ADVANCED ISSUES CONCERNING THE LEMMATISATION OF THE OLD ENGLISH GRADED ADVERBS

Yosra Hamdoun
Universidad de La Rioja

Lemmatisation is still a pending task of historical linguistics, which makes the contribution of a fully lemmatized corpus a necessary source for the study of Old English. This paper aims at presenting the lemmatisation of the Old English adverbs in the comparative forms attested in *The York-Toronto-Helsinki Parsed Corpus of Old English Prose.* This constitutes a pilot study on the assignment of a lemma to an Old English non-verbal category. The starting point has been the automatic extraction of the attested forms assigned the tags ADVR and ADVS. In a second step, each attested form has been manually assigned a lemma provided by the lexical database *Nerthus*. Finally, the results have been checked against two other sources, namely *The Dictionary of Old English* and Seelig's (1930) work on Old English comparative adjectives and adverbs, both of which have contributed to the refinement and completion of the analysis. Overall, this study offers a methodology for lemma assignment that has been proven feasible for the lemmatisation of a non-verbal category and may be applicable to other non-verbal categories. GAUDEAMUS

Keywords: lemmatisation; superlative; comparative; Old English; corpus linguistics

1.   Introduction

This article discusses the results obtained from a pilot study on the lemmatisation of Old English adverbs in the comparative and the superlative forms from a corpus-based perspective. It mainly delves into the methodology applied and the issues that have arisen in its implementation as well as into the advantages of having a lemmatised corpus in a historical language such as Old English.

Lemmatisation can be described as the process by which a uniform heading is assigned to the different elements of a lexical corpus that are represented by the same lexeme (Burkhanov, 1998: 122). This means that lemmatisation gathers under one lexicographical entry all the attested forms in a corpus or database. Through this process it is possible to establish a relationship between the textual attestation *hluddor* and the lemma *hlude* 'loud, aloud'. At the moment there is no fully lemmatised inventory of Old English inflected forms. In order to fill in this research gap, this work presents a methodology for the lemmatisation of Old English comparative and superlative adverbs.

This paper also aims to provide evidence of how a corpus-based approach makes the lemmatisation process more efficient and targets the aspects that need to be considered when dealing with historical corpora. This article is organized as follows: Section 2 revolves around the interdependence of corpus linguistics and historical linguistics and how the latter benefits from corpus-based approaches. Section 3 offers an overview of Old English adverbs, especially of their formation and main aspects concerning adverbial gradation into the comparative and superlative. Section 4 defines and sequences the methodological steps in the lemmatisation process. Section 5 discusses the results and the main difficulties encountered in this process. Finally, Section 6 provides some general conclusions.

2.   Corpus Linguistics and Historical Linguistics

In this section I will first draw a general picture of corpus linguistics. Then, I will briefly review how corpus linguistics has assisted the

study of historical languages. Finally, the main Old English electronic sources will be presented and reviewed.

2.1. A Brief Overview of Corpus Linguistics

Corpus linguistics has traditionally been conceived as a method of research rather than a linguistic discipline. Rissanen (2008: 54) summarised this trend by defining corpus linguistics as a "linguistic study based on a corpus". Corpus linguistics is thus based on the study of a compiling process that must be focused on the selection of a range of relevant texts for the study that will be conducted. Although, as Faaβ (2017: 125) explains, "finding some samples for each of the senses of a word is necessary, but not enough, as also it is deemed relevant how often a linguistic phenomenon occurs". Some reasons why historical corpora face a shortcoming in this aspect include the limited number of surviving texts, the inaccessibility and partial preservation of these, along with copyright problems.

Faaβ (2017: 124-445) observes the need to include data annotation for any corpora with the collaboration of electronic lexicography. For a corpus to become a valuable tool for the study of a historical language, it should be both morphologically and syntactically annotated and fully lemmatised. The annotation of corpora requires an initial process of tokenization, i.e. of identification of word units as tokens. Both morphological and syntactical annotations are crucial tasks when studying morphologically rich languages like Old English. Part-of-speech tagging allows for the disambiguation of homographs and maximizes the accuracy of the corpus content. The resulting systematic, compiled and annotated data can serve both as a source -the data collected allows the inference of properties of the words appearing in a corpus- and as a resource for finding evidence for research. The present study makes use of and contributes to both lines of research of corpus linguistics.

On the one hand, it undertakes a pilot study on the lemmatisation of the Old English adverbs by selecting the inflectional forms as recorded in the *York-Toronto-Helsinki Parsed Corpus of Old English* (see section 4.1 Sources). On the other hand,

this research contributes to the implementation of a lemmatisation methodology in a non-verbal category that can be further applied to the other non-verbal categories that remain unlemmatised for the moment.

2.2.   Corpus Linguistics and Old English

The rise of electronic corpora has made the retrieval and analysis of large amounts of data more efficient and systematic. The evolution of corpora, as López-Couso (2016: 129) asserts, has given rise to efficient research tools that are applicable "to many areas of the historical study of English, including morphosyntax, lexicography, semantics, grammaticalizations, pragmatics and sociolinguistics", besides historical phonology or semantic change.

   The introduction of electronic corpora into historical linguistics allows researchers to study ancient materials with new formats. Among its improvements, electronic corpora facilitate the comparison of past and present stages of the same language in context, and they ultimately provide accessible paths for the collection of written evidence, which is the only material available when studying a historical language. Despite the use of electronic corpora for lexicographical purposes, lexicography is still semi-automatic tasks. Although the extraction of data has been to a great extent automated, the lemmatization and interpretation of data still requires manual revision.

2.3.   Main Old English Electronic Sources

The desire to compile the evolution of a language through the texts of different periods was materialized in the *Helsinki Corpus of English texts* (Kytö and Rissanen: 2008). This structured multi-genre diachronic corpus includes chronologically organized text samples from Old, Middle and Early Modern English, enabling researchers to conduct a diversity of studies with an eye towards diachronic variation and language change.

      The Dictionary of Old English Corpus (DOEC) is one of the first electronic resources that can be used for the study of this historical language. It includes at least one copy of every surviving

text in Old English. The DOEC is, for the moment, the largest corpus of the period, with almost three million words in Old English and nearly one million words in Latin. This is an excellent example of how an extensive dictionary can be constructed on the basis of a corpus.

The YCOE, for its part, includes roughly 1.5 million words and encompasses a variety of text genres, dates of composition and authors. This corpus contains all the major Old English prose works. The singularity of this corpus is that each word is syntactically and morphologically annotated. The *York-Toronto-Helsinki Parsed Corpus of Old English Poetry* includes approximately fifty thousand words and follows the same annotation system as the prose corpus. Both corpora allow automatic searches of syntactic structure, constituent order and lexical items. Section 4.1 will describe this corpus in more detail.

3.   Old English Adverbs

Before delving into the lemmatisation methodology, it is necessary to briefly review the grammatical behaviour of the Old English adverbs from a historical perspective.

The earliest written evidence of the English language that we know of is what has been called Old English. From a linguistic perspective, scholars consider Old English a synthetic language because "there is a close relation between the form and function of the words that is embodied in its rich use of inflections" (Smith 2009: 22). Although Mitchell and Robinson (1985: 62) prefer to qualify it as a "half inflected" one because of the preservation of only four cases from the eight that existed in Indo-European and the extensive use of prepositions. The lexicon of Old English is mainly composed of inherited Germanic words, although there were Latin borrowings. The most productive word formation processes include affixation and compounding.

Adverbs were mostly created through the addition of suffix '-e'. This suffix was particularly productive when deriving adverbs from adjectival stems. For example, *dēop* 'deep' > *dēope* 'deeply'; *biter* 'bitter' > *bit(e)re* 'bitterly'; *clǣn* 'pure' > *clǣne* 'purely'. Some

other adverbs were created through the addition of the suffix '-e' to an adjective ending in 'lic-'. For example, *nytlic* 'useful' > *nytlīce* 'usefully'; *sārlic* 'grievous' > *sārlīce* 'grievously'. Due to the fact that several adverbs were derived in this manner, the suffix '-lice' has resulted in an adverbial marker. In some cases, pairs of adverbs arose with both the '-e' and '-lice' endings, for instance *hearde* and *heardlīce*.

Adverbs only inflect for the comparative and the superlative. The regular comparative and superlative endings for adverbs deriving from adjectives or from lexicalised forms are *-or* and *-ost*; for instance, *oft* 'often' - *oftor* (comparative)*, oftost* (superlative). Alternative endings undergo vowel change, including *-ur* and *-ar* for the comparative and *-ast, -est* and *-ust* for the superlative. Fulk (2018: 240) remarks that a few Old English adverbs -and also some adjectives- form the superlative through double suffixation (*-m-i/est*). This is the case of *innemest* 'innermost' and *yfemest* 'uppermost'. In this regard, Campbell (1959: 278) states that the ending *-mest* is especially common when a compared adjective is derived from an adverb: *inne* 'inside' – *innerra* (adv./adj. comp.) – *innemest* (adv./adj. superl.). The root vowel of several adverbs mutates when forming the comparative and the superlative; an example is *feorr* 'far' - *fierr* – *firrest* (Campbell 1959: 278)*. Other adverbs undergo suppletive comparison (Fulk 2018: 240), that is, they form the comparative and the superlative by taking a different stem form of the adverb. In Old English we find examples such as *yfle* 'evil' - *wiers* - *wierst*; *wel* 'well' - *bet/sēl* - *bet*(*e*)*st/best/sēlest*.

4. Finding and Lemmatising Old English Comparative and Superlative Adverbs

This study focuses on the lemmatisation of the comparative and superlative forms of Old English adverbs. The following sections describe the characteristics of the data sources selected to guide this research (4.1.), and the step-by-step methodology adopted (4.2.).

Previous literature on the lemmatisation of historical languages is quite scarce. This study is based on previous ones conducted by other members of the *Nerthus Project*. These studies

include strong verbs (Metola Rodríguez: 2015, 2017 and 2018), weak verbs (Tío Sáenz: 2015, 2019) and preterite-present, anomalous and contracted verbs (García Fernández: 2019). Although they involved substantially different procedures, a semiautomatic methodology was adopted in all three cases. The present study adopts a three-step semi-automatic methodology which consists of an initial automatic search, followed by a manual lemma assignment and final contrastive analysis of the results with DOE and Seelig. This pilot study focuses on the adverbial category for two main reasons: in quantitative terms, this class is the least numerous of the lexical categories; in qualitative terms, adverbs present a rather limited inflective spectrum, as they may only inflect for the comparative and the superlative.

4.1. Sources

The lemmatisation process requires both lexicographical and textual sources. In order to start lemmatising, two elements are required: an inventory of inflectional forms to lemmatise and a list of headwords that will serve as the lemmas for these forms. To begin with, the YCOE has provided the tagged list of inflectional forms that will be lemmatised. The prose segment is annotated: the Part of Speech files (POS) include the grammatical categories or subcategories and the Parsed annotation file (PSD) identifies the syntactic structure of a sentence, based on the annotation system used by the *Penn- Helsinki Parsed Corpus of Middle English*. Examples (1) and (2) show the morphological and syntactic analysis of the same sentence as represented in the YCOE:

(1) æfter_P þisum_D^D wordum_N^D heo_PRO^N mid_P modes_N^G anrædnesse_N   awrat_VBDI   oðer_ADJ^A   gewrit_N^A coapollo,ApT:20.13.420_ID and_CONJ þæt_D^A geinseglode_VBD coapollo,ApT:20.13.421_Iñls   and_CONJ   sealde_VBD Apollonio_NR^D ._. coapollo,ApT:20.13.422_ID

(2) ((CODE <T06080020900,20.13>)
        (IP-MAT (PP (P +After)
          (NP-DAT (D^D +tisum) (N^D wordum)))
        (NP-NOM (PRO^N heo))
        (PP (P mid)
          (NP (NP-GEN (N^G modes))

```
              (N anr+adnesse)))
        (VBDI awrat)
        (NP-ACC    (ADJ^A    o+der)    (N^A    gewrit)))    (ID
coapollo,ApT:20.13.420))
      ( (IP-MAT (CONJ and)
        (NP-NOM *con*)
        (NP-ACC (D^A +t+at))
        (VBD geinseglode)) (ID coapollo,ApT:20.13.421))
      ( (IP-MAT (CONJ and)
        (NP-NOM *con*)
        (VBD sealde)
        (NP-DAT (NR^D Apollonio))
       (. .)) (ID coapollo,ApT:20.13.422))
```

The lemma list has been retrieved from the Old English lexical database *Nerthus*. *Nerthus* is part of a relational database called *The Grid* (Martín Arista: 2016) that also includes *Freya,* a secondary source database, and *Norna,* a relational database. *Nerthus* files more than 31,000 predicates. For each predicate the database presents information about its alternative spellings, category, translation, inflectional morphology and inflectional forms. *Nerthus* draws on the main sources of reference in traditional Old English lexicography, including *A Concise Dictionary of Anglo-Saxon (and Supplement)* by Clark Hall and Meritt, Bosworth-Toller's (1973) *Anglo-Saxon Dictionary* (including Toller's *Supplement* and Campbell's *Addenda)* and Sweet's (1976) *Student Dictionary of Anglo-Saxon*.

Only three fields of this database have been of interest for this study, viz. predicate, alternative spelling and predicate translation. Figure 1 displays an Excel file with the exported data from *Nerthus*

	A	B	C	
1	predicate	alternative_spellings	predicate_translation	
2	cēne 2	cўne 2	in warlike wise	
3	(ge)bēotlīce		in a threatening manner, threateningly (BT)	
4	(ge)blīðe 2		joyfully, gladly	
5	(ge)brægdenlīce	(ge)bregdenlīce	cunningly, deceitfully (DOE)	
6	(ge)cwēmlīce		graciously, kindly, humbly, satisfactorily	
7	(ge)cyndelīce		naturally (DOE)	
8	(ge)dīegollīce	(ge)dēgollīce, (ge)dēagollīce, (ge)dēogollīce, (ge)dӯgollīce, (ge	secretly; softly (of the voice)	
9	(ge)ðwǣrelīce	(ge)ðwǣrlīce (BT)	in accord (Sweet)	
10	(ge)dwollīce		foolishly, heretically; erroneously, ignorantly, stupi	
11	(ge)ðyldelīce	(ge)ðyldiglīce (BT), (ge)ðyldelice (BT)	patiently, quietly	
12	(ge)ðyldiglīce		patiently (Sweet)	
13	(ge)dyrstiglīce	(ge)dyrstelīce	boldly, daringly (BT)	
14	(ge)efenlīce	emblīce (BT)	equally, evenly, alike; patiently	
15	(ge)endebyrdlīce		in an orderly manner, in order, in succession	
16	(ge)fæstlīce	feastlīce (BT), festlīce (BT)	certainly. fixedly, steadily, constantly; unceasingly;	
17	(ge)flitmǣlum	(ge)flītmǣlum (BT)	contentiously, emulously	
18	(ge)frēolīce	(ge)frīolīce (BT)	freely, readily; as a festival	
19	(ge)fyrn	(ge)firn (BT)	formerly, of old, long ago, once	
20	(ge)hāðeortlīce	(ge)hāthortlīce, (ge)hāthyrtlīce	furiously, ardently, fervently (BT)	
21	(ge)herigendlīce	(ge)hergendlīce (BT)	praiseworthily (BT)	

Figure 1. Nerthus lemma list

The predicate column contains the headwords that will be assigned to the inflectional forms. The central column offers information about the alternative spellings, if any, for each headword. Finally, the rightmost column provides a translation of the OE headword into Present Day English. If compared with the total amount of predicates stored in *Nerthus,* the number of adverbs represents just five per cent of the OE lexicon. A total of 1,755 headwords correspond to the category adverb.

The *Dictionary of Old English* (DOE) has been the main lexicographical source that has guided the validation of the lemmatisation process. The DOE has so far published headwords starting with letters A-I which provides a detailed description of the vocabulary belonging to the Old English period over six centuries (600 – 1150). The entries contain grammatical information of the headword, namely part of speech, gender and grammatical class. In addition, entries gather the attested spellings of the word in the corpus, including the Cameron number, the inflectional forms, dialectal variations, the number of occurrences in the corpus and the meaning accompanied by a few textual citations. The most common meaning of the headword normally appears first, followed by more technical, metaphorical or less common meanings. Two or more entries are created for words belonging to different grammatical categories but which are formally alike. For instance, Figure 2

displays the two entries for *brēme* 'famously', one as an adjective, and the other as adverb.

Figure 2. DOE's sample entries for *breme*

The *Dictionary of Old English Corpus* (DOEC) has also been consulted as it allows to search for words in context and also provides the number of attestations of a form. Finally, Seelig's *Die Komparation der Adjektiva und Adverbien im Altenglischen*, published in 1930, has helped to complete and refine the analysis, especially for letters L to W, as this information is not yet available at the DOE. Seelig compiles all the inflectional forms that belong to the regular comparison of adverbs. These adverbs form the comparative and superlative through the addition of suffixes -*or* and -*ost* respectively. For example, *fæstlīce* 'fast' (*fæstlicor, fæsðlicor, fæstlicost, fæstlicast*); *smale, smæle* 'small' (*smælor, smalost*). The second group is composed of adverbs that undergo a vocalic change in the stem in some of their comparative and superlative forms. Only nine paradigms are listed in this group (1930: 71-74), among which we find *heah, hea* 'high' (*hearor, hyhst*) and *softe* 'soft' (*seft, softor, softost*). The third group of adverbs are irregular in the sense that the comparative and superlative are created from a different stem. A total of six adverbs have been identified by the author as undergoing irregular comparison (1930: 75), including *wyrs* 'worse' (*wærse, wiers, wirs, wyrs, wierst, wyrrest, wyrst*) or *sēl* 'better' (*sæl, sēlast, sēlest, sēlost*).

4.2. Assigning Lemmas to the Old English Adverbs

To obtain the list of inflectional forms that will be lemmatised the following process was completed:

The first step was thus to extract from the YCOE all the words marked with the tags ADVR (adverb in the comparative) and ADVS (adverb in the superlative), together with the contextual information (text name and genre). In order to perform this task in a systematic and efficient way, the inflectional forms were extracted from the POS files with the text editor Notepad++ because this program process heavier files than other text editors. The extraction process begins with a preliminary search on each text to quantify the number of adverbs to be extracted per text. In the next step, a few formal adjustments were needed: sequences '+a', '+d', '+t' were respectively replaced with 'æ', 'ð', 'þ' by using the search and replace engine; both small and capital "RP+" and "$", sequences were replaced with nothing; single spaces were replaced with a paragraph mark, giving rise to a list of words arranged in a column; additionally, a single paragraph mark was used to replace double paragraph marks. The resulting list was sorted alphabetically and all the undesired results such as text codes, stops, semicolons, commas, codes, etc. were eliminated. This list was imported into the first column of an Excel file.

Each Excel page corresponds to a text. In each Excel page, the column containing the imported information was divided into two columns, one containing the inflected form and the other its corresponding tag. Next, two additional columns were added to indicate the name of the text and its genre so that the inflectional form can be easily identified if necessary. Table 1 below displays the information obtained after the extraction process distributed in four columns:

Form	Tag	Text	Genre
ærest	ADVS^T	Covinsal	PROSE
ærost	ADVS^T	Covinsal	PROSE
oftost	ADVS^T	Covinsal	PROSE
seldost	ADVS^T	coboeth.o.02	PROSE
selest	ADVS	COBENRUL	PROSE
teonlycost	ADVS	conicodA	PROSE
ytemest	ADVS^T	cogregdH.o23	PROSE
ytemest	ADVS^L	COBENRUL	PROSE

Table 1. Sample of extracted forms with the ADVS and ADVR tags and the contextual information

As observed in Table 1, the first column lists the inflectional forms extracted from the YCOE that present either comparative or superlative degree. The second column contains the corresponding morphological POS tag. In the table above, we observe that there are locative adverbs, represented by ADVS^L, and also temporal, as the tag ADVS^T indicates. This additional information is especially useful when analysing forms that coincide formally speaking but that may have two or even three adverbial functions. For instance, *ytemest* may have either a locative meaning denoting a specific distance or a temporal one referring to lasting or taking a great amount of time.

Once all the forms have been compiled in an Excel file, another adjustment is required, namely, the normalisation of 'þ' to 'ð', which is carried out with the help of the search engine. Finally, a thorough inspection of the resulting list is conducted to verify that no form has been lost.

As has been described in the previous paragraphs, the extraction of the forms has been mostly automatically performed, but some manual revision was necessary to guarantee a higher

exhaustivity in the analysable material. For the sake of greater clarity, the forms have been separated into two files, one containing the comparatives and the other the superlatives.

There are various ways to lemmatise: by category, by type or by token. In this pilot research, lemmas have been assigned by type. Each of the extracted types has been manually assigned a lemma from the *Nerthus* headword list. At this point, a fifth column (the leftmost one) needs to be incorporated in the Excel file that contains the lemma assigned in each case. Table 2 exemplifies this stage of the process:

Lemmas	Inflectional forms	Tag	Text	Genre
ǣr	*ærest*	ADVS^T	conicodE	PROSE
ǣr	*ærest*	ADVS	coorosiu.o2	PROSE
fullīce	*fullicor*	ADVR	cogregdH.o23	PROSE
fyrn	*firnor*	ADVR^T	cowulf.o34	PROSE
gearwe	*gearor*	ADVR	cocuraC	PROSE
inn	*innor*	ADVR^D	cogregdH.o23	PROSE
inn	*innor*	ADVR^L	colaece.o2	PROSE
lange	*læncg*	ADVR^T	coaelive	PROSE
nēah	*nyhst*	ADVS^L	coorosiu.o2	PROSE
oft	*oftust*	ADVS^T	coverhom	PROSE
rǣdlīce	*rædlicor*	ADVR	cocuraC	PROSE

Table 2: Lemma assignment process

In the first round of lemmatisation, almost eighty percent of the inflectional forms were assigned a lemma, whereas a twenty percent remained unlemmatised and therefore required deeper examination in order to find the adequate lemma. To that end, *The*

Dictionary of Old English heavily contributed to the disambiguation of forms and the identification of lemmas starting with letters A-I. For letters L-W, the online version of Bosworth and Toller's dictionary was most helpful. The next section will discuss the results of the lemmatisation in detail and will compare them with the information offered by the DOE and Seelig.

5.   Analysis of the Results

This section is divided into two: the first part offers a discussion of the results obtained after the lemma assignment process; in the second part these results are compared and validated with the help of the *Dictionary of Old English* for words starting with letters A-I, and with Seelig's work for words starting with letters L-W. The validation of the results is a key step in the lemmatisation as it allows for confirmation that the inflectional forms are gathered under the correct lemma and provides more insight into the normalisation processes that each lexicographical source has employed.

5.1.   Results

The total number of inflectional forms extracted from the YCOE is 2,692, which corresponds to adverbs inflected for the comparative and the superlative. These forms have been found in 96 different prose texts. Each inflectional form was subsequently assigned a lemma from the 1,755 adverbs that constitute the list of headwords provided by the database *Nerthus*. Although the extraction of forms was undertaken without distinguishing comparative from superlative forms, lemma assignment was performed separately. The vast majority of inflectional forms could be lemmatised. The help of lexicographical sources was essential in the disambiguation of roughly 20% of these forms. Only twelve forms represented doubtful cases that required deeper investigation in order to find the appropriate lemma. A total of 181 lemmas were mapped into the 2,692 inflectional forms. The following table shows token distribution per lemma.

Number of inflectional forms per lemma	1	1<=15	15<=100	100<
Number of lemmas	86	72	19	4

Table 3. Number of tokens per lemma

As observed in the table, the number of lemmas decreases as the number of inflectional forms per lemma increases. This means that only a few lemmas, four to be precise, have been assigned more than 100 inflectional forms, whereas the opposite picture is much more common. Almost half of the lemmas (86) have been assigned to an inflectional form that appears only once. Examples of such hapax legomena include *forhæfendlicust* (lemma *forhæfendlīce* 'continently'), *heardlicor* (lemma *heardlīce* 'harshly, severly' or *sceortlicost* (lemma *scortlīce* 'shortly, briefly, soon'). A total of seventy-two lemmas have been assigned to more than one and up to fifteen inflectional forms. Some examples of lemmas gathering two different inflected forms include *ændemest* and *endemest* (lemma *endemest* 'equally, likewise'), *unbeorhtor* and *unbyrhtor* (lemma *unbeorhte* 'not brightly'), *undeorest* and *undeoror* (lemma *undeore* 'cheaply'), among others. This data proves that 158 out of 181 lemmas are associated with a relatively low variety of inflectional forms, most of which are assigned to either one or two different forms. Nineteen lemmas have been associated with more than fifteen and less than one hundred tokens each. In this case, the inflectional forms gather under one same lemma present a considerable degree of spelling variation, giving rise to between seven and eleven different spellings, while the number of tokens per type remains low. For instance, the lemma *ēaðe* has been assigned to the graded forms *eað, eaþost, eaþust, eð, eðest, eþost, ið, iðesð, ieð, yð* and *yþest*. From this list, the forms with the highest number of occurrences are *eð* and *ieð*, adding up to ten and eleven respectively.

Finally, the lemmas *ær* 'before', *swīðe* 'very much, exceedingly', *leng* 'longer' and *bet* 'good' have an extensive range of spelling variation. The four lemmas have been assigned to more than one hundred occurrences. Lemmas *leng* and *bet* do not present as many alternative spellings as *ær* and *swiðe* do. To begin with, adverb *leng* has four distinct inflectional forms, namely *lencg, leng,*

lengc and *lenge*, the most frequent of which is *leng* with 110 occurrences. The lemma *bet*, in turn, has eight distinct inflectional forms, namely *best, bet, betesð, best, bet, betest, betst, bett*, being *bet* the form with the most occurrences (75). On the other hand, the adverbs *ær* and *swyðe* gather the highest number of inflectional forms. *Ær* has been assigned to the following comparative forms: *ær, æror, ærre, ærror* and *ærur*, and the following superlative forms: *ærœst, ærast, æresð, ærest, ærost, ærst, ærust, æryst, æst, arrest, erest* and *erost*. A quantitative approach enables the identification of the forms that are more widely spread in the literature of the period. For instance, the superlative form *ærest* gathers the highest number of tokens (600), which are distributed in sixty texts. Correspondingly, the lemma *swyðe* gathers seventeen inflectional forms. These forms are *swiðer, swiðor, swiður, swyðer, swyðere, swyðor* and *swyður* for the comparative and *swiðast, swiðest, swiðosð, swidost, swiðost, swiðusð, swiðust, swyðast, swyðost* and *swyðust* for the superlative. Of these, the most widely spread is the comparative form *swiðor*, with 275 tokens distributed throughout thirty-five texts.

The YCOE's morphological tags may further specify the type of adverb, namely directional, locative and temporal, which permits disambiguation in certain contexts. These overspecified tags have the following distribution among the inflected adverbs. Regarding the comparatives, 1,083 forms are assigned the underspecified ADVR tag, while the remaining 342 are divided into directional (16 tokens), locative (86 tokens) and temporal (240 tokens). As evinced by the figures, comparative adverbs with a temporal meaning constitute the group with the highest amount of occurrences and include the following forms: *ær, æror, ærre, ærror, ærur, firnor, hraðor, læng, længc, længe, lator, lencg, leng, lengc, ofter, oftor* and *seldor*. Concerning superlatives, a total of 477 forms are assigned the tag ADVS, whereas 762 have a temporal meaning (ADVS^T) and twenty-five a locative one (ADVS^L). Superlatives with a temporal meaning constitute the group with the highest amount of occurrences and include the following forms: *ærœst ærast, æresð, ærest, ærest, ærost, ærst, ærust, æryst, æst, arest, erest, erost, fyrmest, fyrmust, fyrst, længast, længest, længst, latost, lengest, lengst, nyhst, oftosð,oftost, oftust, seldost, siðestan* and *ytemest*.

5.2. Validation and Inferences

During the lemmatisation process not all the inflectional forms could be assigned a lemma from the *Nerthus* lemma list. As it will be explained afterwards, in some cases the list did not offer an adequate lemma, and in other cases the lack of formal transparency of the form hindered the process. The DOE and Seelig's work, together with additional lexicographical sources, have assisted this process.

A comparative study with other sources has made possible a more accurate view of the differences and similarities, in the way they analyse and organize information. Moreover, this comparative analysis served to validate the assignment of lemmas and justify the choice of one lemma over another, while also revealing that the inflectional forms provided by the YCOE, though fairly comprehensive, do not compile the entirety of comparative and superlative adverbial forms. These forms attested by DOE and Seelig but not attested by the YCOE are listed with their corresponding lemmas in Appendix 1. The DOE identifies a total of 186 and Seelig of 189 additional forms.

In order to give a broader picture of the process, two columns have been added to the previous table that describe which forms are lemmatised (or not) by the different sources This is illustrated in Table 4:

Inflectional form	Tag	Text code	Text genre	Seelig		DOE	
				Lemma	Inflected form	Lemma	Inflected form
Andgitfullicost	ADVS	Coprefcura	PROSE	✓	X	X	✓
Biorhtost	ADVS	Coverhom	PROSE	✓	X	✓	✓
Beorhtost	ADVS	Conicoda	PROSE	✓	✓	✓	✓
Beorhtre	ADVR	Comart3.O23	PROSE	✓	X	✓	X
Deoror	ADVR	Colaw6atr.O3	PROSE	✓	✓	R	✓
Ðristlicor	PROSE	Cootest	PROSE	X	X	X	X
Hatust	ADVS	Colacnu.O23	PROSE	✓	✓	✓	X

Table 4. Comparing the results of the lemmatisation with other sources

The additional columns integrated specify whether the lemma and the inflectional form are attested by each of the sources. The

symbol ✓ is used to indicate that a lemma or an inflectional form appears in these sources, preserves the same status and is formally alike; if not, the symbol X is used instead. For example, neither *andgitfullicost* nor *biorhtos*t and *beorhtre* are compiled by Seelig, while in the case of the DOE, *andgitfullicost* is assigned a lemma with a slightly different spelling, *andgytfullīce*. The lemma *beorhte*, on the other hand, is attested by both sources; the DOE also attests the two inflectional forms *biorhtost* and *beorhtost,* and Seelig attests only *beorhtost*; but neither source attests the inflectional form *beorhtre*. Something similar occurs with lemma *deore* and the inflectional form *deoror*; while the inflectional form has been attested by both Seelig and the DOE, the latter regards the form *deore* as an alternative spelling of headword *dyre*. This demonstrates that different criteria have been followed for the selection of headwords in *Nerthus* and the DOE.

The remainder of this section will give an overview of the distribution of comparative and superlative forms and will finally focus on doubtful cases. Starting with the comparative adverbs, a total of 1,425 were mapped into 136 lemmas. The DOE attested 85 distinct comparative forms (420 tokens), while Seelig compiled 99 types (785 tokens)[1]. A total of nine forms have been attested by neither source. These words are listed together with the lemma assigned from the list of headwords: *beorhtre* (*beorhte* 'brightly'), *eðost* (*ēaðe* 'easily'), *gearnlicor* (*geornlice* 'earnestly, diligently'), *geredelicor* (*gerǣdelīce* 'wisely, prudently'), *ðristelicor* (*ðrīstlīce* 'boldly'), *ðristlicor* (*ðrīstlīce* 'boldly, confidently'), *ðwyrlicost* (*ðwēorlīce* 'insolently').

When comparing the number of attested forms by DOE and Seelig, it must be born in mind that the DOE does not normalize 'ð' into 'þ'. The list of inflectional forms that this dictionary attests is, therefore, more exhaustive, as it distinguishes pairs of words such as *hraðe* and *hraþe, raðe* and *raþe*, *hrǣðe* and *hrǣþe* or *reþe* and *reðe*.

---

[1] Note that the DOE has for the moment only published letters A to I and thus lemmas starting for letters L to W cannot be attested by this source.

As for the superlative adverbs, they amount to 886 inflectional forms, which have been mapped into 76 lemmas. A total of 858 tokens (70 types) are also compiled by the DOE, while Seelig attests 768 tokens (36 types). Three superlative forms have not been attested by either source, though a lemma (in brackets) has been conveniently assigned to them. These forms are *eðost* (*eaðe* 'easily, lightly'), *gewissot* (*wise* 'wisely'), *ðwirlicost* (*ðweorlice* 'insolently').

The DOE has contributed enormously to the disambiguation of several comparative and superlative forms and to the subsequent assignment of a lemma. The following forms have been assigned a lemma from the DOE: *betest, betst, bet, bett, betest, betesð* (*bet* 'good'), *betere* (*betere* 'better'), *endenexð* (*endenēxt* 'last, final'), *firnor* (*fyrn* 'at first'), *fyrrest* (*feor* 'far, far away'). It can be observed that most of these forms present either vocalic change in the stem or the comparison has been created from a different stem.

The analysis has also revealed that there may be a mismatch between DOE's list of entries and *Nerthus'* list of lemmas. In most cases, this mismatch originates from the fact that one of the lemmas has been considered an alternative spelling of another. For instance, the DOE has assigned lemma *ārwurðlīce* to the inflectional form *arwurðlicor,* while *Nerthus* considers *ārwurðlīce* as an alternative spelling of the headword *ārweorðlīce* 'reverentially'. Other examples are *arwurðlicor* and *arwurðlicost*, which have been assigned the lemma *ārweorðlīce* 'honourably' from *Nerthus*, corresponding with DOE's lemma *ārwurðlīce; deoror* (*dēore* 'dearly'; DOE *dȳre*), *emnar* (*efne* 'even'; DOE *efne, emne*), *estelicor* (*estelice* 'corteously; luxuriously'; DOE *estlice*), *fægerost* (*fægre* 'fairly, elegantly', *fægere* DOE).

Other inflectional forms required closer inspection as their lexical category may not be adverbial. This is the case of the forms *leofost, liffest* and *liofast*. The fact that neither *Nerthus* nor Seelig offered a suitable lemma aroused suspicion, that is why it was deemed appropriate to verify their lexical status in context. The following citations correspond to the occurrences of these forms in the DOEC: *þonne hit wære leofost gehealden* (WHom 13 B2.3.1 [0004 (12)]) 'when it would most dearly be held'; *min bearn liffest gedoan* (Ch 1510 (Rob 6) B15.6.27 [0002 (4)]) 'my child has done

the quickest'; *swæ him liofast sie* (Ch 1510 (Rob 6) B15.6.27 [0004 (11)]) 'as it may best please them'. As evinced in the examples, they perform an adverbial function, however these are the only three occurrences that have been tagged by the YCOE as superlative adverbs, the rest being adjectives. This leads us to suggest that even if the function they fulfill is adverbial, they are adjectives. This hypothesis is supported by Bosworth and Toller's dictionary, which confirms their adjectival status.

Likewise, *endenexð*, which has been tagged as a superlative adverb by the YCOE, is actually an adjective. This conclusion was reached after verifying its status in different sources. Firstly, *Nerthus* does not provide any lemma that could be assigned to this form. In addition, the DOE considers this form an alternative spelling of the adjectival headword *endenēxt.*

Overall, this section has presented the quantitative and qualitative results of the lemmatisation of the Old English comparative and superlative adverbs by adopting the methodology presented in section 4.2. The contrastive analysis with DOE and Seelig has provided mutual feedback and has helped verify and refine the results of the analysis.

6.   Conclusions

This article has contributed to the design and implementation of a methodology for the lemmatisation of the Old English adverbs inflected in the comparative and superlative as attested by *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. If compared with the verbal lexicon, already lemmatised in previous works, adverbs present a substantially lower degree of variation and opacity, which has motivated a different methodology, more appropriate for this class.

The lemmatising methodology can be summarised in three main stages: firstly, the extraction of the 2,692 inflected adverbs from the YCOE through a fully automatic procedure; secondly, the lemmatisation of these forms through the manual assignment of a lemma from the list of headwords supplied by *Nerthus*; finally, the comparison of the results with a lexicographical and a secondary

source. This three-step procedure combines a selection of sources and of analytical methods that are yet far from being completely automatic.

The difficulties encountered during the process are in line with the presence of ambiguous forms that can be assigned to more than one lemma and, to a lesser extent, to forms that have been originally wrongly analysed and are not adverbs. Furthermore, this work has also identified all those inflected forms that were given by the DOE and Seelig but were not part of the YCOE's inventory, these make a total of 376 types that contribute to completing the adverbial paradigms. For this reason, additional sources have been consulted, including Old English grammars and dictionaries of reference in the language; in other cases, doubtful forms have been analysed in context to determine their meaning and function as attested in citations.

Considering the previous works and the study presented here, it is possible to make further advances in the lemmatisation of the pending major categories, including nouns, adjectives and non-graded adverbs. In addition, a work of these characteristics has direct implications in the field of corpus linguistics as it has proved the feasibility of lemmatising a historical corpus.

References

Bosworth, Joseph and Toller, Northcote T. 1973. (1898) *An Anglo-Saxon Dictionary*. Oxford: Oxford University Press.

Burkhanov, Igor. 1998. *Lexicography. A Dictionary of basic terminology*. Wydawn: Wyzszej Szkoly Pedagogicznej Rzeszowie.

Campbell, Alistair. 1959. *Old English Grammar*. Oxford: Oxford University Press.

Claridge, Claudia. 2008. "Historical corpora". *Corpus Linguistics. An international Handbook*. Eds. Anke Lüdeling, Merja Kytö and Tony McEnery. Berlin: Mouton de Gruyter: 242-254.

Faaβ, Gertrud. 2017. "Lexicography and corpus linguistics". *The Routledge Handbook of Lexicography*. Ed. Pedro A. Fuertes-Oliveira. Abingdon: Routledge: 123-425.

Fulk, Robert D. 2018. *A Comparative Grammar of the Early Germanic Languages.* Amsterdam/Philadelphia: John Benjamins Publishing Company: 237-240.

García Fernández, Laura. 2018. "Preterite-present verb lemmas from a corpus of Old English". *Verbs, Clauses and Constructions: Functional and Typological Approaches.* Eds. Pilar Guerrero Medina, Roberto Torre Alonso and Raquel Vea Escarza. Newcastle: Cambridge Scholars Publishing: 59-76.

García Fernández, Laura. 2019. *The lemmatisation of the verbal lexicon of Old English on a relational database. Preterite-present, contracted, anomalous and strong VII verbs*. PhD Dissertation, Department of Modern Languages: University of La Rioja.

Clark Hall, John. R. 1984 (1960). *A Concise Anglo-Saxon Dictionary. With a supplement by H. D. Meritt*. Toronto: University of Toronto Press.

Healey, diPaolo A. Price Wilkin, John and Xiang, Xin. 2009. *Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Healey, A. diPaolo et al. 2018. *Dictionary of Old English: A to I online.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

López-Couso, María José. 2016. "Corpora and online resources in English historical linguistics". *The Cambridge Handbook of English Historical Linguistics*. Eds Merja Kytö and Päivi Pahta. Cambridge: Cambridge University Press:127-145.

Martín Arista, F. Javier. ed, Laura García Fernández, Miguel Lacalle Palacios, Ana E. Ojanguren López and Esaul. Ruiz Narbona. 2016. *Nerthus V3. Online Lexical Database of Old English.* University of La Rioja

Metola Rodríguez, Darío. 2015. *Lemmatisation of Old English Strong Verbs on a Lexical Database.* PhD Dissertation, Department of Modern Languages, University of La Rioja.

Metola Rodríguez, Darío. 2017. "Strong Verb Lemmas from a Corpus of Old English. Advances and issues". *Revista de Lingüística y Lenguas Aplicadas,* 12: 65- 76.

Metola Rodríguez, Darío. 2018. "Corpus lemmatisation compatible with treebanks. Old English strong verbs". *Electronic Lexicography of Old English,* 1: 55-66.

Mitchell, Bruce and Fred C. Robinson. 1985 (1964). *A Guide to Old English.* Oxford: Blackwell.

Taylor, Ann, Anthony Warner, Susan Pintzuk and Frank Beths (eds.) 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose.* Department of Language and Linguistic Science, University of York.

Rissanen, Matti. 2008. "Corpus Linguistics and historical linguistics". *Corpus Linguistics. An international Handbook.* Eds. Lüdeling, Anke., Kytö, Merja and Tony McEnery. Berlin: Mouton de Gruyter: 53-68.

Seelig, Fritz. 1930. *Die Komparation der Adjektiva und Adverbien im Altenglischen.* Heidelberg: Winter.

Smith, Jeremy J. 2009. *Old English. A Linguistic Introduction.* Cambridge: Cambridge University Press.

Sweet, Henry. 1976 (1896). *The Student's Dictionary of Anglo-Saxon.* Cambridge: Cambridge University Press.

The York-Toronto-Helsinki Parsed Corpus of Old English Prose. 2003. Department of Language and Linguistic Science, University of York.

Tío Sáenz, Marta .2015. "Advances and Issues in the Lemmatisation of Old English Verbs on a Lexical Database". *Revista Electrónica de Lingüística Aplicada*, 14(1): 1-12.

Tío Sáenz, Marta .2019. "A semi-automatic lemmatisation procedure: Old English class 1 and class 2 weak verbs". *Electronic Lexicography of Old English.* 1: 67-76.

Appendix 1

Inflected forms attested by Seelig and DOE and not found in the YCOE[2]

A) Forms Attested by Seelig (1930) but not found in YCOE:

Inflectional Form	Lemma
æres (s)	ær
æris (s)	ær
ærist (s)	ær
andgifullicast (s)	andgitfullice
angsumest (s)	angsume
anlicost (s)	anlice
arlicor (c)	arlice
arwyrðlicost (s)	arodlice
atelicor (c)	atollice
baldlicor (c)	beadlice
baldlicost (c)	beadlice
beorhtast (s)	beorhte
bliðelicor (c)	blidelice
carfullicor (c)	carfullice
cuðlicost (s)	cudlice
cymlicor (c)	cudlice
cystiglicor (c)	cystiglice
þæslicor (c)	ðæslice

Inflectional Form	Lemma
þæslicost (s)	ðæslice
þearflicast (s)	ðearflice
þearlicor (c)	ðearllice
þearlicost (s)	ðearllice
deopost (s)	deope
deoplicost (s)	deoplice
deopplicor (c)	deoplice
deorost (s)	deore
deorlicost (s)	deorlice
drihtenlicor (c)	drihtenlice
þryðlicost (s)	ðryðlice
æð (c)	eaðe
eaðost (s)	eaðe
eaðusð (s)	eaðe
eaðust (s)	eaðe
ið (c)	eaðe
yðæst (s)	eaðe
yðast (s)	eaðe
yðost (s)	eaðe
fæstlicast (s)	fæstlice

---

[2] Next to the inflectional form it has been specified whether the form is comparative (c) or superlative (s).

fæstlicost (s)	fæstlice	gesundfullicost (s)	gesundfullice
feor (c)	feor	gewislicost (s)	gewislice
fior (c)	feor	gleawast (s)	gleawe
fyer (c)	feor	gleawlicor (c)	gleawlice
forðor (c)	forð	grimlicor (c)	grimlice
fracelicor (c)	forðlice	grimmost (s)	grimme
firmest (s)	fore	hadrost (s)	hadre
first (s)	fore	hatlicor (c)	hatlice
fyrmæst (s)	fore	hatlicur (c)	hatlice
forðlicor (c)	fracoðlice	heaor (c)	heah
fracedlicor (c)	freolice	hear (c)	heah
fromlicast (s)	freolice	hyhst (s)	heah
fromlicor (c)	fromlice	healicust (s)	healice
fullfremedlicor (c)	fullfremedlice	heardor (c)	hearde
fullicur (c)	fullice	hefgor (c)	hefige
gearwast (s)	gearwe	hludur (c)	hlude
gearwost (s)	gearwe	hlutrost (s)	hlutre
gehyðlicor (c)	gehyðelice	hraður (c)	hlutre
geliccast (s)	gelice	hwætlicor (c)	hraðe, hræðe
gelimplicur (c)	gelimplice	hwonlicost (s)	hwonlice
gelustfullicor (c)	gelustfullice	inlocast (s)	inlice
gemetlicost (s)	gemetlice	innemest (s)	inne
genihtsumlicor (c)	genihtsumlice	inweardlicor (c)	inweardlice
geornnost (s)	georne	inweardlicost (s)	inweardlice
geornust (s)	georne	laðlicost (s)	laðlice
gerisenlicor (c)	gerisenlice	læsast (s)	læs
gerisenlicost (s)	gerisenlice	læsest (s)	læs
gerisenlicur (c)	gerisenlice	læst (s)	læs
		lætlicor (c)	lætlice

alenge (c)	lange, longe
lengost (s)	lange, longe
leong (c)	lange, longe
hlætmeste (s)	late
hlætmesto (s)	late
lætmest (s)	late
lætmesta (s)	late
latest (s)	late
leohtor (c)	leohte
lihtluccost (s)	leohtlice
luflicor (c)	luflice
litelicost (s)	lytig-lytelice
lytelicost (s)	lytig-lytelice
mæstlicust (s)	ma
mærlicor (c)	mærlice
meahtelicor (c)	meahtelice, mihtlice
mihtlicor (c)	mildheortlice
anihst (s)	neah, neh
næar (c)	neah, neh
nycst (s)	neah, neh
nyr (c)	neah, neh
neoðer (c)	neodlice
neodlicor (c)	neodlice
nioðoror (c)	nið und niðer
niwlicor (c)	niwlice
niwlicost (s)	niwlice

norðmest (s)	norð
ofostlicor (c)	ofostice
oftast (s)	oft
orsorhlicur (c)	orsorglice
raðust (s)	raðe, hraðe
rædlicost (s)	rædlice
raðost (s)	ramlice
ramlicor (c)	ramlice
recenust (s)	recene
regollicor (c)	regollice
riclicost (s)	riclice
rihtast (s)	rihte
rihtlicast (s)	rihtlice
sarlicast (s)	sarlice
scerpest (s)	scearpe
scortlicor (c)	scortlice
sæl (c)	sel
selast (s)	sel
seolest (s)	sel
seldnor (c)	seld -
seft (c)	sið
siðor (c)	sið
siðor (c)	sið
softost (s)	sið
slawlicor (c)	slawlice
sniomor (c)	sneome,sniome
snotorlicor (c)	snotorlice
snotorlicost (s)	snotorlice
soðlicost (s)	sodlice
softost (s)	softe

styðlicor (c)	stidlice
strangor (c)	strange
stranglicost (s)	stranglice
stuntlicor (c)	stunlice
swetolor (c)	sweotole
swætolocor (c)	sweotollice
sweotolicor (c)	sweotollice
swiðast (s)	swiðe, swyðe
swiðust (s)	swiðe, swyðe
tearlicer (c)	teartlice
teartliclur (c)	teartlice
tidlicor (c)	tidlice
todæledlicor (c)	todæledlice
tolcendlicor (c)	tolcendlice
trumlicor (c)	trumlice
tylg (c)	tulge
tylgust (s)	tulge
ofor (c)	ufor
uferur (c)	ufor
yfemesð (s)	ufor
ungerædelicost (s)	ungerædelice
ungetæslicost (s)	ungetæslice
unswiðor (c)	unswiðe
waccor (c)	wace
waclicost (s)	waclice
weorðfulicor (c)	weorðfullice
weorðelicor (c)	weorðlice, wurðlice

weorðlicost (s)	weorðlice, wurðlice
wurðlicor (c)	weorðlice, wurðlice
gewidost (s)	wide
widor (c)	wide
widost (s)	wide
widre (c)	wide
wærse (c)	wiers
wyrrest (s)	wiers
wyrst (s)	wiers
wrætlicost (s)	wrætlice
wunderlicor (c)	wundorlice

B) Forms Attested by DOE but not found in YCOE.

Inflectional Form	Lemma
æðellice (c)	æðellice
æþellicor, (c)	æðellice
æðellicor, (c)	æðellice
æþelicor, (c)	æðellice
æðelucur (c)	æðellice
hær (c)	ǣr
ęr (c)	ǣr
ar (c)	ǣr
aar (c)	ǣr
ærrur (c)	ǣr
ærre (c)	ǣr
eror (c)	ǣr
awor (c)	ǣr
here (c)	ǣr
hærest (s)	ǣr
æresð (s)	ǣr
æres (s)	ǣr
ærets (s)	ǣr
ærist (s)	ǣr
aerist (s)	ǣr
æris (s)	ǣr
æryst (s)	ǣr
ærst (s)	ǣr
aerst (s)	ǣr
eræst (s)	ǣr
earest (s)	ǣr

andgietfullicost	andgietfullīce
arwurdlicor (c)	ārweorðlīce
arweorðlycor (c)	ārweorðlīce
baldlicor (c)	bealdlīce
bealdlicost (s)	bealdlīce
baldlicost (s)	bealdlīce
beortur (c)	beorhte
beorhtast (s)	beorhte
berrhtost (s)	beorhte
brihtlycor (c)	beorhtlīce
beorhtlicor (c)	beorhtlīce
best (s)	bet
bezt (s)	bet
bezte (s)	bet
bæst (s)	bet
beotost (s)	bet
cuþlicor (c)	cūðlīce
cyðlicor (c)	cūðlīce
cuðlicost (s)	cūðlīce
deopper, deoppur (c)	dēope
deowwor (c)	dēope
deopost (s)	dēope
deopplicor (c)	dēoplīce
deoplicur (c)	dēoplīce
deoplicost (s)	dēoplīce
deorost (s)	dēoplīce

derast (s)	dēore	eaðelucust (s)	ēaðelīce
domlicost (s)	dōmlīce	eaðelicost; (s)	ēaðelīce
ieðest (s)	ēaðe	eaðelicust (s)	ēaðelīce
yþæst (s)	ēaðe	eadmodlucor	ēaðmōdlīce
yþast (s)	ēaðe	ælmæst (s)	eallmǣst
yþust (s)	ēaðe	earfoþlicor	earfoðlīce
yþost (s)	ēaðe	efnast (s)	efne
eaðest (s)	ēaðe	efnost (s)	efne
eaþest (s)	ēaðe	egeleaslycor (c)	egelēaslīce
eþ (s)	ēaðe	ægeleaslycor (c)	egelēaslīce
iðesð (s)	ēaðe	egeleaslecor (c)	egelēaslīce
eþest (s)	ēaðe	ælcor (c)	elcor
yþ (s)	ēaðe	ylcor (c)	elcor
æð (s)	ēaðe	fægrost, (s)	ēstelīce
eaþ (s)	ēaðe	fægeror (c)	ēstelīce
eað (s)	ēaðe	fægror, (c)	ēstelīce
eaðor (c)	ēaðe	fægrur, (c)	fægre
eaþor (c)	ēaðe	færlicor (c)	fǣrlīce
eaðust (s)	ēaðe	festlycor (c)	fæstlice
eaður (c)	ēaðe	fæstlicost (s)	fæstlice
eaðr (c)	ēaðe	fæstlicas (s)	fæstlice
eðor (c)	ēaðe	fyrrer (c)	feorr
eaþelicor (c)	ēaðelīce	fyer (c)	feorr
eaþlicor (c)	ēaðelīce	ferrer (c)	feorr
eþelicor (c)	ēaðelīce	fær (c)	feorr
eþelicur (c)	ēaðelīce	fierr (c)	feorr
yþelicor (c)	ēaðelīce	fīr (c)	feorr
eþelicost (s)	ēaðelīce	furþor, (c)	forð
æðelicest (s)	ēaðelīce		

furþar, (c)	forð		gearnor (c)	georne
forþor, (c)	forð		geornæst (s)	georne
fyrþmest (s)	forð		georner, (c)	georne
furþur, (c)	forð		geornere (c)	georne
furþer (c)	forð		geornest, (s)	georne
forðor, (c)	forð		geornnost (s)	georne
forður (c)	forð		geornus (s)	georne
freolukeost (s)	frēolīce (ge)		geornfullicur, (c)	georne
frelukest (s)	frēolīce (ge)		geornfullicer (c)	georne
frelikest (s)	frēolīce (ge)		geornlicer (c)	geornlice
freolutust (s)	frēolīce (ge)		geornlicur (c)	geornlice
frelubest (s)	frēolīce (ge)		geornlecor (c)	geornlice
fullicur (c)	fullīce		geornlucor (c)	geornlice
fullecar (s)	fullīce		geornlycor (c)	geornlice
fyrmæst (s)	fyrmest		georlicor (c)	geornlice
fyrmost (s)	fyrmest		geornlucost (s)	geornlice
fyrmyst (s)	fyrmest		georlicost (s)	geornlice
fywmyst (s)	fyrmest		grimlicor. (c)	grimlīce
firmest (s)	fyrmest		hælicer (c)	hēalīce
fyrnor (c)	fyrn		healicust (s)	hēalīce
gearwost, (s)	gearwe		heardor (c)	hearde
gearwast (s)	gearwe		heardlicur (c)	heardlīce
gehændast (s)	gehende		hefilicor (c)	hefiglīce
gehændor (c)	gehende		hætelicor (c)	hetelice
gehendust (s)	gehende		hetolycor (c)	hetelice
geendost (s)	gehende		hiwcuþlicor, (c)	hīwcūðlīce
ihendost (s)	gehende		hiowcuðlucor (c)	hīwcūðlīce
gehændost (s)	gehende			
eornnost (s)	georne			

hludur (c)	hlūde	hrædlicur (c)	hrædlīce
hludast (s)	hlūde	rædlicer (c)	hrædlīce
hraþor (c)	hraðe	hærdlicor (c)	hrædlīce
raþor (c)	hraðe	rædlycor (c)	hrædlīce
hraður (c)	hraðe	rædlicost (s)	hrædlīce
raþur, (c)	hraðe	hreadlicost (s)	hrædlīce
hraðer (c)	hraðe	hwonlicer (c)	hwōnlīce
raþer (c)	hraðe	hwonlycor (c)	hwōnlīce
hræþor (c)	hraðe	hwonlicost (s)	hwōnlīce
hroþor (c)	hraðe	hwonlicest (s)	hwōnlīce
hroðor (c)	hraðe		