0	e		1	Inf		hl	æ			h	an
35	hlehhan	s	6	0	e		2	Inf		hl	æ			hh	an
36	sceððan	s	6	0	f		0	Inf		sc	e			ðð	an
36	sceððan	s	6	0	f		0	PaInSg1		sc	ó			d	
36	sceððan	s	6	0	f		0	PaInPl		sc	ó			d	on
36	sceððan	s	6	0	f		1	Inf		sc	ie			ðð	an
37	scippan	s	6	0	g		0	Inf		sc	ie			pp	an
37	scippan	s	6	0	g		0	PaInSg1		sc	ó			p	
37	scippan	s	6	0	g		0	PaInPl		sc	ó			p	on
37	scippan	s	6	0	g		0	Pp		sc	ea			p	en
37	scippan	s	6	0	g		1	Inf		sc	i			pp	an
38	stæppan	s	6	0	h		0	Inf		st	æ			pp	an
38	stæppan	s	6	0	h		0	PaInSg1		st	ó			p	
38	stæppan	s	6	0	h		0	PaInPl		st	ó			p	on
38	stæppan	s	6	0	h		0	Pp		st	a			p	en
38	stæppan	s	6	0	h		1	Pp		st	æ			p	en
39	swerian	s	6	0	i		0	Inf		sw	e	r			an
39	swerian	s	6	0	i		0	PaInSg1		sw	ó	r			
39	swerian	s	6	0	i		0	PaInPl		sw	ó	r			on
39	swerian	s	6	0	i		0	Pp		sw	o	r			en
40	hátan	s	7	0	a		0	Inf		h	á			t	an
40	hátan	s	7	0	a		0	PaInSg1		h	é			t	
40	hátan	s	7	0	a		0	PaInPl		h	é			t	on
40	hátan	s	7	0	a		0	Pp		h	á			t	en
40	hátan	s	7	0	a		1	PaInSg1		h	é			ht	

40	hátan	s	7	0	a		1	PaInPl		h	é		ht		on
66	lácan	s	7	0	a		0	Inf		l	á		c		an
66	lácan	s	7	0	a		0	PaInSg1		l	é		c		
66	lácan	s	7	0	a		0	PaInPl		l	é		c		on
66	lácan	s	7	0	a		0	Pp		l	á		c		en
66	lácan	s	7	0	a		1	PaInSg1		l	eo		lc		
66	lácan	s	7	0	a		1	PaInPl		l	eo		lc		on
67	scéadan	s	7	0	a		0	Inf		sc	éa	d			an
67	scéadan	s	7	0	a		0	PaInSg1		sc	éa	d			
67	scéadan	s	7	0	a		0	PaInPl		sc	éa	d			on
67	scéadan	s	7	0	a		0	Pp		sc	á	d			en
67	scéadan	s	7	0	a		1	Inf		sc	á	d			an
67	scéadan	s	7	0	a		1	PaInSg1		sc	é	d			
67	scéadan	s	7	0	a		1	PaInPl		sc	é	d			on
41	lǽtan	s	7	0	b		0	Inf		l	ǽ		t		an
41	lǽtan	s	7	0	b		0	PaInSg1		l	é		t		
41	lǽtan	s	7	0	b		0	PaInPl		l	é		t		on
41	lǽtan	s	7	0	b		0	Pp		l	ǽ		t		en
41	lǽtan	s	7	0	b		1	PaInSg1		l	eo		t		
41	lǽtan	s	7	0	b		1	PaInPl		l	eo		rt		on
68	drǽdan	s	7	0	b		0	Inf		dr	ǽ		d		an
68	drǽdan	s	7	0	b		0	PaInSg1		dr	é		d		
68	drǽdan	s	7	0	b		0	PaInPl		dr	é		d		on
68	drǽdan	s	7	0	b		0	Pp		dr	ǽ		d		en
68	drǽdan	s	7	0	b		1	PaInSg1		dr	eo		rd		
68	drǽdan	s	7	0	b		1	PaInPl		dr	eo		rd		on
69	rǽdan	s	7	0	b		0	Inf		r	ǽ		d		an
69	rǽdan	s	7	0	b		0	PaInSg1		r	é		d		
69	rǽdan	s	7	0	b		0	PaInPl		r	é		d		on
69	rǽdan	s	7	0	b		0	Pp		r	ǽ		d		en
69	rǽdan	s	7	0	b		1	PaInSg1		r	eo		rd		
69	rǽdan	s	7	0	b		1	PaInPl		r	eo		rd		on
70	slǽpan	s	7	0	b		0	Inf		sl	ǽ	p			an
70	slǽpan	s	7	0	b		0	PaInSg1		sl	é	p			
70	slǽpan	s	7	0	b		0	PaInPl		sl	é	p			on
70	slǽpan	s	7	0	b		0	Pp		sl	ǽ	p			en
70	slǽpan	s	7	0	b		1	PaInSg1		sl	eo	p			
70	slǽpan	s	7	0	b		1	PaInPl		sl	eo	p			on
71	blandan	s	7	0	b		0	Inf		bl	a	nd			an
71	blandan	s	7	0	b		0	PaInSg1		bl	é	nd			
71	blandan	s	7	0	b		0	PaInPl		bl	é	nd			on
71	blandan	s	7	0	b		0	Pp		bl	a	nd			en
42	fón	s	7	0	c	514	0	Inf		f	ó				n
42	fón	s	7	0	c	514	1	Inf		f	ó		h		n
42	fón	s	7	0	c	514	0	PaInSg1		f	é		ng		
42	fón	s	7	0	c	514	0	PaInPl		f	é		ng		on

42	fón	s	7	0	c	514	0	Pp		f	a		ng		en
43	bannan	s	7	0	d	515	0	Inf		b	a	nn			an
43	bannan	s	7	0	d	515	0	PaInSg1		b	éo	nn			
43	bannan	s	7	0	d	515	0	PaInPl		b	éo	nn			on
43	bannan	s	7	0	d	515	0	Pp		b	a	nn			en
43	bannan	s	7	0	d	515	1	PaInSg1		b	éo		n		
43	bannan	s	7	0	d	515	1	PaInPl		b	éo		n		on
44	gangan	s	7	0	e		0	Inf		g	a	ng			an
44	gangan	s	7	0	e		0	PaInSg1		g	éo	ng			
44	gangan	s	7	0	e		0	PaInPl		g	éo	ng			on
44	gangan	s	7	0	e		0	Pp		g	a	ng			en
44	gangan	s	7	0	e		1	PaInSg1		g	íe	ng			
44	gangan	s	7	0	e		1	PaInPl		g	íe	ng			on
45	fealdan	s	7	0	f	516	0	Inf		f	ea	ld			an
45	fealdan	s	7	0	f	516	0	PaInSg1		f	éo	ld			
45	fealdan	s	7	0	f	516	0	PaInPl		f	éo	ld			on
45	fealdan	s	7	0	f	516	0	Pp		f	ea	ld			en
46	bláwan	s	7	0	g	517	0	Inf		bl	á	w			an
46	bláwan	s	7	0	g	517	0	PaInSg1		bl	éo	w			
46	bláwan	s	7	0	g	517	0	PaInPl		bl	éo	w			on
46	bláwan	s	7	0	g	517	0	Pp		bl	á	w			en
72	hléapan	s	7	0	h	518	0	Inf		hl	éa	p			an
72	hléapan	s	7	0	h	518	0	PaInSg1		hl	éo	p			
72	hléapan	s	7	0	h	518	0	PaInPl		hl	éo	p			on
72	hléapan	s	7	0	h	518	0	Pp		hl	éa	p			en
47	béatan	s	7	0	h		0	Inf		b	éa		t		an
47	béatan	s	7	0	h		0	PaInSg1		b	éo		t		
47	béatan	s	7	0	h		0	PaInPl		b	éo		t		on
47	béatan	s	7	0	h		0	Pp		b	éa		t		en
47	béatan	s	7	0	h		1	PaInSg1		b	éo		ft		
47	béatan	s	7	0	h		1	PaInPl		b	éo		ft		on
48	blótan	s	7	0	i	519	0	Inf		bl	ó	t			an
48	blótan	s	7	0	i	519	0	PaInSg1		bl	éo	t			
48	blótan	s	7	0	i	519	0	PaInPl		bl	éo	t			on
48	blótan	s	7	0	i	519	0	Pp		bl	ó	t			en
49	stígan	s	1	0	b		0	Inf		st	í		g		an
49	stígan	s	1	0	b		0	PaInSg1		st	á		g		
49	stígan	s	1	0	b		0	PaInPl		st	i		g		on
49	stígan	s	1	0	b		0	Pp		st	i		g		en
49	stígan	s	1	0	b		1	Inf		st	í		h		an
49	stígan	s	1	0	b		1	PaInSg1		st	á		h		
51	míðan	s	1	0	c		0	Inf		m	í	ð			an
51	míðan	s	1	0	c		0	PaInSg1		m	á	ð			
51	míðan	s	1	0	c		0	PaInPl		m	i	ð			on
51	míðan	s	1	0	c		0	Pp		m	i	ð			en
52	wríðan	s	1	0	c		0	Inf		wr	í	ð			an

52	wríðan	s	1	0	c		0	PaInSg1	wr	á		ð		
52	wríðan	s	1	0	c		0	PaInPl	wr	i		ð		on
52	wríðan	s	1	0	c		0	Pp	wr	i		ð		en
53	hnígan	s	1	0	a		0	Inf	hn	í		g		an
53	hnígan	s	1	0	a		0	PaInSg1	hn	á		g		
53	hnígan	s	1	0	a		0	PaInPl	hn	i		g		on
53	hnígan	s	1	0	a		0	Pp	hn	i		g		en
53	hnígan	s	1	0	a		1	PaInSg1	hn	á		h		
53	hnígan	s	1	0	a		1	Inf	hn	í		h		an
54	mígan	s	1	0	a		0	Inf	m	í		g		an
54	mígan	s	1	0	a		0	PaInSg1	m	á		g		
54	mígan	s	1	0	a		0	PaInPl	m	i		g		on
54	mígan	s	1	0	a		0	Pp	m	i		g		en
54	mígan	s	1	0	a		1	Inf	m	í		h		an
54	mígan	s	1	0	a		1	PaInSg1	m	á		h		
55	sígan	s	1	0	a		0	Inf	s	í		g		an
55	sígan	s	1	0	a		0	PaInSg1	s	á		g		
55	sígan	s	1	0	a		0	PaInPl	s	i		g		on
55	sígan	s	1	0	a		0	Pp	s	i		g		en
55	sígan	s	1	0	a		1	PaInSg1	s	á		h		
55	sígan	s	1	0	a		1	Inf	s	í		h		an
0	weorðan	s	3	0	e	500	0	Inf	w	eo		rð		an
0	weorðan	s	3	0	e	500	0	PaInSg1	w	ea		rð		
0	weorðan	s	3	0	e	500	0	PaInPl	w	u		rd		on
0	weorðan	s	3	0	e	500	0	Pp	w	o		rd		en
0	weorðan	s	3	0	e	500	1	Inf	w	ea		rð		an
73	nerian	w	1	A	a	525	0	Inf	n	e	r			ian
73	nerian	w	1	A	a	525	0	PsInSg2	n	e	r			est
73	nerian	w	1	A	a	525	0	PaInSg1	n	e	r		ed	e
73	nerian	w	1	A	a	525	1	PaInSg1	n	e	r		od	e
73	nerian	w	1	A	a	525	1	Inf	n	e	r	ig		an
73	nerian	w	1	A	a	525	2	Inf	n	e	r	g		an
73	nerian	w	1	A	a	525	3	Inf	n	e	r	ige		an
73	nerian	w	1	A	a	525	1	PsInSg2	n	e	r	ig		est
73	nerian	w	1	A	a	525	2	PsInSg2	n	e	r	g		est
73	nerian	w	1	A	a	525	3	PsInSg2	n	e	r	ige		est
73	nerian	w	1	A	a	525	4	Inf	n	e	r			an
73	nerian	w	1	A	a	525	1	PsInSg2	n	e	r			es
74	fremman	w	1	A	b	526	0	Inf	fr	e	mm			an
74	fremman	w	1	A	b	526	0	PsInSg2	fr	e	m			est
74	fremman	w	1	A	b	526	0	PaInSg1	fr	e	m		ed	e
74	fremman	w	1	A	b	526	0	PaInSg1	fr	e	m		od	e
74	fremman	w	1	A	b	526	1	Inf	fr	e	m			an
74	fremman	w	1	A	b	526	2	Inf	fr	e	m	i		an
75	settan	w	1	A	c	527	0	Inf	s	e	tt			an
75	settan	w	1	A	c	527	0	PsInSg2	s	e	t			st

75	settan	w	1	A	c	527	0	PaInSg1		s	e	t		t	e
75	settan	w	1	A	c	527	1	Pp	ge	s	e	t		ed	
75	settan	w	1	A	c	527	1	Pp	ge	s	e	t		od	
75	settan	w	1	A	c	527	2	Pp	ge	s	e	t			
76	déman	w	1	B	a	530	0	Inf		d	é	m			an
76	déman	w	1	B	a	530	0	PsInSg2		d	é	m			est
76	déman	w	1	B	a	530	0	PaInSg1		d	é	m		d	e
77	drencan	w	1	B	b	531	0	Inf		dr	e	nc			an
77	drencan	w	1	B	b	531	0	PsInSg2		dr	e	nc			est
77	drencan	w	1	B	b	531	0	PaInSg1		dr	e	nc		t	e
78	hyngrian	w	1	B	c	532	0	Inf		h	y	ngr			ian
78	hyngrian	w	1	B	c	532	0	PsInSg2		h	y	ngr			est
78	hyngrian	w	1	B	c	532	0	PaInSg1		h	y	ngr		ed	e
78	hyngrian	w	1	B	c	532	1	PaInSg1		h	y	ngr		od	e
78	hyngrian	w	1	B	c	532	1	Inf		h	y	ngr			an
79	gierwan	w	1	B	d	533	0	Inf		g	ie	r	w		an
79	gierwan	w	1	B	d	533	0	PsInSg2		g	ie	r			est
79	gierwan	w	1	B	d	533	0	PaInSg1		g	ie	r		ed	e
79	gierwan	w	1	B	d	533	1	PaInSg1		g	ie	r		od	e
79	gierwan	w	1	B	d	533	1	Inf		g	ie	r	wi		an
79	gierwan	w	1	B	d	533	2	Inf		g	ie	r	i		an
79	gierwan	w	1	B	d	533	3	Inf		g	ie	r			an
80	bycgan	w	1	B			0	Inf		b	y		cg		an
80	bycgan	w	1	B			0	PsInSg2		b	y		cg		est
80	bycgan	w	1	B			0	PaInSg1		b	o		h	t	e
93	wyrcan	w	1	B			0	Inf		w	y		rc		an
93	wyrcan	w	1	B			0	PsInSg2		w	y		rc		est
93	wyrcan	w	1	B			0	PaInSg1		w	o		rh	t	e
81	cweccan	w	1	B			0	Inf		cw	e		cc		an
81	cweccan	w	1	B			0	PsInSg2		cw	e		cc		est
81	cweccan	w	1	B			0	PaInSg1		cw	ea		h	t	e
81	cweccan	w	1	B			1	PaInSg1		cw	e		h	t	e
95	dreccan	w	1	B			0	Inf		dr	e		cc		an
95	dreccan	w	1	B			0	PsInSg2		dr	e		cc		est
95	dreccan	w	1	B			0	PaInSg1		dr	ea		h	t	e
95	dreccan	w	1	B			1	PaInSg1		dr	e		h	t	e
96	leccan	w	1	B			0	Inf		l	e		cc		an
96	leccan	w	1	B			0	PsInSg2		l	e		cc		est
96	leccan	w	1	B			0	PaInSg1		l	ea		h	t	e
96	leccan	w	1	B			1	PaInSg1		l	e		h	t	e
97	reccan	w	1	B			0	Inf		r	e		cc		an
97	reccan	w	1	B			0	PsInSg2		r	e		cc		est
97	reccan	w	1	B			0	PaInSg1		r	ea		h	t	e
97	reccan	w	1	B			1	PaInSg1		r	e		h	t	e
98	streccan	w	1	B			0	Inf		str	e		cc		an
98	streccan	w	1	B			0	PsInSg2		str	e		cc		est

98	streccan	w	1	B			0	PaInSg1		str	ea		h	t	e
98	streccan	w	1	B			1	PaInSg1		str	e		h	t	e
99	ðeccan	w	1	B			0	Inf		ð	e		cc		an
99	ðeccan	w	1	B			0	PsInSg2		ð	e		cc		est
99	ðeccan	w	1	B			0	PaInSg1		ð	ea		h	t	e
99	ðeccan	w	1	B			1	PaInSg1		ð	e		h	t	e
100	weccan	w	1	B			0	Inf		w	e		cc		an
100	weccan	w	1	B			0	PsInSg2		w	e		cc		est
100	weccan	w	1	B			0	PaInSg1		w	ea		h	t	e
100	weccan	w	1	B			1	PaInSg1		w	e		h	t	e
106	pǣcan	w	1	B			0	Inf		p	ǣ		c		an
106	pǣcan	w	1	B			0	PsInSg2		p	ǣ		c		est
106	pǣcan	w	1	B			0	PaInSg1		p	ǣ		h	t	e
106	pǣcan	w	1	B			1	PaInSg1		p	ǣ		c	t	e
107	wǣcan	w	1	B			0	Inf		w	ǣ		c		an
107	wǣcan	w	1	B			0	PsInSg2		w	ǣ		c		est
107	wǣcan	w	1	B			0	PaInSg1		w	ǣ		h	t	e
107	wǣcan	w	1	B			1	PaInSg1		w	ǣ		c	t	e
108	ícan	w	1	B			0	Inf			í		c		an
108	ícan	w	1	B			0	PsInSg2			í		c		est
108	ícan	w	1	B			0	PaInSg1			í		h	t	e
108	ícan	w	1	B			1	PaInSg1			í		c	t	e
108	ícan	w	1	B			2	Inf			é		c		an
108	ícan	w	1	B			2	PsInSg2			é		c		est
108	ícan	w	1	B			2	PaInSg1			é		h	t	e
108	ícan	w	1	B			3	PaInSg1			é		c	t	e
109	lǣcan	w	1	B			0	Inf		l	ǣ		c		an
109	lǣcan	w	1	B			0	PsInSg2		l	ǣ		c		est
109	lǣcan	w	1	B			0	PaInSg1		l	ǣ		h	t	e
109	lǣcan	w	1	B			1	PaInSg1		l	ǣ		c	t	e
110	leccan	w	1	B			0	Inf		l	e		cc		an
110	leccan	w	1	B			0	PsInSg2		l	e		cc		est
110	leccan	w	1	B			0	PaInSg1		l	e		h	t	e
110	leccan	w	1	B			1	PaInSg1		l	e		c	t	e
111	sícan	w	1	B			0	Inf		s	ý		c		an
111	sícan	w	1	B			0	PsInSg2		s	ý		c		est
111	sícan	w	1	B			0	PaInSg1		s	ý		h	t	e
111	sícan	w	1	B			1	PaInSg1		s	ý		c	t	e
112	ðryccan	w	1	B			0	Inf		ðr	y		cc		an
112	ðryccan	w	1	B			0	PsInSg2		ðr	y		cc		est
112	ðryccan	w	1	B			0	PaInSg1		ðr	y		h	t	e
112	ðryccan	w	1	B			1	PaInSg1		ðr	y		c	t	e
113	wleccan	w	1	B			0	Inf		wl	e		cc		an
113	wleccan	w	1	B			0	PsInSg2		wl	e		cc		est
113	wleccan	w	1	B			0	PaInSg1		wl	e		h	t	e
113	wleccan	w	1	B			1	PaInSg1		wl	e		cc	t	e

82	cwellan	w	1	B			0	Inf		cw	e		ll		an
82	cwellan	w	1	B			0	PsInSg2		cw	e		ll		est
82	cwellan	w	1	B			0	PaInSg1		cw	ea		l	d	e
82	cwellan	w	1	B			1	PaInSg1		cw	e		l	ed	e
82	cwellan	w	1	B			2	PaInSg1		cw	e		l	od	e
101	dwellan	w	1	B			0	Inf		dw	e		ll		an
101	dwellan	w	1	B			0	PsInSg2		dw	e		ll		est
101	dwellan	w	1	B			0	PaInSg1		dw	ea		l	d	e
101	dwellan	w	1	B			1	PaInSg1		dw	e		l	ed	e
101	dwellan	w	1	B			2	PaInSg1		dw	e		l	od	e
102	sellan	w	1	B			0	Inf		s	e		ll		an
102	sellan	w	1	B			0	PsInSg2		s	e		ll		est
102	sellan	w	1	B			0	PaInSg1		s	ea		l	d	e
102	sellan	w	1	B			1	PaInSg1		s	e		l	ed	e
102	sellan	w	1	B			2	PaInSg1		s	e		l	od	e
102	sellan	w	1	B			1	Inf		s	ie		ll		an
102	sellan	w	1	B			2	Inf		s	y		ll		an
102	sellan	w	1	B			1	PsInSg2		s	ie		ll		est
102	sellan	w	1	B			2	PsInSg2		s	y		ll		est
103	stellan	w	1	B			0	Inf		st	e		ll		an
103	stellan	w	1	B			0	PsInSg2		st	e		ll		est
103	stellan	w	1	B			0	PaInSg1		st	ea		l	d	e
103	stellan	w	1	B			1	PaInSg1		st	e		l	ed	e
103	stellan	w	1	B			2	PaInSg1		st	e		l	od	e
104	tellan	w	1	B			0	Inf		t	e		ll		an
104	tellan	w	1	B			0	PsInSg2		t	e		ll		est
104	tellan	w	1	B			0	PaInSg1		t	ea		l	d	e
104	tellan	w	1	B			1	PaInSg1		t	e		l	ed	e
104	tellan	w	1	B			2	PaInSg1		t	e		l	od	e
83	rǽcan	w	1	B			0	Inf		r	ǽ		c		an
83	rǽcan	w	1	B			0	PsInSg2		r	ǽ		c		est
83	rǽcan	w	1	B			0	PaInSg1		r	ǽ		h	t	e
83	rǽcan	w	1	B			1	PaInSg1		r	á		h	t	e
105	tǽcan	w	1	B			0	Inf		t	ǽ		c		an
105	tǽcan	w	1	B			0	PsInSg2		t	ǽ		c		est
105	tǽcan	w	1	B			0	PaInSg1		t	ǽ		h	t	e
105	tǽcan	w	1	B			1	PaInSg1		t	á		h	t	e
84	brengan	w	1	B			0	Inf		br	i		ng		an
84	brengan	w	1	B			0	PsInSg2		br	i		ng		est
84	brengan	w	1	B			0	PaInSg1		br	ó		h	t	e
84	brengan	w	1	B			1	Inf		br	e		ng		an
84	brengan	w	1	B			1	PsInSg2		br	e		ng		est
94	ðencan	w	1	B			0	Inf		ð	e		nc		an
94	ðencan	w	1	B			0	PsInSg2		ð	e		nc		est
94	ðencan	w	1	B			0	PaInSg1		ð	ó		h	t	e
85	ðyncan	w	1	B			0	Inf		ð	y		nc		an

169

85	ðyncan	w	1	B			0	PsInSg2		ð	y		nc		est
85	ðyncan	w	1	B			0	PaInSg1		ð	ú		h	t	e
86	sécan	w	1	B			0	Inf		s	é		c		an
86	sécan	w	1	B			0	PsInSg2		s	é		c		est
86	sécan	w	1	B			0	PaInSg1		s	ó		h	t	e
87	sealfian	w	2	0	a	536	0	Inf		s	ea	lf			ian
87	sealfian	w	2	0	a	536	0	PsInSg2		s	ea	lf			ast
87	sealfian	w	2	0	a	536	0	PsInSg3		s	ea	lf			að
87	sealfian	w	2	0	a	536	0	ImSg		s	ea	lf			a
87	sealfian	w	2	0	a	536	0	PaInSg1		s	ea	lf		od	e
87	sealfian	w	2	0	a	536	1	PaInSg1		s	ea	lf		ed	e
87	sealfian	w	2	0	a	536	1	PsInSg2		s	ea	lf			as
87	sealfian	w	2	0	a	536	1	Inf		s	ea	lf	ig		an
87	sealfian	w	2	0	a	536	2	Inf		s	ea	lf	ige		an
87	sealfian	w	2	0	a	536	3	Inf		s	ea	lf			an
87	sealfian	w	2	0	a	536	2	PsInSg2		s	ea	lf	ig		ast
87	sealfian	w	2	0	a	536	3	PsInSg2		s	ea	lf	ige		ast
87	sealfian	w	2	0	a	536	4	PsInSg2		s	ea	lf	ig		as
87	sealfian	w	2	0	a	536	5	PsInSg2		s	ea	lf	ige		as
87	sealfian	w	2	0	a	536	1	PsInSg3		s	ea	lf	ig		að
87	sealfian	w	2	0	a	536	2	PsInSg3		s	ea	lf	ige		að
87	sealfian	w	2	0	a	536	1	ImSg		s	ea	lf	ig		a
87	sealfian	w	2	0	a	536	2	ImSg		s	ea	lf	ige		a
87	sealfian	w	2	0	a	536	3	PsInSg3		s	ea	lf			iað
87	sealfian	w	2	0	a	536	3	ImSg		s	ea	lf			ia
88	twéogan	w	2		b	537	0	Inf		tw	éo		g		an
88	twéogan	w	2		b	537	0	PsInSg2		tw	éo				st
88	twéogan	w	2		b	537	0	PaInSg1		tw	éo			d	e
88	twéogan	w	2		b	537	1	Inf		tw	ío		g		an
88	twéogan	w	2		b	537	1	PsInSg2		tw	ío				st
88	twéogan	w	2		b	537	1	PaInSg1		tw	ío			d	e
89	habban	w	3		a		0	Inf		h	a		bb		an
89	habban	w	3		a		0	InflIn		h	a		bb		anne
89	habban	w	3		a		0	PsInSg1		h	æ		bb		e
89	habban	w	3		a		0	PsInSg2		h	a		f		ast
89	habban	w	3		a		0	PsInSg3		h	a		f		að
89	habban	w	3		a		0	PsInPl		h	a		bb		að
89	habban	w	3		a		0	PaInSg1		h	æ		f	d	e
89	habban	w	3		a		0	PaInSg2		h	æ		f	d	est
89	habban	w	3		a		0	PaInSg3		h	æ		f	d	e
89	habban	w	3		a		0	PaInPl		h	æ		f	d	on
89	habban	w	3		a		0	Pp	ge	h	æ		f	d	
89	habban	w	3		a		0	PsPa		h	æ		bb		ende
89	habban	w	3		a		0	PsSuSg		h	æ		bb		e
89	habban	w	3		a		0	PsSuPl		h	æ		bb		en
89	habban	w	3		a		0	PaSuSg		h	æ		f	d	e

89	habban	w	3		a	0	PaSuPl		h	æ		f	d	en
89	habban	w	3		a	0	ImpSg		h	a		f		a
89	habban	w	3		a	0	ImpPl		h	a		bb		að
89	habban	w	3		a	1	PsInSg2		h	a		f		as
89	habban	w	3		a	2	PsInSg2		h	æ		f		st
89	habban	w	3		a	1	PsInSg3		h	æ		f		ð
89	habban	w	3		a	1	PaInSg2		h	æ		f	d	es
90	libban	w	3		b	0	Inf		l	i		bb		an
90	libban	w	3		b	0	InflIn		l	i		bb		anne
90	libban	w	3		b	0	PsInSg1		l	i		bb		e
90	libban	w	3		b	0	PsInSg2		l	io		f		ast
90	libban	w	3		b	0	PsInSg3		l	io		f		að
90	libban	w	3		b	0	PsInPl		l	io		f		að
90	libban	w	3		b	0	PaInSg1		l	i		f	d	e
90	libban	w	3		b	0	PaInSg2		l	i		f	d	est
90	libban	w	3		b	0	PaInSg3		l	i		f	d	e
90	libban	w	3		b	0	PaInPl		l	i		f	d	on
90	libban	w	3		b	0	Pp	ge	l	i		f	d	
90	libban	w	3		b	0	PsPa		l	i		bb		ende
90	libban	w	3		b	1	PsInSg2		l	io		f		as
90	libban	w	3		b	1	PaInSg2		l	i		f	d	es
90	libban	w	3		b	0	PsSuSg		l	i		bb		e
90	libban	w	3		b	0	PsSuPl		l	i		bb		en
90	libban	w	3		b	0	PaSuSg		l	i		f	d	e
90	libban	w	3		b	0	PaSuPl		l	i		f	d	en
90	libban	w	3		b	0	ImpSg		l	io		f		a
90	libban	w	3		b	0	ImpPl		l	i		bb		að
91	secgan	w	3		c	0	Inf		s	e		cg		an
91	secgan	w	3		c	0	InflIn		s	e		cg		anne
91	secgan	w	3		c	0	PsInSg1		s	e		cg		e
91	secgan	w	3		c	0	PsInSg2		s	a		g		ast
91	secgan	w	3		c	0	PsInSg3		s	a		g		að
91	secgan	w	3		c	0	PsInPl		s	e		cg		að
91	secgan	w	3		c	0	PaInSg1		s	æ		g	d	e
91	secgan	w	3		c	0	PaInSg2		s	æ		g	d	est
91	secgan	w	3		c	0	PaInSg3		s	æ		g	d	e
91	secgan	w	3		c	0	PaInPl		s	æ		g	d	on
91	secgan	w	3		c	0	Pp	ge	s	æ		g	d	
91	secgan	w	3		c	0	PsPa		s	e		cg		ende
91	secgan	w	3		c	0	PsSuSg		s	e		cg		e
91	secgan	w	3		c	0	PsSuPl		s	e		cg		en
91	secgan	w	3		c	0	PaSuSg		s	æ		g	d	e
91	secgan	w	3		c	0	PaSuPl		s	æ		g	d	en
91	secgan	w	3		c	0	ImpSg		s	a		g		a
91	secgan	w	3		c	0	ImpPl		s	e		cg		að
91	secgan	w	3		c	1	Inf		s	e		cg		ean

91	secgan	w	3		c		1	PsInSg2		s	a		g		as
91	secgan	w	3		c		2	PsInSg2		s	æ		g		st
91	secgan	w	3		c		1	PsInSg3		s	æ		g		ð
91	secgan	w	3		c		1	PsInPl		s	e		cg		eað
91	secgan	w	3		c		1	PaInSg2		s	æ		g	d	es
91	secgan	w	3		c		1	ImpPl		s	e		cg		eað
91	secgan	w	3		c		1	ImpSg		s	æ		g		e
92	hycgan	w	3		d		0	Inf		h	y		cg		an
92	hycgan	w	3		d		0	InflIn		h	y		cg		anne
92	hycgan	w	3		d		0	PsInSg1		h	y		cg		e
92	hycgan	w	3		d		0	PsInSg2		h	y		g		est
92	hycgan	w	3		d		0	PsInSg3		h	y		g		eð
92	hycgan	w	3		d		0	PsInPl		h	y		cg		að
92	hycgan	w	3		d		0	PaInSg1		h	o		g	d	e
92	hycgan	w	3		d		0	PaInSg2		h	o		g	d	est
92	hycgan	w	3		d		0	PaInSg3		h	o		g	d	e
92	hycgan	w	3		d		0	PaInPl		h	o		g	d	on
92	hycgan	w	3		d		0	Pp	ge	h	o		g	od	
92	hycgan	w	3		d		1	Pp	ge	h	o		g	ed	
92	hycgan	w	3		d		0	PresPar		h	y		cg		ende
92	hycgan	w	3		d		0	PsSuSg		h	y		cg		e
92	hycgan	w	3		d		0	PsSuPl		h	y		cg		en
92	hycgan	w	3		d		0	PaSuSg		h	o		g	d	e
92	hycgan	w	3		d		0	PaSuPl		h	o		g	d	en
92	hycgan	w	3		d		0	ImpSg		h	o		g		a
92	hycgan	w	3		d		0	ImpPl		h	y		cg		að
92	hycgan	w	3		d		1	Inf		h	y		cg		ean
92	hycgan	w	3		d		1	InflIn		h	y		cg		eanne
92	hycgan	w	3		d		1	PsInPl		h	y		cg		eað
92	hycgan	w	3		d		1	PsInSg3		h	y		g		ð
92	hycgan	w	3		d		2	PsInSg3		h	o		g		að
92	hycgan	w	3		d		1	PsInSg2		h	y		g		st
92	hycgan	w	3		d		2	PsInSg2		h	o		g		ast
92	hycgan	w	3		d		3	PsInSg2		h	o		g		as
92	hycgan	w	3		d		1	PaInSg2		h	o		g	d	es
92	hycgan	w	3		d		1	ImpSg		h	y		g		e
92	hycgan	w	3		d		1	ImpPl		h	y		cg		eað
114	witan	pp	1			540	0	Inf		w	i		t		an
114	witan	pp	1			540	0	PsPa		w	i		t		ende
114	witan	pp	1			540	0	PsInSg1		w	á		t		
114	witan	pp	1			540	0	PsInSg2		w	á				st
114	witan	pp	1			540	0	PsInSg3		w	á		t		
114	witan	pp	1			540	0	PsInPl		w	i		t		on
114	witan	pp	1			540	0	ImpSg		w	i		t		e
114	witan	pp	1			540	0	ImpPl		w	i		t		að
114	witan	pp	1			540	0	PsSuSg		w	i		t		e

114	witan	pp	1			540	0	PsSuPl		w	i		t		en
114	witan	pp	1			540		PaInSg1		w	i		st		e
114	witan	pp	1			540		PaInPl		w	i		st		on
114	witan	pp	1			540	0	Pp		w	i		t		en
114	witan	pp	1			540	1	Inf		w	io		t		an
114	witan	pp	1			540	2	Inf		w	ie		t		an
114	witan	pp	1			540	1	PsInPl		w	io		t		un
114	witan	pp	1			540	2	PsInPl		w	ie		t		un
114	witan	pp	1			540	3	PsInPl		w	u		t		on
114	witan	pp	1			540	4	PsInPl		w	io		t		on
114	witan	pp	1			540	5	PsInPl		w	ie		t		on
114	witan	pp	1			540	1	ImpSg		w	i		t		
114	witan	pp	1			540	1	PaInSg1		w	i		ss		e
114	witan	pp	1			540	1	PaInPl		w	i		ss		on
115	dugan	pp	2			541	0	Inf		d	u		g		an
115	dugan	pp	2			541	0	PsPa		d	u		g		ende
115	dugan	pp	2			541	0	PsInSg1		d	éa		g		
115	dugan	pp	2			541	0	PsInSg2		d	éa		g		e
115	dugan	pp	2			541	0	PsInSg3		d	éa		g		
115	dugan	pp	2			541	0	PsInPl		d	u		g		on
115	dugan	pp	2			541	0	ImpSg		d	u		g		e
115	dugan	pp	2			541	0	ImpPl		d	u		g		að
115	dugan	pp	2			541	0	PsSuSg		d	y		g		e
115	dugan	pp	2			541	0	PsSuPl		d	y		g		en
115	dugan	pp	2			541	0	PaInSg1		d	o		ht		e
115	dugan	pp	2			541	0	PaInPl		d	o		ht		on
115	dugan	pp	2			541	0	Pp		d	o		ht		e
115	dugan	pp	2			541	1	PsInSg1		d	éa		h		
115	dugan	pp	2			541	2	PsInSg1		d	é		g		
115	dugan	pp	2			541	1	PsInSg3		d	éa		h		
115	dugan	pp	2			541	2	PsInSg3		d	é		g		
115	dugan	pp	2			541	1	PsSuSg		d	u		g		e
115	dugan	pp	2			541	1	PsPa		d	u		g		unde
116	unnan	pp	3		a	542	0	Inf			u		nn		an
116	unnan	pp	3		a	542	0	PsInSg1			a		nn		
116	unnan	pp	3		a	542	0	PsInSg2			a		nn		e
116	unnan	pp	3		a	542	0	PsInSg3			a		nn		
116	unnan	pp	3		a	542	0	PsInPl			u		nn		on
116	unnan	pp	3		a	542	0	PaInSg1			ú		ð		e
116	unnan	pp	3		a	542	0	PaInPl			ú		ð		on
116	unnan	pp	3		a	542	0	Pp			u		nn		en
116	unnan	pp	3		a	542	0	ImpSg			u		nn		e
116	unnan	pp	3		a	542	0	ImpPl			u		nn		að
116	unnan	pp	3		a	542	0	PsSuSg			u		nn		e
116	unnan	pp	3		a	542	0	PsSuPl			u		nn		en
116	unnan	pp	3		a	Camp	0	PsPa			u		nn		ende

116	unnan	pp	3		a	542	1	PsInSg1			a		n		
116	unnan	pp	3		a	542	2	PsInSg1			o		nn		
116	unnan	pp	3		a	542	3	PsInSg1			o		n		
116	unnan	pp	3		a	542	1	PsInSg3			a		n		
116	unnan	pp	3		a	542	2	PsInSg3			o		nn		
116	unnan	pp	3		a	542	3	PsInSg3			o		n		
116	unnan	pp	3		a	542	1	ImpSg	ge		u		nn		
116	unnan	pp	3		a	542	2	ImpSg	g		io		nn		
116	unnan	pp	3		a	542	1	PsSuSg	ge		u		nn		
116	unnan	pp	3		a	542	2	PsSuSg	g		io		nn		
117	cunnan	pp	3		b	542	0	Inf		c	u		nn		an
117	cunnan	pp	3		b	542	0	PsInSg1		c	a		nn		
117	cunnan	pp	3		b	542	0	PsInSg2		c	a		n		st
117	cunnan	pp	3		b	542	0	PsInSg3		c	a		nn		
117	cunnan	pp	3		b	542	0	PsInPl		c	u		nn		on
117	cunnan	pp	3		b	542	0	ImpSg		c	u		nn		e
117	cunnan	pp	3		b	542	0	ImpPl		c	u		nn		að
117	cunnan	pp	3		b	542	0	PsPa		c	u		nn		ende
117	cunnan	pp	3		b	542	0	PaInSg1		c	ú		ð		e
117	cunnan	pp	3		b	542	0	PaInPl		c	ú		ð		on
117	cunnan	pp	3		b	542	0	Pp		c	u		nn		en
117	cunnan	pp	3		b	Camp	0	PsSuSg		c	u		nn		e
117	cunnan	pp	3		b	Camp	0	PsSuPl		c	u		nn		en
117	cunnan	pp	3		b	542	1	PsInSg1		c	a		n		
117	cunnan	pp	3		b	542	2	PsInSg1		c	o		nn		
117	cunnan	pp	3		b	542	3	PsInSg1		c	o		nn		
117	cunnan	pp	3		b	542	1	PsInSg2		c	o		n		st
117	cunnan	pp	3		b	542	1	PsInSg3		c	a		n		
117	cunnan	pp	3		b	542	2	PsInSg3		c	o		nn		
117	cunnan	pp	3		b	542	3	PsInSg3		c	o		nn		
117	cunnan	pp	3		b	Camp	1	Pp	on	c	u		nn		en
117	cunnan	pp	3		b	Camp	1	PsInPl		c	u		nn		að
117	cunnan	pp	3		b	Camp	2	PsInPl		c	u		nn		eð
118	ðurfan	pp	3		c	542	0	Inf		ð	u		rf		an
118	ðurfan	pp	3		c	542	0	PsInSg1		ð	ea		rf		
118	ðurfan	pp	3		c	542	0	PsInSg2		ð	ea		rft		t
118	ðurfan	pp	3		c	Camp	0	PsInSg3		ð	ea		rf		
118	ðurfan	pp	3		c	542	0	PsInPl		ð	u		rf		on
118	ðurfan	pp	3		c	542	0	PaInSg1		ð	o		rft		e
118	ðurfan	pp	3		c	542	0	PaInPl		ð	o		rft		on
118	ðurfan	pp	3		c	542	0	Pp		ð	u		rf		en
118	ðurfan	pp	3		c	542	0	ImpSg		ð	u		rf		e
118	ðurfan	pp	3		c	542	0	ImpPl		ð	u		rf		að
118	ðurfan	pp	3		c	542	0	PsPa		ð	ea		rf		ende
118	ðurfan	pp	3		c	Camp	0	PsSuSg		ð	u		rf		e
118	ðurfan	pp	3		c	Camp	0	PsSuPl		ð	u		rf		en

118	ðurfan	pp	3		c	Camp	1	PsSuSg		ð	y		rf		e
118	ðurfan	pp	3		c	Camp	1	PsPa		ð	o		rf		ende
118	ðurfan	pp	3		c	Camp	2	PsPa		ð	u		rf		ende
118	ðurfan	pp	3		c	Camp	1	PsInSg3		ð	o		rf		eð
118	ðurfan	pp	3		c	Camp	1	PsInPl		ð	o		rf		eð
118	ðurfan	pp	3		c	Camp	2	PsInPl		ð	u		rf		u
119	durran	pp	3		d	542	0	Inf		d	u		rr		an
119	durran	pp	3		d	542	0	PsInSg1		d	ea		r		
119	durran	pp	3		d	542	0	PsInSg2		d	ea		r		st
119	durran	pp	3		d	Camp	0	PsInSg3		d	ea		rr		
119	durran	pp	3		d	542	0	PsInPl		d	u		rr		on
119	durran	pp	3		d	542	0	PaInSg1		d	o		rst		e
119	durran	pp	3		d	542	0	PaInPl		d	o		rst		on
119	durran	pp	3		d	542	0	Pp		d	ea		r		en
119	durran	pp	3		d	542	0	ImpSg		d	ea		r		e
119	durran	pp	3		d	542	0	ImpPl		d	ea		r		að
119	durran	pp	3		d	542	0	PsPa		d	ea		r		ende
119	durran	pp	3		d	Camp	0	PsSuSg		d	u		rr		e
119	durran	pp	3		d	Camp	0	PsSuPl		d	u		rr		en
119	durran	pp	3		d	542	1	Inf		d	ea		rr		an
119	durran	pp	3		d	542	2	Inf		d	ea		rr		an
119	durran	pp	3		d	542	1	PsInSg1		d	ea		rr		
119	durran	pp	3		d	Camp	1	PsSuSg		d	y		rr		e
119	durran	pp	3		d	Camp	1	PaInSg1		d	u		rst		e
119	durran	pp	3		d	Camp	2	PaInSg1		d	y		rst		e
119	durran	pp	3		d	Camp	3	PaInSg1		d	y		st		e
119	durran	pp	3		d	Camp	4	PaInSg1		d	a		rst		e
119	durran	pp	3		d	Camp	1	PaInPl		d	u		rst		on
119	durran	pp	3		d	Camp	2	PaInPl		ð	y		rst		on
119	durran	pp	3		d	Camp	3	PaInPl		d	y		st		on
119	durran	pp	3		d	Camp	4	PaInPl		d	a		rst		on
119	durran	pp	3		d	Camp	2	PsInSg1		d	a		rr		
120	sculan	pp	4		a	543	0	Inf		sc	u		l		an
120	sculan	pp	4		a	543	0	PsInSg1		sc	ea		l		
120	sculan	pp	4		a	543	0	PsInSg2		sc	ea		l		t
120	sculan	pp	4		a	Camp	0	PsInSg3		sc	ea		l		
120	sculan	pp	4		a	543	0	PsInPl		sc	u		l		on
120	sculan	pp	4		a	543	0	PaInSg1		sc	eo		ld		e
120	sculan	pp	4		a	543	0	PaInPl		sc	eo		ld		on
120	sculan	pp	4		a	543	1	PaInSg1		sc	o		ld		e
120	sculan	pp	4		a	543	1	PaInPl		sc	o		ld		on
120	sculan	pp	4		a	543	2	PaInSg1		sc	ea		ld		e
120	sculan	pp	4		a	543	2	PaInPl		sc	ea		ld		on
120	sculan	pp	4		a	543	2	PaInSg1		sc	a		ld		e
120	sculan	pp	4		a	543	2	PaInPl		sc	a		ld		on
120	sculan	pp	4		a	543	2	PaInSg1		sc	u		ld		e

120	sculan	pp	4		a	543	2	PaInPl		sc	u		ld		on
120	sculan	pp	4		a	543	0	Pp		sc	u		l		en
120	sculan	pp	4		a	543	0	ImpSg		sc	u		l		e
120	sculan	pp	4		a	543	0	ImpPl		sc	u		l		að
120	sculan	pp	4		a	543	0	PsPa		sc	u		l		ende
120	sculan	pp	4		a	Camp	0	PsSuSg		sc	u		l		e
120	sculan	pp	4		a	Camp	0	PsSuPl		sc	u		l		en
120	sculan	pp	4		a	Camp	1	PsSuSg		sc	y		ld		e
120	sculan	pp	4		a	543	1	Inf		sc	eo		l		an
120	sculan	pp	4		a	543	1	PsInPl		sc	eo		l		on
120	sculan	pp	4		a	Camp	2	PsSuSg		sc	y		l		e
120	sculan	pp	4		a	Camp	1	PsInSg2		sc	a		l		t
120	sculan	pp	4		a	Camp	1	PsInSg3		sc	a		l		
120	sculan	pp	4		a	Camp	1	PsInSg1		sc	a		l		
120	sculan	pp	4		a	Camp	2	PsInPl		sc	y		l		un
120	sculan	pp	4		a	Camp	3	PsInPl		sc	i		l		un
120	sculan	pp	4		a	Camp	4	PsInPl		sc	iu		l		un
120	sculan	pp	4		a	Camp	5	PsInPl		sc	io		l		un
121	munan	pp	4		b	543	0	Inf		m	u		n		an
121	munan	pp	4		b	543	0	PsInSg1		m	a		n		
121	munan	pp	4		b	543	0	PsInSg2		m	a		n		st
121	munan	pp	4		b	Camp	0	PsInSg3	ge	m	a		n		
121	munan	pp	4		b	543	0	PsInPl		m	u		n		on
121	munan	pp	4		b	Camp	0	PaInSg1	ge	m	u		nd		e
121	munan	pp	4		b	Camp	0	PaInPl	ge	m	u		nd		on
121	munan	pp	4		b	543	0	Pp	ge	m	u		n		en
121	munan	pp	4		b	543	0	ImpSg		m	u		n		
121	munan	pp	4		b	Camp	0	ImpPl	ge	m	u		n		að
121	munan	pp	4		b	543	0	PsPa		m	u		n		ende
121	munan	pp	4		b	Camp	0	PsSuSg		m	u		n		e
121	munan	pp	4		b	Camp	0	PsSuPl		m	u		n		en
121	munan	pp	4		b	543	1	Inf		m	o		n		
121	munan	pp	4		b	543	1	PsInSg1		m	o		n		st
121	munan	pp	4		b	543	1	ImpSg		m	y		n		e
121	munan	pp	4		b	543	2	ImpSg		m	u		n		e
121	munan	pp	4		b	Camp	3	ImpSg	ge	m	u		n		
121	munan	pp	4		b	Camp	2	Inf		m	u		n		an
121	munan	pp	4		b	Camp	1	PsInSg2	ge	m	a		n		st
121	munan	pp	4		b	Camp	1	PsInPl	ge	m	u		n		on
121	munan	pp	4		b	Camp	1	PsSuSg		m	y		n		e
121	munan	pp	4		b	Camp	1	PsPa	ge	m	u		n		ende
121	munan	pp	4		b	Camp	2	PsInSg1	ge	m	u		n		u
121	munan	pp	4		b	Camp	2	PsInSg2	ge	m	y		n		es
121	munan	pp	4		b	Camp	3	PsInSg2	ge	m	y		n		est
121	munan	pp	4		b	Camp	1	PsSuPl	ge	m	y		n		en
121	munan	pp	4		b	Camp	4	ImpSg	ge	m	y		n		e

122	magan	pp	5			544	0	Inf		m	a		g	an
122	magan	pp	5			544	0	PsInSg1		m	æ		g	
122	magan	pp	5			544	0	PsInSg2		m	ea		h	t
122	magan	pp	5			Camp	0	PsInSg3		m	æ		g	
122	magan	pp	5			544	0	PsInPl		m	a		g	on
122	magan	pp	5			Camp	0	PaInSg1		m	ea		ht	e
122	magan	pp	5			544	0	PaInPl		m	ea		ht	on
122	magan	pp	5			544	0	Pp		m	a		g	en
122	magan	pp	5			544	0	ImpSg		m	a		g	e
122	magan	pp	5			544	0	ImpPl		m	a		g	að
122	magan	pp	5			544	0	PsSuSg		m	æ		g	e
122	magan	pp	5			544	0	PsSuPl		m	æ		g	en
122	magan	pp	5			544	0	PsPa		m	a		g	ende
122	magan	pp	5			Camp	1	PaInSg1		m	e		ht	e
122	magan	pp	5			544	1	PaInPl		m	e		ht	on
122	magan	pp	5			Camp	2	PaInSg1		m	i		ht	e
122	magan	pp	5			544	2	PaInPl		m	i		ht	on
122	magan	pp	5			Camp	3	PaInSg1		m	æ		ht	e
122	magan	pp	5			Camp	3	PaInPl		m	æ		ht	on
122	magan	pp	5			544	1	PsInSg2		m	i		h	t
122	magan	pp	5			Camp	1	PsInPl		m	æ		g	on
122	magan	pp	5			Camp	1	PsSuSg		m	a		g	e
122	magan	pp	5			Camp	2	PsSuSg		m	u		g	e
122	magan	pp	5			Camp	1	PsInSg3		m	e		g	
122	magan	pp	5			Camp	2	PsInSg2		m	æ		h	t
122	magan	pp	5			Camp	3	PsInSg2		m	e		h	t
122	magan	pp	5			Camp	3	PsSuSg		m	ei		g	e
122	magan	pp	5			Camp	1	PsPa		m	e		g	ende
123	nugan	pp	5			544	0	Inf		n	u		g	an
123	nugan	pp	5			544	0	PsInSg1		n	ea		h	
123	nugan	pp	5			544	0	PsInSg2		n	ea		h	t
123	nugan	pp	5			544	0	PsInSg3	ge	n	ea		h	
123	nugan	pp	5			544	0	PsInPl	ge	n	u		g	on
123	nugan	pp	5			544	0	PaInSg1		n	o		ht	e
123	nugan	pp	5			544	0	PaInPl		n	o		ht	on
123	nugan	pp	5			544	0	Pp	ge	n	u		g	en
123	nugan	pp	5			544	0	ImpSg		n	u		g	e
123	nugan	pp	5			544	0	ImpPl		n	u		g	að
123	nugan	pp	5			544	0	PsSuSg	ge	n	u		g	e
123	nugan	pp	5			544	0	PsSuPl		n	u		g	en
123	nugan	pp	5			544	0	PsPa		n	u		g	ende
123	nugan	pp	5			544	1	PsInSg3	be	n	ea		h	
123	nugan	pp	5			544	1	PsSuSg	be	n	u		g	e
123	nugan	pp	5			544	1	PsInPl	be	n	u		g	on
124	mótan	pp	6		a	545	0	Inf		m	ó		t	an
124	mótan	pp	6		a	545	0	PsInSg1		m	ó		t	

124	mótan	pp	6		a	545	0	PsInSg2	m	ó			st
124	mótan	pp	6		a	545	0	PsInSg3	m	ó		t	
124	mótan	pp	6		a	545	0	PsInPl	m	ó		t	on
124	mótan	pp	6		a	545	0	PaInSg1	m	ó		st	on
124	mótan	pp	6		a	545	0	PaInPl	m	ó		st	on
124	mótan	pp	6		a	545	0	Pp	m	ó		t	en
124	mótan	pp	6		a	545	0	ImpSg	m	ó		t	e
124	mótan	pp	6		a	545	0	ImpPl	m	ó		t	að
124	mótan	pp	6		a	545	0	PsSuSg	m	ó		t	e
124	mótan	pp	6		a	545	0	PsSuPl	m	ó		t	en
124	mótan	pp	6		a	545	0	PsPa	m	ó		t	ende
125	ágan	pp	6		b	546	0	Inf		á		g	an
125	ágan	pp	6		b	546	0	PsInSg1		á		g	
125	ágan	pp	6		b	546	0	PsInSg2		á		h	st
125	ágan	pp	6		b	546	0	PsInSg3		á		g	
125	ágan	pp	6		b	546	0	PsInPl		á		g	on
125	ágan	pp	6		b	546	0	PaInSg1		á		ht	e
125	ágan	pp	6		b	546	0	PaInPl		á		ht	on
125	ágan	pp	6		b	546	0	Pp		á		g	en
125	ágan	pp	6		b	546	0	ImpSg		á		g	e
125	ágan	pp	6		b	546	0	ImpPl		á		g	að
125	ágan	pp	6		b	546	0	PsSuSg		á		g	e
125	ágan	pp	6		b	546	0	PsSuPl		á		g	en
125	ágan	pp	6		b	546	0	PsPa		á		g	ende
125	ágan	pp	6		b	546	1	PsInSg1		á		h	
125	ágan	pp	6		b	546	1	PsInSg3		á		g	
125	ágan	pp	6		b	546	1	Pp		ǽ		g	en
125	ágan	pp	6		b	546	1	PsInSg2		á		h	t
126	béon	a				548	0	Inf	b	é			on
126	béon	a				548	0	InflInf	w	o		s	anne
126	béon	a				548	0	PsInSg1		eo		m	
126	béon	a				548	0	PsInSg2		ea		r	t
126	béon	a				548	0	PsInSg3		i		s	
126	béon	a				548	0	PsInPl	s	i		n	t
126	béon	a				548	0	PaInSg1	w	æ		s	
126	béon	a				548	0	PaInSg2	w	ǽ		r	on
126	béon	a				548	0	PaInSg3	w	æ		s	
126	béon	a				548	0	PaInPl	w	ǽ		r	on
126	béon	a				548	0	Pp	b	é			on
126	béon	a				548	0	ImpSg	b	ío			
126	béon	a				548	0	ImpPl	b	ío			ð
126	béon	a				548	0	PsSuSg	s	í			e
126	béon	a				548	0	PsSuPl	s	í			en
126	béon	a				548	0	PaSuSg	w	ǽ		r	e
126	béon	a				548	0	PaSuPl	w	ǽ		r	en
126	béon	a				548	0	PsPa	b	ío			nde

126	béon	a				548	1	Inf		b	í				on
126	béon	a				548	2	Inf		w	e		s		an
126	béon	a				548	1	InflInf		w	o		ss		anne
126	béon	a				548	2	InflInf		w	o		s		ane
126	béon	a				548	1	PsInSg1			ea		m		
126	béon	a				548	2	PsInSg1			a		m		
126	béon	a				548	3	PsInSg1		b	ío				
126	béon	a				548	4	PsInSg1		b	éo				
126	béon	a				548	5	PsInSg1		b	ío		m		
126	béon	a				548	6	PsInSg1		n	ea		m		
126	béon	a				548	1	PsInSg2			ea		r		ð
126	béon	a				548	2	PsInSg2			a		r		ð
126	béon	a				548	3	PsInSg2		b	i				st
126	béon	a				548	4	PsInSg2		b	i				s
126	béon	a				548	1	PsInSg3		b	i				ð
126	béon	a				548	2	PsInSg3			i		s		
126	béon	a				548	3	PsInSg3		n	i		s		
126	béon	a				548	1	PsInPl		s	i		nd		on
126	béon	a				548	2	PsInPl		s	i		nd		un
126	béon	a				548	3	PsInPl		s	i		nd		
126	béon	a				548	4	PsInPl			ea		r		on
126	béon	a				548	5	PsInPl			a		r		on
126	béon	a				548	6	PsInPl			ea		r		un
126	béon	a				548	7	PsInPl			a		r		un
126	béon	a				548	8	PsInPl		b	ío				ð
126	béon	a				548	9	PsInPl		b	éo				ð
126	béon	a				548	10	PsInPl		b	io		ð		on
126	béon	a				548	11	PsInPl		b	i		ð		on
126	béon	a				548	12	PsInPl		b	io		ð		un
126	béon	a				548	13	PsInPl		b	i		ð		un
126	béon	a				548	14	PsInPl		s	ie		nt		
126	béon	a				548	15	PsInPl		s	ie		nd		on
126	béon	a				548	1	PsSuSg		s	í				
126	béon	a				548	2	PsSuSg		b	ío				
126	béon	a				548	3	PsSuSg		b	éo				
126	béon	a				548	1	PsSuPl		s	í				n
126	béon	a				548	2	PsSuPl		b	ío				n
126	béon	a				548	3	PsSuPl		b	éo				n
126	béon	a				548	1	ImpSg		b	éo				
126	béon	a				548	2	ImpSg		w	e		s		
126	béon	a				548	1	ImpPl		b	éo				ð
126	béon	a				548	2	ImpPl		w	e		s		að
126	béon	a				548	1	PsPa		b	éo				nde
126	béon	a				548	2	PsPa		w	e		s		ende
127	dón	a				549	0	Inf		d	ó				n
127	dón	a				549	0	InflInf		d	ó				anne

127	dón	a				549	0	PsInSg1		d	ó			
127	dón	a				549	0	PsInSg2		d	é			st
127	dón	a				549	0	PsInSg3		d	é			ð
127	dón	a				549	0	PsInPl		d	ó			ð
127	dón	a				549	0	PaInSg1		d	y	d		e
127	dón	a				549	0	PaInSg2		d	y	d		es
127	dón	a				549	0	PaInSg3		d	y	d		e
127	dón	a				549	0	PaInPl		d	y	d		on
127	dón	a				549	0	Pp	ge	d	ó			n
127	dón	a				549	0	ImpSg		d	ó			
127	dón	a				549	0	ImpPl		d	ó			ð
127	dón	a				549	0	PsSuSg		d	ó			
127	dón	a				549	0	PsSuPl		d	ó			n
127	dón	a				549	0	PaSuSg		d	y	d		e
127	dón	a				549	0	PaSuPl		d	y	d		en
127	dón	a				549	0	PsPa		d	ó			nde
127	dón	a				549	1	PaInSg2		d	y	d		est
127	dón	a				549	1	PsInSg1		d	ó	m		
127	dón	a				549	1	PsInSg2		d	œ			s
127	dón	a				549	2	PsInSg2		d	œ			st
127	dón	a				549	1	PsInSg3		d	œ			ð
127	dón	a				549	2	PsInSg3		d	œ			s
127	dón	a				549	1	ImpSg		d	ó			a
127	dón	a				549	1	ImpPl		d	ó			að
127	dón	a				549	1	Inf		d	ó			a
127	dón	a				549	2	Inf		d	ó			an
127	dón	a				549	1	PaInPl		d	ǽ	d		on
127	dón	a				549	1	InflInf		d	ó			anne
127	dón	a				549	2	InflInf		d	ó			ne
127	dón	a				549	3	InflInf		d	ó			enne
127	dón	a				449	1	PsSug		d	ǽ	d		e
127	dón	a				549	1	Pp	ge	d	é			n
127	dón	a				549	2	Pp	ge	d	œ			n
127	dón	a				Camp	1	PsSuSg		d	ó			e
127	dón	a				Camp	1	PsPa		d	ó			ende
127	dón	a				Camp	1	PaInSg1		d	e	d		e
128	gán	a				550	0	Inf		g	á			n
128	gán	a				550	0	InflInf		g	á			nne
128	gán	a				550	0	PsInSg1		g	á			
128	gán	a				550	0	PsInSg2		g	ǽ			st
128	gán	a				550	0	PsInSg3		g	ǽ			ð
128	gán	a				550	0	PsInPl		g	á			ð
128	gán	a				550	0	PaInSg1			éo	d		e
128	gán	a				550	0	PaInSg2			éo	d		es
128	gán	a				550	0	PaInSg3			éo	d		e
128	gán	a				550	0	PaInPl			éo	d		on

128	gán	a				550	0	Pp	ge	g	á				n
128	gán	a				550	0	ImpSg		g	á				
128	gán	a				550	0	ImpPl		g	á				ð
128	gán	a				550	0	PsSuSg		g	á				
128	gán	a				550	0	PsSuPl		g	á				n
128	gán	a				550	0	PaSuSg			éo		d		e
128	gán	a				550	0	PaSuPl			éo		d		en
128	gán	a				Camp	0	PsPa	ge	g	á		nd		e
128	gán	a				550	1	PaInSg2			éo		d		est
128	gán	a				Camp	1	PaInSg1			éa		d		e
128	gán	a				Camp	1	PaInSg3			éa		d		e
128	gán	a				Camp	1	Pp	ge		éa				d
128	gán	a				Camp	1	PsInSg1	ge	g	aa				
128	gán	a				Camp	2	PsInSg1	ge	g	á				n
128	gán	a				Camp	1	PsInSg3		g	á				ð
128	gán	a				Camp	2	PsInSg3		g	é				ð
128	gán	a				Camp	1	ImpSg	ge	g	aa				
128	gán	a				Camp	2	ImpSg		g	á				e
128	gán	a				Camp	1	ImpPl		g	ǽ				ð
128	gán	a				Camp	1	PsSuSg	ge	g	á				e
128	gán	a				Camp	1	PsSuPl		g	ǽ				n
128	gán	a				Camp	1	Inf		g	ǽ				
128	gán	a				Camp	2	Inf	ge	g	á				
128	gán	a				550	1	InflInf		g	á				ne
128	gán	a				550	2	InflInf		g	á				nde
129	willan	a				551	0	Inf		w	i		ll		an
129	willan	a				551	0	PsInSg1		w	i		ll		e
129	willan	a				551	0	PsInSg2		w	i		l		t
129	willan	a				551	0	PsInSg3		w	i		l		e
129	willan	a				551	0	PsInPl		w	i		ll		að
129	willan	a				551	0	PaInSg1		w	o		ld		e
129	willan	a				551	0	PaInSg2		w	o		ld		es
129	willan	a				551	0	PaInSg3		w	o		ld		e
129	willan	a				551	0	PaInPl		w	o		ld		on
129	willan	a				551	0	Pp	ge	w	o		ld		en
129	willan	a				551	0	ImpSg		w	e		ll		e
129	willan	a				Camp	0	ImpPl		w	e		ll		að
129	willan	a				551	0	PsSuSg		w	i		ll		e
129	willan	a				551	0	PsSuPl		w	i		ll		en
129	willan	a				551	0	PaSuSg		w	o		ld		e
129	willan	a				551	0	PaSuPl		w	o		ld		en
129	willan	a				551	0	PsPa		w	i		ll		ende
129	willan	a				551	1	PsInSg2		w	i		ll		e
129	willan	a				551	1	PsSuSg		w	i		l		e
129	willan	a				551	1	PaInSg2		w	o		ld		est
129	willan	a				Camp	1	PsInSg3		w	i		ll		e

129	willan	a				Camp	1	PsInSg1	w	i		ll		a
129	willan	a				Camp	2	PsInSg1	w	i		ll		io
129	willan	a				Camp	1	PsPa	w	e		ll		ende
129	willan	a				Camp	1	PaInSg1	w	a		ld		e
129	willan	a				Camp	3	PsInSg1	w	i		ll		o
129	willan	a				Camp	2	PsInSg3	w	i		l		
129	willan	a				Camp	3	PsInSg3	w	i		ll		
129	willan	a				Camp	1	PsInPl	w	a		ll		að
129	willan	a				Camp	2	PsInPl	w	a		ll		as
129	willan	a				Camp	3	PsInPl	w	a		ll		on
129	willan	a				Camp	2	PsSuSg	w	æ		ll		e
129	willan	a				Camp	3	PsSuSg	w	e		ll		e
130	nytan	pp	1			540	0	Inf	n	y		t		an
130	nytan	pp	1			540	0	PsPa	n	y		t		ende
130	nytan	pp	1			540	0	PsInSg1	n	á		t		
130	nytan	pp	1			540	0	PsInSg2	n	á				st
130	nytan	pp	1			540	0	PsInSg3	n	á		t		
130	nytan	pp	1			540	0	PsInPl	n	y		t		on
130	nytan	pp	1			540	0	ImpSg	n	y		t		e
130	nytan	pp	1			540	0	ImpPl	n	y		t		að
130	nytan	pp	1			540	0	PsSuSg	n	y		t		e
130	nytan	pp	1			540	0	PsSuPl	n	y		t		en
130	nytan	pp	1			540		PaInSg1	n	y		st		e
130	nytan	pp	1			540		PaInPl	n	y		st		on
130	nytan	pp	1			540	0	Pp	n	y		t		en
130	nytan	pp	1			540	1	Inf	n	io		t		an
130	nytan	pp	1			540	2	Inf	n	ie		t		an
130	nytan	pp	1			540	1	PsInPl	n	io		t		un
130	nytan	pp	1			540	2	PsInPl	n	ie		t		un
130	nytan	pp	1			540	3	PsInPl	n	u		t		on
130	nytan	pp	1			540	4	PsInPl	n	io		t		on
130	nytan	pp	1			540	5	PsInPl	n	ie		t		on
130	nytan	pp	1			540	1	ImpSg	n	y		t		
130	nytan	pp	1			540	1	PaInSg1	n	y		ss		e
130	nytan	pp	1			540	1	PaInPl	n	y		ss		on
131	nágan	pp	6		b	546	0	Inf	n	á		g		an
131	nágan	pp	6		b	546	0	PsInSg1	n	á		g		
131	nágan	pp	6		b	546	0	PsInSg2	n	á		h		st
131	nágan	pp	6		b	546	0	PsInSg3	n	á		g		
131	nágan	pp	6		b	546	0	PsInPl	n	á		g		on
131	nágan	pp	6		b	546	0	PaInSg1	n	á		ht		e
131	nágan	pp	6		b	546	0	PaInPl	n	á		ht		on
131	nágan	pp	6		b	546	0	Pp	n	á		g		en
131	nágan	pp	6		b	546	0	ImpSg	n	á		g		e
131	nágan	pp	6		b	546	0	ImpPl	n	á		g		að
131	nágan	pp	6		b	546	0	PsSuSg	n	á		g		e

131	nágan	pp	6		b	546	0	PsSuPl		n	á		g		en
131	nágan	pp	6		b	546	0	PsPa		n	á		g		ende
131	nágan	pp	6		b	546	1	PsInSg1		n	á		h		
131	nágan	pp	6		b	546	1	PsInSg3		n	á		g		
131	nágan	pp	6		b	546	1	Pp		n	ǽ		g		en
131	nágan	pp	6		b	546	1	PsInSg2		n	á		h		t
132	nyllan	a				551	0	Inf		n	y		ll		an
132	nyllan	a				551	0	PsInSg1		n	y		ll		e
132	nyllan	a				551	0	PsInSg2		n	y		l		t
132	nyllan	a				551	0	PsInSg3		n	y		l		e
132	nyllan	a				551	0	PsInPl		n	y		ll		að
132	nyllan	a				551	0	PaInSg1		n	o		ld		e
132	nyllan	a				551	0	PaInSg2		n	o		ld		es
132	nyllan	a				551	0	PaInSg3		n	o		ld		e
132	nyllan	a				551	0	PaInPl		n	o		ld		on
132	nyllan	a				551	0	Pp	ge	n	o		ld		en
132	nyllan	a				551	0	ImpSg		n	e		ll		e
132	nyllan	a				Camp	0	ImpPl		n	e		ll		að
132	nyllan	a				551	0	PsSuSg		n	y		ll		e
132	nyllan	a				551	0	PsSuPl		n	y		ll		en
132	nyllan	a				551	0	PaSuSg		n	o		ld		e
132	nyllan	a				551	0	PaSuPl		n	o		ld		en
132	nyllan	a				551	0	PsPa		n	y		ll		ende
132	nyllan	a				551	1	PsInSg2		n	y		ll		e
132	nyllan	a				551	1	PsSuSg		n	y		l		e
132	nyllan	a				551	1	PaInSg2		n	o		ld		est
132	nyllan	a				Camp	1	PsInSg3		n	y		ll		e
132	nyllan	a				Camp	1	PsInSg1		n	y		ll		a
132	nyllan	a				Camp	2	PsInSg1		n	y		ll		io
132	nyllan	a				Camp	1	PsPa		n	e		ll		ende
132	nyllan	a				Camp	1	PaInSg1		n	a		ld		e
132	nyllan	a				Camp	3	PsInSg1		n	y		ll		o
132	nyllan	a				Camp	2	PsInSg3		n	y		l		
132	nyllan	a				Camp	3	PsInSg3		n	y		ll		
132	nyllan	a				Camp	1	PsInPl		n	a		ll		að
132	nyllan	a				Camp	2	PsInPl		n	a		ll		as
132	nyllan	a				Camp	3	PsInPl		n	a		ll		on
132	nyllan	a				Camp	2	PsSuSg		n	æ		ll		e
132	nyllan	a				Camp	3	PsSuSg		n	e		ll		e

12.3.   Screenshots of the Analyser Output and Interface

æ   þ

hū mæg māre scamu þurh Godes yrre

Analyze form: [   ] [Analyze]

☐ (force variant spellings) ☐ (include derived forms) ☐ (show sql)

○ **COM**
  ○ *cuman*
    ○ verb
      ○ infl. like cuman - conjugate
        ○ Sg. 1.p. Ind. Pret. (s, 4, b, cóm - ID:15917, freq: 2599, in texts: 378, prob: 100)
        ○ Sg. 3.p. Ind. Pret. (s, 4, b, cóm - ID:15918, freq: 2599, in texts: 378, prob: 100)

○ **NIHT**
  ○ *neaht*
    ○ noun
      ○ infl. like bóc - decline
        ○ Sg. F. Nom. (niht - ID:517, freq: 1106, in texts: 282, prob: 100)
        ○ Sg. F. Acc. (niht - ID:518, freq: 1106, in texts: 282, prob: 100)
        ○ Sg. F. Gen. (niht - ID:519, freq: 1106, in texts: 282, prob: 100)
        ○ Sg. F. Dat. (niht - ID:523, freq: 1106, in texts: 282, prob: 100)
        ○ Pl. F. Nom. (niht - ID:524, freq: 1106, in texts: 282, prob: 100)
        ○ Pl. F. Acc. (niht - ID:525, freq: 1106, in texts: 282, prob: 100)
  ○ *nihtian*
    ○ verb
      ○ infl. like sealfian - conjugate
        ○ Sg. 3.p. Ind. Pres. (w, 2, a, nihtt - ID:731105, freq: 1106, in texts: 282, prob: 98)
        ○ Sg. Imp. (w, 2, a, niht - ID:731109, freq: 1106, in texts: 282, prob: 100)
        ○ Pret. Part. (w, 2, a, nihtt - ID:731167, freq: 1106, in texts: 282, prob: 98)
  ○ *niht*
    ○ noun
      ○ infl. like hof - decline
        ○ Sg. N. Nom. (strong, niht - ID:13938165, freq: 1106, in texts: 282, prob: 100)
        ○ Sg. N. Acc. (strong, niht - ID:13938166, freq: 1106, in texts: 282, prob: 100)

○ **ON**
  ○ *on*
    ○ preposition
      ○ infl. like on - decline
        ○ (on - ID:4676, freq: 84713, in texts: 2217, prob: 100)
  ○ *unnan*
    ○ verb
      ○ infl. like unnan - conjugate
        ○ Sg. 1.p. Ind. Pres. (pp, 3, a, onn - ID:47159, freq: 84713, in texts: 2217, prob: 99)
        ○ Sg. 3.p. Ind. Pres. (pp, 3, a, onn - ID:47163, freq: 84713, in texts: 2217, prob: 99)

- WANRE
  - *wan*
    - adjective
      - infl. like til - decline
        - Sg. F. Gen. Pos. (strong, wanre - ID:2429155, freq: 1, in texts: 1, prob: 100)
        - Sg. F. Dat. Pos. (strong, wanre - ID:2429157, freq: 1, in texts: 1, prob: 100)
      - infl. like blinda - decline
        - Sg. N. Nom. Comp. (weak, wanre - ID:2429211, freq: 1, in texts: 1, prob: 100)
        - Sg. N. Acc. Comp. (weak, wanre - ID:2429212, freq: 1, in texts: 1, prob: 100)
        - Sg. F. Nom. Comp. (weak, wanre - ID:2429215, freq: 1, in texts: 1, prob: 100)
  - *wana*
    - adjective
      - infl. like blinda - decline
        - Sg. N. Nom. Comp. (weak, wanre - ID:2429602, freq: 1, in texts: 1, prob: 100)
        - Sg. N. Acc. Comp. (weak, wanre - ID:2429603, freq: 1, in texts: 1, prob: 100)
        - Sg. F. Nom. Comp. (weak, wanre - ID:2429606, freq: 1, in texts: 1, prob: 100)
  - *wana*
    - adjective
      - infl. like blinda - decline
        - Sg. N. Nom. Comp. (weak, wanre - ID:2430248, freq: 1, in texts: 1, prob: 100)
        - Sg. N. Acc. Comp. (weak, wanre - ID:2430249, freq: 1, in texts: 1, prob: 100)
        - Sg. F. Nom. Comp. (weak, wanre - ID:2430252, freq: 1, in texts: 1, prob: 100)
  - *wann*
    - adjective
      - infl. like blind - decline
        - Sg. F. Gen. Pos. (strong, wannre - ID:2433121, freq: 1, in texts: 1, prob: 99)
        - Sg. F. Dat. Pos. (strong, wannre - ID:2433125, freq: 1, in texts: 1, prob: 99)
      - infl. like blinda - decline
        - Sg. N. Nom. Comp. (weak, wannre - ID:2433337, freq: 1, in texts: 1, prob: 99)
        - Sg. N. Acc. Comp. (weak, wannre - ID:2433339, freq: 1, in texts: 1, prob: 99)
        - Sg. F. Nom. Comp. (weak, wannre - ID:2433345, freq: 1, in texts: 1, prob: 99)

Statistics: tokens/types: 4/4; not identified tokens/types:0/0; identified lemmata: 10; identified forms:32
verb: 3/7, noun: 2/8, preposition: 1/1, adjective: 4/16,
Running time of the script was: 0.023303031921387 seconds

Declination of: *neaht*
Paradigm: bóc
Show/Hide Variants

- ○ (514) **Sg. F. Acc.** - *neaht*
- ○ (513) **Sg. F. Nom.** - *neaht*
- ○ (526) **Pl. F. Gen.** - *neahta*
- ○ (520) **Sg. F. Gen.** - *neahte*
- ○ (527) **Pl. F. Dat.** - *neahtum*
- ○ (525) **Pl. F. Acc.** - *niht*
- ○ (524) **Pl. F. Nom.** - *niht*
- ○ (518) **Sg. F. Acc.** - *niht*
- ○ (523) **Sg. F. Dat.** - *niht*
- ○ (519) **Sg. F. Gen.** - *niht*
- ○ (517) **Sg. F. Nom.** - *niht*
- ○ (13938166) **Sg. N. Acc.** - *niht*
- ○ (13938165) **Sg. N. Nom.** - *niht*
- ○ (530) **Pl. F. Gen.** - *nihta*
- ○ (13938173) **Pl. N. Gen.** - *nihta*
- ○ (522) **Sg. F. Gen.** - *nihte*
- ○ (13938168) **Sg. N. Dat.** - *nihte*
- ○ (13938167) **Sg. N. Gen.** - *nihtes*
- ○ (13938172) **Pl. N. Acc.** - *nihto*
- ○ (13938170) **Pl. N. Nom.** - *nihto*
- ○ (13938171) **Pl. N. Acc.** - *nihtu*
- ○ (13938169) **Pl. N. Nom.** - *nihtu*
- ○ (531) **Pl. F. Dat.** - *nihtum*
- ○ (13938174) **Pl. N. Dat.** - *nihtum*
- ○ (516) **Sg. F. Acc.** - *næht*
- ○ (515) **Sg. F. Nom.** - *næht*
- ○ (528) **Pl. F. Gen.** - *næhta*
- ○ (521) **Sg. F. Gen.** - *næhte*
- ○ (529) **Pl. F. Dat.** - *næhtum*

12.4.   Texts Used to Testing the Analyser

12.4.1.   House on the Rock

Ælc þāra þe þās mīn word gehīerþ, and þā wyrcþ, biþ gelīc þæm wīsan were, sē his hūs ofer stān getimbrode. Þā cōm þær regen and micel flōd, and þær blēowon windas, and āhruron on þæt hūs, and hit nā ne fēoll: sōþlīce hit wæs ofer stān getimbrod.

And ælc þāra þe gehīerþ þās mīn word, and þā ne wyrcþ, sē biþ gelīc þæm dysigan menn, þe getimbrode his hūs ofer sandcēosol. Þā rīnde hit, and þær cōmon flōd, and blēowon windas, and āhruron on þæt hūs, and þæt hūs fēoll; and his hryre wæs micel.

(Čermák & Znojemská, 2001)

12.4.2.   Wise and Foolish Virgins

Ðonne biþ heofona rīce gelīc þæm tīen fæmnum, þe þā lēohtfatu nāmon, and fērdon ongēan þone brydguman and þā bryd. Hiera fīf wæron dysige, and fīf glēawe. Ac þā fīf dysigan nāmon lēohtfatu, and ne nāmon nānne ele mid him; þā glēawan nāmon ele on hiera fatum mid þæm lēohtfatum. Þā se brydguma ielde, þā hnappodon hīe ealle, and slæpon. Witodlīce tō midre nihte man hrīemde, and cwæþ: 'Nū se brydguma cymþ, faraþ him tōgēanes.' Þā ārison ealle þā fæmnan, and glengdon hiera lēohtfatu. Þā cwædon þā dysigan to þæm wīsum: 'Sellaþ ūs of ēowrum ele, for þæm ūre lēohtfatu sind ācwenctu.' Þā andswarodon þā glēawan, and cwædon: 'Nese; þy læs þe wē and gē næbben genōg. Gāþ to þæm cīependum, and bycgaþ ēow ele.' Witodlīce, þā hīe fērdon, and woldon bycgan, þā cōm se brydguma; and þā þe gearwe wæron ēodon inn mid him to þæm gieftum; and sēo duru wæs belocen. Þā æt nīehstan cōmon þā ōþre fæmnan, and cwædon: 'Dryhten, dryhten, læt ūs inn.' Þā andswarode hē him, and cwæþ: 'Sōþ ic ēow secge, ne cann ic ēow.' Witodlīce, waciaþ, for þæm þe gē nyton ne þone dæg ne þā tīd.

(Čermák & Znojemská, 2001)

12.4.3.   The Sun and the Moon

Sēo sunne gæð betwēonan heofonan and eorðan: on dæg bufon eorðan and on niht under ðysse eorðan. Æfre hēo byð yrnende ymbe þās eorþan, and ealswā lēohte scīnð under þære eorðan on niht, swā swā hēo on dæg dēð bufan ūrum hēafdum.

On þā healfe þe hēo scīnð, þær byð dæg; and on þā healfe þe hēo ne scīnð, ðær byð niht. Wē hātaþ ānne dæg fram sunnan upgange oð æfen.

Dæghwāmlīce ðæs mōnan lēoht byþ weaxende oððe waniende. Simle hē went his hrycg tō ðære sunnan. Hwonne hē weaxeð is hē full. Sōðlīce ðæs mōnan gēar hæfð seofon and twēntig daga and eahta tīda. On ðām fyrste hē underyrnð ealle þā twelf tācna þe sēo sunne undergæð twelf mōnað.

On ðām gēare sind getealde twelf mōnðas, and twā and fīftig wucena, þrēo hund daga and fīf and syxtig daga and ðær tō ēacan syx tīda.

(Čermák & Znojemská, 2001)

12.4.4. The Voyages of Ohthere and Wulfstan

Ohthere sæde his hlaforde, Ælfrede cyninge, þæt he ealra Norðmonna norþmest bude.

He cwæþ þæt he bude on þæm lande norþweardum wiþ þa Westsæ.

He sæde þeah þæt þæt land sie swiþe lang norþ þonan;

Ac hit is eal weste, buton on feawum stowum styccemælum wiciaþ Finnas, on huntoþe on wintra, and on sumera on fiscaþe be þære sæ.

He sæde þæt he æt sumum cirre wolde fandian hu longe þæt land norþryhte læge, oþþe hwæþer ænig mon be norþan þæm westenne bude.

Þa for he norþryhte be þæm lande.

Let him ealne weg þæt weste land on þæt steorbord, ond þa widsæ on þæt bæcbord þrie dagas.

Þa wæs he swa feor norþ swa þa hwælhuntan firrest faraþ.

Þa for he þa giet norþryhte swa feor swa he meahte on þæm oþrum þrim dagum gesiglan.

Þa beag þæt land þær eastryhte, oþþe seo sæ in on þæt lond, he nysse hwæþer, buton he wisse þæt he þær bad westan windes ond hwon norþan, ond siglde þa east be lande swa swa he meahte on feower dagum gesiglan.

Þa sceolde he þær bidan ryhtnorþanwindes, for þæm þæt land beag þær suþryhte, oþþe seo sæ in on þæt land, he nysse hwæþer.

Þa siglde he þonan suþryhte be lande swa swa he mehte on fif dagum gesiglan.

Þa læg þær an micel ea up in on þæt land.

Þa cirdon hie up in on þa ea, for þæm hie ne dorston forþ bi þære ea siglan for unfriþe; for þæm þæt land wæs eall gebun on oþre healfe þære eas.

Ne mette he ær nan gebun land, siþþan he from his agnum ham for; ac him wæs ealne weg weste land on þæt steorbord, butan fiscerum and fugelerum ond huntum, ond þæt wæron eall Finnas; ond him wæs a widsæ on þæt bæcbord.

Þa Beormas hæfdon swiþe wel gebun hira land: ac hie ne dorston þæron cuman.

Ac þara Terfinna land wæs eal weste, buton þær huntan gewicodon, oþþe fisceras, oþþe fugelera.

(McGillivray & Chevallier, 2014)

12.4.5. Cynewulf and Cyneheard

754. Her Cuþred forþferde, & Cyneheard onfeng biscepdome æfter Hunferþe on Wintanceastre; & Cantwara burg forbærn þy geare, & Sigebryht feng to Wesseaxna rice, & heold an gear.

755. Her Cynewulf benam Sigebryht his rices & West Seaxna wiotan for unryhtum dædum, buton Hamtunscire; & he hæfde þa oþ he ofslog þone aldormon þe him lengest

wunode; & hiene þa Cynewulf on Andred adræfde, & he þær wunade oþ þæt hiene an swan ofstang æt Pryfetes flodan;

& he wræc þone aldormon Cumbran; & se Cynewulf oft miclum gefeohtum feaht wiþ Bretwalum; & ymb x x x i wintra þæs þe he rice hæfde, he wolde adræfan anne æþeling se was Cyneheard haten, & se Cyneheard wæs þæs Sigebryhtes broþur;

& þa geascode he þone cyning lytle werode on wifcyþþe on Merantune, & hine þær berad, & þone bur utan beeode ær hine þa men onfunden þe mid þam kyninge wærun;

& þa ongeat se cyning þæt, & he on þa duru eode, & þa unheanlice hine werede, oþ he on þone æþeling locude, & þa utræsde on hine, & hine miclum gewundode.

& hie alle on þone Cyning wærun feohtende oþ þæt hie hine ofslægenne hæfdon; & þa on þæs wifes gebærum onfundon þæs cyninges þegnas þa unstilnesse & þa þider urnon swa hwelc swa þonne gearo wearþ & radost; & hiera se æþeling gehwelcum feoh & feorh gebead, & hiera nænig hit geþicgean nolde.

Ac hie simle feohtende wæran oþ hie alle lægon butan anum Bryttiscum gisle, & se swiþe gewundad wæs.

Ða on morgenne gehierdun þæt þæs cyninges þegnas þe him beæftan wærun þæt se cyning ofslægen wæs, þa ridon hie þider, & his aldormon Osric, & Wiferþ his þegn, & þa men þe he beæftan him læfde ær, & þone æþeling on þære byrig metton þær se cyning ofslægen læg, & þa gatu him to belocen hæfdon & þa þær to eodon;

(McGillivray, Cynewulf and Cyneheard, 2014)

12.4.6. Beowulf

	Com on wanre niht
scriðan sceadugenga.	Sceotend swæfon,
þa þæt hornreced	healdan scoldon,
ealle buton anum.	Þæt wæs yldum cuþ;
þæt hie ne moste,	þa metod nolde,
se scynscaþa	under sceadu bregdan;
ac he wæccende	wraþum on andan
bad bolgenmod	beadwa geþinges.
Ða com of more	under misthleoþum
Grendel gongan,	godes yrre bær;
mynte se manscaða	manna cynnes
sumne besyrwan	in sele þam hean.
Wod under wolcnum	to þæs þe he winreced,
goldsele gumena,	gearwost wisse,
fættum fahne.	Ne wæs þæt forma sið

þæt he Hroþgares ham gesohte;
næfre he on aldordagum ær ne siþðan
heardran hæle, healðegnas fand.
Com þa to recede rinc siðian,
dreamum bedæled. Duru sona onarn,
fyrbendum fæst, syþðan he hire folmum [æthr]an;
onbræd þa bealohydig, ða he [ge]bolgen wæs,
recedes muþan. Raþe æfter þon
on fagne flor feond treddode,
eode yrremod; him of eagum stod
ligge gelicost leoht unfæger.
Geseah he in recede rinca manige,
swefan sibbegedriht samod ætgædere,
magorinca heap. Þa his mod ahlog;
mynte þæt he gedælde, ærþon dæg cwome,
atol aglæca, anra gehwylces
lif wið lice, þa him alumpen wæs
wistfylle wen. Ne wæs þæt wyrd þa gen
þæt he ma moste manna cynnes
ðicgean ofer þa niht. Þryðswyð beheold
mæg Higelaces, hu se manscaða
under færgripum gefaran wolde.
Ne þæt se aglæca yldan þohte,
ac he gefeng hraðe forman siðe
slæpendne rinc, slat unwearnum,
bat banlocan, blod edrum dranc,
synsnædum swealh; sona hæfde
unlyfigendes eal gefeormod,
fet ond folma. Forð near ætstop,
nam þa mid handa higeþihtigne
rinc on ræste, ræhte ongean
feond mid folme; he onfeng hraþe

inwitþancum	ond wið earm gesæt.
Sona þæt onfunde	fyrena hyrde
þæt he ne mette	middangeardes,
eorþan sceata,	on elran men
mundgripe maran.	He on mode wearð
forht on ferhðe;	no þy ær fram meahte.
Hyge wæs him hinfus,	wolde on heolster fleon,
secan deofla gedræg;	ne wæs his drohtoð þær
swylce he on ealderdagum	ær gemette.

(Kiernan, 2011)

12.4.7.   Sermo Lupi ad Anglos

And lā, hū mæg māre scamu þurh Godes yrre mannum gelimpan þonne ūs dēð gelōme for āgenum gewyrhtum? Ðēh þræla hwylc hlāforde æthlēape and of crīstendōme tō wīcinge weorþe, and hit æfter þām eft geweorþe þæt wæpngewrixl weorðe gemæne þegene and þræle, gif þræl þæne þegen fullīce āfylle, licge ægylde ealre his mægðe; and, gif se þegen þæne þræl þe hē ær āhte fūllīce āfylle, gylde þegengylde. Ful earhlice laga and scandlice nydgyld þurh Godes yrre ūs syn gemæne, understande sē þe cunne; and fela ungelimpa gelimpð þysse þēode oft and gelōme. Ne dohte hit nū lange inne ne ūte, ac wæs here and hete on gewelhwilcan ende oft and gelōme, and Engle nū lange eal sigelēase and tō swyþe geyrigde þurh Godes yrre; and flotmen swā strange þurh Godes þafunge þæt oft on gefeohte ān fēseð tyne, and hwīlum læs, hwīlum mā, eal for ūrum synnum. And oft tyne oððe twelfe, ælc æfter ōðrum, scendað tō bysmore þæs þegenes cwenan, and hwīlum his dohtor oððe nydmāgan, þær hē on lōcað, þe læt hine sylfne rancne and rīcne and genōh gōdne ær þæt gewurde. And oft þræl þæne þegen þe ær wæs his hlāford cnyt swīþe fæste and wyrcð him tō þræle þurh Godes yrre. Wālā þære yrmðe and wālā þære woroldscame þe nū habbað Engle, eal þurh Godes yrre! Oft twēgen sæmæn, oððe þry hwīlum, drīfað þā drāfe crīstenra manna fram sæ tō sæ, ūt þurh þās þēode, gewelede tōgædere, ūs eallum tō woroldscame, gif wē on eornost ænige cūþon āriht understandan; ac ealne þæne bysmor þe wē oft þoliað wē gyldað mid weorðscipe þām þe ūs scendað: wē him gyldað singallīce, and hy ūs hynað dæghwāmlīce; hy hergiað and hy bernað, rypað and rēafiað and tō scipe lædað; and lā, hwæt is ænig ōðer on eallum þām gelimpum būtan Godes yrre ofer þās þēode swutol and gesæne?

(Čermák & Znojemská, 2001)

12.4.8.   Lindisfarne Gospels

behealdas þæt soðfæstnise iuerre gie doas before monnum þæt gie se geseno from him eaðe mæg mearde nabbas ge mið fader iurre se ðe in heafnas is

mið ðy ðonne ðu doas ælmessa nelle ðu bema vel stocc singa before ðec suæ legeras gewyrcas in somnungum ond in londum vel in gemærum þæt hia se gearðad from monnum soðlice vel soð is ic cueðo iuh to hie gefengon mearde hiora

ðu vel ðeh uutedlice wyrcende ða ællmissa nyta winstra ðin huæt wyrcas vel doas suuðra ðin

þæt sie ællmessa ðin in degelnisse ond fader ðin se ðe gesiið in degelnisse forgeldeð ðe

mið ðy gie gebiddas ne wosas ge suæ legeras ða ðe lufas in somnungum ond huommum ðara plæcena vel worðum stondes vel stondende gebiddas vel to gebiddanne þæt hia gesene sie from monnum soðlic ic cueð iuch to onfengon mearde heara

ðu uutedlice mið ðy gie gebiddes vel ingeong vel inga in cotte ðinum ond gesparrado dure ðin gbidd fæder ðinne in degolnis ond fæder ðin se ðe gesiið vel locas in degelnisse forgeldes ðe

(Sweet, 1887)

12.4.9. Rushworth Gospels

behaldeþ þæt ge eowre soþfestnisse ne doan fore monnum þæt ge sie geseane from heom elles vel elcur ge ne habbaþ lean vel mearde mid eower fæder þæne þe in heofunum is.

forþon þonne þu wirce ælmisse ne blau þu beman for þe swa liceteras doan in heora somnungum ond in tunum þæt hie sie weorþade from monnum soþ ic sæcge eow hie onfengun heora lean

ðe þonne wircendum ælmesse nyte se winstrae hond þin hwat þin sio swiþre dóa

þæt þin ælmes sie in degulnisse ond þin fæder se þe gesið in degulnisse geldeþ ðe

ond þonne ge bidde eow ne beoþ ge swa liceteras þa þe lufigaþ stalle vel stonde in gesomnungum ond in hwommum worþana stondende him gebidde þæt hie sie gesenæ from monnum soþ ic sæcge eow hie onfengun heora lean

þu þonne þonne þu gebidde ga in þine cofan ond betun þine dure bidde þin fæder ond þin fæder se þe gesihð in degulnisse geldeþ ðe

(Sweet, 1887)

12.4.10. Oswulf's Charter

Ic Osuulf aldormonn mid Godes gæfe, ond Beornðryð min gemecca sellað to Cantuarabyrg, to Cristes cirican, ðæt lond æt Stanhamstede, swuluncga, Gode allmehtgum ond ðere halgon gesomnuncgæ, fore hyhte and fore aedleane ðæs aecan ond ðaes towardon lifes, ond fore uncerra saula hela ond uncerra bearna. Ond mid micelre eaðmodnisse biddað ðæt wit moten bion on ðem gemanon ðe ðaer Godes ðiowas siondan, ond ða menn ða ðaer hlafordas wæron, ond ðara monna ðe hiora lond to ðaere cirican saldon; ond ðættæ mon unce tide ymb tuælf monað mon geuueorðiae on godcundum godum ond æc on aelmessan suæ mon hiora doeð.

Ic ðonne Uulfred mid Godes gaefe arc. epis. ðas forecuaedenan uuord fulliae, ond bebeode ðæt mon ymb tuælf monað hiora tid boega ðus geuuerðiae to anes daeges to Osuulfes tide ge mid godcundum godum ge mid aelmessan, ge aec mid higna suesendum. Ðonne bebeode ic ðaet mon ðas ðing selle ymb tuælf monað of Liminum, ðe ðis forecuaedene lond to limpeð, of ðaem ilcan londe æt Stanhamstede: huaetenra hlafa ond clenra, ond án hriðer dugunde, ond iiii scep, ond tua flicca, ond goes, ond hennfuglas, ond pund caeses, gif hit fuguldaeg sie. Gif hit ðonne festendæg sie, selle mon uuege cæsa, ond fisces, ond butran, ond aegera ðaet mon begeotan maege, ond ombra godes uuelesces aloð

ðet limpeð to mittum, ond mittan fulne huniges, oðða tuegen uuines, sue hwaeder suae mon ðonne maege. Ond of higna gemenum godum ðaer aet ham mon geselle gesuflra hlafa to aelmessan for hiora saula, suae mon aet hlaforda tidum doeð. Ond ðas forecuedenan suesenda all agefe mon ðem reogolwarde, ond he brytnie swæ higum maest red sie ond ðaem sawlum soelest. Aec mon ðaet weax ágæfe to ciricican, ond hiora sawlum nytt gedoe ðe hit man fore doeð. Aec ic bebeode minum aefterfylgendum ðe ðaet lond hebben aet Burnan ðaet hiae simle ymb monað foran to ðære tide gegeorwien ten hund hlafa ond swae feola sufla, ond ðet mon gedele to almessan aet ðere tide fore mine sawle ond Osuulfes ond Beornðryðe aet Cristes cirican, ond him se reogolword on byrg gebeode foran to hwonne sio tid sie. Aec ic bidde higon ðette hie ðas godcundan god gedon aet ðere tide fore hiora sawlum: ðaet eghwilc messepriost gesinge fore Osuulfes sawle twa messan, twa fore Beornðryðe sawle; ond aeghwilc diacon arede twa passione fore his sawle, twa fore hire; ond eghwilc Godes ðiow gesinge twa fiftig fore his sawle, twa fore hire, ðaette ge fore uueorolde sien geblitsade mid ðem weoroldcundum godum ond hiora saula mid ðem godcundum godum. Aec ic biddo higon ðaet ge me gemynen aet ðere tide mid suilce godcunde gode suolce iow cynlic ðynce, ic ðe ðas gesettnesse sette ge hueder ge for higna lufon ge ðeara saula ðe haer beforan hiora namon auuritene siondon.Ðis is gesetnes Osuulfes ond Bearnðryðe.

(Sweet, 1887)

12.5. Sample of the Input Dictionary

ID	lemma	wright	noun	pron.	adj.	verb	part.	adv.	prep.	conj.	interj.	num.	w. vb.	s. vb.	c. vb.	pp. vb.	a. vb.	un. vb.	m. n.	f. n.	n. n.	u. n.
008031	DRȲ	142;388	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
008032	drý-cræft	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
008033	drý-cræftig	NULL	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
008035	drýfan	NULL	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
008036	drygan	530	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
008040	dryht	390	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
008041	Dryht	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
008042	dryht-bearn	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
008043	dryht-cwén	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
008044	dryht-ealdor	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
008045	dryhten	288;340;563	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
008047	dryhten-bealo	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
008048	dryhten-dóm	NULL	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

12.6.   Source Code and the Analyser Access

Source code and the input files are provided with the paper in electronic form only. Once the application is deemed secure and for publication, it will be accessible (including its source codes) at: http://bosworthtoller.com/analyser.