In Press

ADVANCES IN THE AUTOMATIC LEMMATIZATION OF OLD ENGLISH: CLASS V STRONG VERBS (L-Y)

Roberto Torre Alonso
Universidad de La Rioja

Abstract: The grammatical description of Old English lacks complete and systematic lemmatization, which hinders Natural Language Processing studies in this language, as they strongly rely on the existence of large, annotated corpora. Moreover, the inflectional features of Old English preclude token-based automatic lemmatization. Therefore, specifically goal-oriented applications must be developed to account for the automatic lemmatization of specific variable categories. This article designs an automatic lemmatizer within the framework of Morphological Generation to address the type-based lemmatization of Old English class V strong verbs (L-Y). The lemmatizer is implemented with rules that account for inflectional, derivational and morphophonological variation. The generated forms are compared with the most relevant corpora of Old English for validation before being assigned a lemma. The lemmatizer is successful in supplying form-lemma associations not yet accounted for in the literature, and in identifying mismatches and areas for manual revision.

Keywords: Old English; lemmatization; strong verb; Natural Language Generation; Morphological Generation.

1. AIMS, RELEVANCE AND SCOPE

This article takes issue with the automatic generation and lemmatization of inflectional forms for Old English (hereafter OE). More specifically, its aim is to design and implement an automatic lemmatizer of OE Class V strong verbs (L-Y) based on the Morphological Generation and assess its accuracy.

While there has been much philological and lexicographical discussion on the morphological development of OE, there have not been parallel advances in the computational study of the language. Apart from the digitized versions of the *Dictionary of Old English Corpus* –henceforth DOEC– (Healey et al., 2004) and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* –hereafter YCOE– (Taylor et al., 2003); the online versions of the *Dictionary of Old English* –henceforth DOE– (Healey et al., 2018) and of *An Anglo-Saxon Dictionary* (Tichy & Rocek, 2019); and the series of studies (Mateo Mendaza, 2016; Novo Urraca, 2016; Martín Arista & Ojanguren López, 2018; Vea Escarza, 2018; or Vázquez González & Barðdal, 2019) based on the knowledge base *The Grid* (Martín Arista, 2013), not many advances have been made in computer-based studies of OE, at a time when Natural Language Processing (NLP) approaches are becoming central to linguistic analysis.

Within NLP, Natural Language Generation (NLG) is the subfield of computational linguistics that deals with the the development of computer sytems that are able to generate understandable texts in human language is. These systems produce texts "from some underlying non-linguistic representation of information" (Reiter & Dale, 1997:1). For so doing, computers need to be trained and supplied with vast amounts of linguistic data. Such training is the domain of interest of Natural Language Understanding (NLU) which may have different approaches, including Conceptual Dependency (Schank, 1972); Naïve Semantics (Dahlgren, 1988); Commonsense Reasoning (Mueller, 2014) or neural networks (Liu et al., 2015).

There are several reasons that preclude NLU and NLG of OE. First, the dearth of textual material. The extant OE word stock is limited to the data stored in the DOEC, –3,000,000 words in approximately 3,000 texts. Partial corpora of OE include the *YCOE* –1,500,000 words–; the *Helsinki Corpus of English texts* (Rissanen et al., 2001) –300,000 words– and the *York-Helsinki Parsed Corpus of Old English Poetry* (Pintzuk & Plug, 2001) –70,000 words. Second, the lack of a lemmatization standard. The cited corpora are not lemmatized, and only the York-Helsinki corpora offer linguistic metadata, including morphological tagging and syntactic parsing. On its part, textual editions with glossaries and traditional dictionaries like Bosworth and Toller's (1973) *An Anglo-Saxon Dictionary*,

or Clark Hall's (1996) *A Concise Anglo-Saxon Dictionary* show variation and inconsistencies as regards lemma assignment, with spelling variation and cross-referencing standing out as the major issues. The DOE by Healey et al. (2018), offers a more systematic approach to lemma selection, and displays lists of attested forms associated to a lemma. However, at present, it has only been published up to the letter I. The only lemmatized corpus of OE published so far is Javier Martín Arista et al.'s (2021) *An open access annotated parallel corpus Old English-English*. This resource provides glosses, translation and lemma for a selection of OE texts. However, the published version only accounts for 110,000 word tokens. Third, the overlapping of spelling forms and the coalescence of several morphophonological evolutions which give way to formally ambiguous forms which limit the possibilities for machine training.

Notwithstanding these limitations, Hedderich et al. (2021) survey several models to account for NLP in contexts with low-resource settings as the one just described. Among the cited procedures they include Distant or Weak Supervision (Craven & Kumlien, 1999), which uses unlabelled text to which labels might be added. The corresponding labels are obtained through a (semi-)automatic process from an external source of information i.e. knowledge bases or dictionaries. While distant supervision has traditionally been used for Relation Extraction (RE) (Hoffman et al., 2011), it has also been used for automatic annotation, thus, Li et al. (2012) on Part-of-Speech (POS) tagging.

To the best of my knowledge, just a handful of works have been published dealing with the automatic lemmatization of OE and none has used Weak Supervision models. Metola Rodríguez (2015, 2017), Tío Sáenz (2019) and García Fernández (2020) have tackled the lemmatization of strong verbs, weak verbs, and anomalous and contracted verbs respectively. While differing in the methodological approaches due to the particular features of each verb class and in the scope of their research, these authors develop semi-automatic search systems to identify potential verb forms in the corpus. The retrieved data are manually supervised and compared with the existing dictionaries and corpora for validation and correction, thus contributing to the improvement of the search system and the increase of the degree of accuracy. Hamdoun Bghiyel (2020) has explored the lemmatization of the comparative and superlative degrees of the adverbs filed in the YCOE, while Novo Urraca and Ojanguren López (2018) have incorporated lemma assignment to the YCOE syntactic analysis. All these works are based upon the knowledge base *The Grid* (Martín Arista, 2013), which is formed by the lexical database *Nerthus*, a dictionary of Old English storing ca. 30,000 files; the dictionary database *Freya* (ca. 35,000 files), which provides information on secondary source indexing; the database of primary sources *Idunn*, which keeps at least one digitized file of each Old English text along with glossaries and translations; *Rindr*, an interface based on a concordance and an index of all the major corpora of Old English; and the lemmatizer *Norna* (ca. 190,000 files), based on an indexed concordance of the DOEC.

Against the background described, this article develops a Weak Supervision Model that can provide a label (lemma) for each of the types (abstract form) of the word tokens (textual occurrences) identified in either the DOEC or the YCOE. Automatic annotation may range from simple text string matching to the elaboration of complex pipelines. I shall follow the former approach in this research. However, given the current state of description of OE, there is no complete and reliable external source from which an exhaustive list of inflectional forms can be obtained. Thus, I have opted for developing a tool within the framework of Morphological Generation (MG) to create a set of word forms subject to being labelled with a Class V strong verb lemma. MG is "the task of producing the appropriate inflected form of a lemma in a given textual context and according to some morphological features" (Ferrés et al., 2017:110). Figure 1 shows an example of MG in OE. The lemma *þicgan* 'to take, receive', inflected for *person* (second), *number* (singular), *tense* (preterite), *mode* (indicative), with breaking of the stem vowel generates the inflected form *þeage* 'you took'.
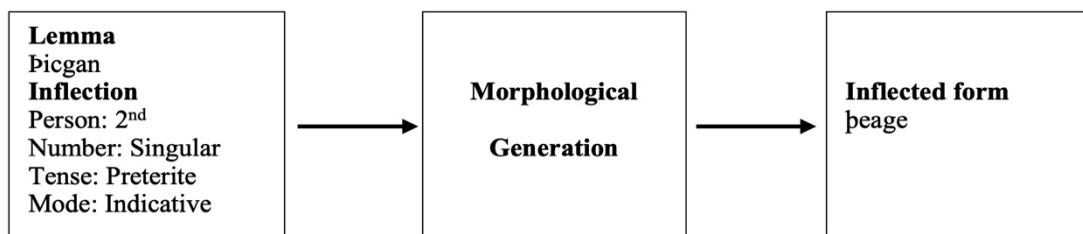

Figure 1. Morphological generation of the OE verb *þicgan*.

The MG Lemmatizer is able to (i) generate inflectional forms and apply morphophonological variations at inflectional ending and word stem levels; (ii) derive complex forms from simplex inflections; (iii) compare the generated forms with the selected corpora and (iv) assign lemma when a match is found.

The scope of the research is limited to Class V strong verb beginning with the letters L-Y excluding the contracted verbs therein included. The strong verb category has been chosen for its predominant position in the lexical creation of OE as Kastovsky (1992) and Martín Arista (2012), among others, remark. Class V has been selected for the *a priori* stable character of the vocalic changes that define it. The alphabetical limit is imposed by the current state of publication of the DOE.

The remainder of the article is organized as follows. Section 2 offers an overview of the OE strong verb system with an emphasis on Class V and lists the verbs selected for analysis. Section 3 describes the methodological steps followed in each of the stages of the lemmatization process. The results are offered in section 4 along with a discussion on the advances and limits of MG automatic lemmatization. To round off, Section 5 presents some conclusions and advances promising paths to be explored in future research.

2. THE OE STRONG VERB SYSTEM: AN OVERVIEW

The OE strong verb system is characterized by the existence of seven classes organized around the changes in the stem vowel in the different verb forms and the retention of inflectional endings. Vowel changes are grouped in four different grades, namely the infinitive grade for the present tense (vowel 1); the preterite 1 grade for the 1st and 3rd singular forms (vowel 2); the preterite 2 grade, for the other forms of the preterite (vowel 3) and the past participle grade (vowel 4). Each combination of these four grades constitutes an *ablaut* or *gradation* pattern. Figure 2 summarizes the traditional classification of OE verbs.

Class	Vowel 1	Vowel 2	Vowel 3	Vowel 4	Translation
I	*rīdan*	*rād*	*ridon*	*geriden*	'ride'
II	*crēopan*	*crēap*	*crupon*	*gecropen*	'creep'
IIIa	*bindan*	*band*	*bundon*	*gebunden*	'bind'
IIIb	*weorpan*	*wearp*	*wurpon*	*geworpen*	'cast'
IIIc	*helpan*	*healp*	*hulpon*	*geholpen*	'help'
IV	*stelan*	*stæl*	*stǣlon*	*gestolen*	'steal'
V	*sprecan*	*spræc*	*sprǣcon*	*gesprecen*	'speak'
VI	*faran*	*fōr*	*fōron*	*gefaren*	'travel'
VIIa	*hātan*	*hēt*	*hēton*	*Gehāten*	'be named'
VIIb	*healdan*	*hēold*	*hēoldon*	*gehealden*	'hold'

Figure 2. The traditional classification of OE strong verbs (adapted from von Mengden, 2011).

The first six classes evolve from the earlier Indo-European *ablaut* series. These apophonic changes conveyed morphological significance associated to particular stress patterns. Eventually, they came to be the only morphological marker when stress shifts ceased to be productive. The reconstructed Proto-Germanic class V *ablaut* system (Mailhammer, 2007) is presented in Figure 3.

Class	ablaut pattern	root	vowel 1	vowel 2	vowel3	vowel 4
V	e-a-ē-e	CVC	CeC	CaC	CēC	CuC

Figure 3. Proto-Germanic class V ablaut pattern (adapted from Mailhammer, 2007).

The seventh class is formed by a group of reduplicating verbs. These verbs duplicated part of their stem when inflected for perfective forms. When the duplicated syllable was lost in Northwest Germanic dialects, these verbs still displayed a stem vowel distribution that paralleled that of the *ablauting* verbs –a generalized *a/*e alternation– and eventually became part of the strong verb system (Adamczyk, 2002:26).

All in all, the traditional classification in seven classes of the OE strong system is widely accepted. In von Mengden's (2011:123-124) words:

> The strong verbs of Old English with their classification into seven classes and their categorisation according to 'ablaut vowels', have constituted one of the most established and, so it seems, least disputed sections of the grammatical description of Old English.

Levin (1964), Laing & Lass (2010) and von Mengden (2011) make claims against this historical classification and attempt to postulate a different representation. Von Mengden's (2011) approach aims at incorporating the *i*-mutated vowel of the 2nd and 3rd persons into the system under the label *vowel 1' or 5*, on the basis that such apophonic variation conveys morphological meaning. Figure 4 shows this model.

Class	1 (inf.)	1' or 5	2 (pret. 1)	3 (pret. 2)	4 (past part.)	Translation
V	*sprec-an*	*spric-þ*	*spræc*	*spræc-on*	*ge-sprec-en*	'speak'

Figure 4. von Mengden's (2011) class V strong verbs with apophonic variants.

Its part, Levin (1964) postulates a system based on a fully synchronic analysis which takes the vocalism of the preterit as the ruling criterion. With such premises, Levin (1964:59) establishes a 28-class system which rearranges verbs into a completely new classification. Thus, he establishes a fourth class with four different subclasses, as shown by Figure 5.

Class	Gradation	Vowel 1	Vowel 2	Vowel 3	Vowel 4	Translation
4a	e-æ-ǣ-o	*beran*	*bær*	*bǣron*	*(ge)boren*	'bear'
4b	e-æ-ǣ-e	*metan*	*mæt*	*mǣton*	*(ge)meten*	'measure'
4c	ēo-ea-ǣ-e	*sēon*	*seah*	*sǣgon*	*(ge)segen*	'see'
4d	i-æ-ǣ-e	*biddan*	*bæd*	*bǣdon*	*(ge)beden*	'pray'

Figure 5. Levin's (1964) class 4 strong verbs.

Whereas Levin's (1964) 4a class corresponds to the traditional class IV, classes 4b-d group those verbs that are historically considered class V. Thus, 4b groups verbs that follow the standard gradation; 4c the contracted verbs; and 4d those class V verbs with a modified infinitive vowel, which, furthermore, present weak present paradigms.

Laing and Lass (2010) argue against the standard gradation patterns in classes IV and V claiming that an infinitive vocalism *ēo* should be postulated in view of the data obtained from Kuhn's (1965) *Vespasian Psalter*. Such vocalism corresponds to the analogical extension of the vocalic mutation e>eo produced by *u*-mutation when *e* is followed by a back vowel, as in the present indicative plural form *sprecað > spreocað*.

For the purpose of this research, I shall stick to the historical classification for several reasons. First, the presence of *i*-mutated vocalisms does not rule out the existence of older, unmutated forms. Second, the distinctions put forward by Levin (1964) do not apply in this research. On the one hand, contracted verbs fall out of scope. On the other, once implemented with the lemma and the basic inflection, the MG lemmatizer follows a strict sequence of rules. This means that strong verbs with weak present or preterite paradigms are treated as fully strong verbs. This allows potential research on the development of weak forms on a contextual basis. Third, the set of rules developed for generating vocalic and consonantal mutations (Section 3.2) include the analogical extension of *u*-mutation to the present system. Therefore, the generation of mutated and non-mutated inflectional forms maximizes data retrieval and does not restrict the findings to one specific dialect or text.

As it stands, following the premises stated in Section 1, which include limiting the scope of the research to the letters not yet published by the DOE and the exclusion of contracted verbs, the following verbs have been chosen from Martín Arista et al.'s (2021) list of lemmas: *lesan* 'gather', *licgan* 'lie', *metan* 'mete', *nesan* 'escape', *recan* 'proceed hastily', *repan* 'reap', *screpan* 'scrape', *sittan* 'sit', *sprecan* 'speak', *stecan* 'stick', *swefan* 'sleep', *tredan* 'tread', *þicgan* 'take', *wefan* 'weave', *wegan* 'carry' and *wrecan* 'drive'.

3. METHODOLOGICAL PROCEDURE

This section describes the different steps taken to develop the specific set of MG rules to be implemented in the lemmatizer with the following organization. 3.1 describes the MG rules addressed at inflectional endings. 3.2 addresses the rules accounting for word-internal mutations. 3.3 tackles the creation of derived forms. To round off 3.4 describes the automation of the attestation of forms.

3.1 MG of class V verbs: Inflection

The selected verbs described in section 2 are inflected for infinitive, present indicative (person and number), preterite indicative (person and number), present subjunctive (number), preterite subjunctive (number), inflected infinitive, present participle, past participle and imperative (number). These paradigms are reconstructed on the

basis of well-attested inflectional patterns, but may include non-attested forms. Further, these initial paradigms do not account for morphophonological processes of vocalic alternation, syncopation, assimilation or simplification of consonant groups. They constitute, however, an initial schema upon which MG rules can operate to generate forms showing diachronic modifications, including assimilations and weakening of consonants and weakening and elision of vowels in unaccented position, which affect inflectional endings, and consonant changes resulting from the application of Verner's rule, which affect the stem. The inflection of *sittan* 'to sit' is illustrated in (1).

(1)

Infinitive	*sittan*	Pres. subj. (sg.)	*sitte*
Inflected Infinitive	*sittenne*	Pres. subj. (pl.)	*sitten*
Pres. ind. (1st sg)	*sitte*	Pret. subj. (sg.)	*sætte*
Pres. ind. (2nd sg.)	*sittest*	Pret. subj. (pl.)	*sætten*
Pres. ind. (3rd sg)	*sitteþ/sitteð*	Pres. part.	*sittend*
Pres. ind. (pl.)	*sittaþ/sittað*	Past part.	*setten*
Pret. ind. (1st/3rd sg.)	*sitt*	Imperative (sg.)	*sitt*
Pret. ind. (2nd sg.)	*sætte*	Imperative (pl.)	*sittaþ/sittað*
Pret. ind. (pl.)	*sætton*		

Two aspects deserve comment here. First, the limited inflectional system of OE strong verbs gives rise to the formation of formally ambiguous forms within the paradigm. As the goal of this article is to tackle type-base lemmatization, the morphological tagging of the inflected forms is of little importance. Duplicated forms will be deleted in a later stage (see Section 3.3). Second, forms containing the spelling <þ> are also inflected with the spelling <ð> to maximise their attestation. Thus, the pair *sittaþ, sittað*.

From the paradigm given in (1), alternating and mutated forms are generated. In a first step, rules are designed to account for diachronic and diatopic variation of inflectional endings. To codify these rules, I draw on Campbell (1987:299-300), whose account of the most common changes is given in (2).

(2)
a. -e >-æ/-o/-u

b. -est > -ist
    -est > -st
    -dst > tst > -st
    -þst > -sst > -st
    -þs > -ts
    -ngst > -ncst
    -gst > -hst > -xt

c. -eþ > -iþ
    -eþ > -þ
    -eþ/-iþ > -et/-it
    -et/-it > t
    -tþ/-dþ > tt
    -tt > -t
    -sþ > -st
    -gþ > hþ
    -ngþ > -ncþ
    -þþ > t

d. -on >-an > -un

e. -enne > -anne > -onne

f. -end > -and > -ind

g. -en > -in > -æn

(2a, b, c) display changes in the present indicative 1st, 2nd, and 3rd person singular, respectively; (2d) shows changes in the preterite indicative plural form; (2e) exemplifies changes in the inflected infinitive ending while (2f) shows changes in the present participle. Finally, (2g) accounts for changes in the past participle ending.

The non-finite forms, that is, the infinitive, the present participle and the past participle of the OE verb paradigm can be inflected. The inflected infinitive has been accounted for in (1). To generate the inflected forms of the participles, the endings *-ø*; *-ne*; *-es*; *-um*; *-e*; *-ra*; *-u*; *-re*; *-a*; *-an*; *-ena* have been attached to the basic forms, following Campbel (1987: 266-272).

3.2 MG of class V verbs: Mutation

This subsection presents the rules designed to account for the mutated forms which deviate from those given in (2). These rules include the generation of *i*-mutated forms in the 2nd and 3rd persons singular of the present indicative, and other changes caused by the phonological environments as *u*-mutation, breaking, analogical extensions and the effects of Verner's law. Empirical evidence supporting these rules comes from Campbell (1987:313), Krygier (1994:52) and Laing and Lass (2010:146). Figure 6 summarizes the rules implemented in this study.

Rule No.	Rule Description
Rule #1	*-e-* > *-i-* in 2nd and 3rd person singular (*i*-mutation)
Rule #2	*-i-* > *-y-* in 2nd and 3rd person singular
Rule #3	*-e-* > *-ie-* > -after initial *g-*
Rule #4	*-e-* > *-eo-* in present indicative (*u*-mutation)
Rule #5	*-eo-* > *-ea-* - in present indicative
Rule #6	Extension of *-eo-* to the present system
Rule #7	Extension of *-ea-* to the present system
Rule #8	*-e-* > *-æ-* in the present system
Rule #9	*-e-* > *-oe-* after -w-
Rule #10	*-æ-* > *-e-* in preterit system
Rule #11	*-æ-* > *-ea-* in preterit forms (*breaking*)
Rule #12	*-æ-* > *-a-* in preterit forms before /G/-/x/
Rule #13	*-g-* > *-h-* in final position
Rule #14	*-e-* > *-o-* in past participle
Rule #15	*-e-* > *-oe* in past participle
Rule #16	*-þ-* > *-d-* (Verner's law)
Rule #17	*-s-* > *-r-* in pret. Pl. and past part. (Verner's law)
Rule #18	spr-> sp-

Figure 6. Rule implementation for class V stem mutations.

While rules #3 and #16 apply to other class V verbs, they do not apply to any of the verbs selected for this research. Furthermore, according to Krygier (1994:52) and Campbell (1987:313) the rhotacism generated by Verner's law is not operative in *fnesan* 'breathe hard', *lesan* 'gather', *nesan* 'escape', which extend the spirant sound to the whole paradigm. Nevertheless, Rule #17 has been generated by analogy with other verb classes – thus, class II *ceosan-ceas-curon-coren* 'to choose'– to maximize data retrieval.

3.3 MG of class V verbs: Derivation

After the implementation of the inflectional rules that generate the inflectional forms of the simplex verbs, I shall turn to the description of the process followed to generate their morphologically complex counterparts. For so doing, preverbal elements need to be attached to the simplex forms. The collection of preverbal items includes those elements L-Y described in Metola Rodríguez (2015) and García Fernández (2020). To these, I add the prefix *ge-* which, despite falling out of the alphabetical scope of this research, participates in the formation of participial forms. Given that the selected preverbal elements are subject to spelling variation and that the research is type-based, they have been arranged in canonical (lemma) and non-canonical forms. When at stake, the grapheme <þ> has been chosen as canonical. (3) lists the selected preverbal lemmas (in bold) along with their alternative spellings.

(3)

ge-(*cg-*, *g-*, *ga-*, *gæ-*, *gæn-*, *gær-*, *gad-*, *gan-*, *gar-*, *ged-*, *gen-*, *gem-*, *ger-*, *gi-*, *gif-*, *gim-*, *gy-*); ***med-***(*me-*, *met-*, *mi-*, *mid-*, *miŏ-*, *miþ-*, *mod-*); ***mis-***(*miss-*, *mus-*); ***niþer-***(*neoþer-*, *nioþer-*, *nyþer-*, *nieþer-*, *niþor-*, *niŏer neoŏer-*, *nioŏer-*, *nyŏer-*, *nieŏer-*, *niŏor-*); ***o-; of-***(*æf-*, *af-*, *off-*); ***ofer-***(*eofer-*, *eofor-*, *ofær-*, *ofern-*, *ofor-*, *of'-*, *ofyr-*, *ouer-*, *ouyr-*); ***on-***; ***or-***; ***oþ-***(*oeþ-*, *oŏ-*, *oeŏ-*); ***onweg-***(*anweg-*, *aweg-*, *unweg-*); ***riht-***(*reht-*, *reoht-*, *rieht-*, *ryht-*); ***sam-***; ***sin-***; ***sub-***; ***to-***; ***twi-***(*twig-*, *twy-*); ***þri-***(*þry-*, *þrie-*, *ŏri-*, *ŏry-*, *ŏrie-*); ***þurh-***(*þorh-*, *ŏurh-*, *ŏorh-*); ***un-***; ***under-***(*und-*, *undern-*,

ynder-); ***up-***(*upp-*); ***ut-***(*utt-*, *vt-*); ***uþ-***(*uð-*); ***wan-***; ***wiþ-***(*wið-*); ***wiþer-***(*wiþere-*, *wiþyr-*, *wiðer-*, *wiðere-*, *wiðyr-*); ***ymb-***(*ym-*, *ymbe-*, *emb-*, *embe-*, *eme-*, *imb-*)

Once all the forms have been generated, they are searched for duplicated generations. Whenever several formally ambiguous generations share the same lemma, instances are reduced to just one occurrence. However, if two or more identical forms are generated in different paradigms, they are kept, so that the assignment of competing lemmas is guaranteed. Disambiguation would come from contextual analysis if the generated form was attested in the corpora. Compare the cases in (4).

(4)
nesan > *nesað* (pres. ind. pl.); *nesan* > *nesað* (imp. pl)
miswefan > *misswefend* (pres. part.); *misswefan* > *misswefend* (pres. part.)

The forms in 6a are simplified to just one occurrence, while the two instances of *miswefend* in 6b are maintained, associated to the lemmas *miswefan* and misswefan respectively.

3.4 Automatic attestation of generated forms

The final step involves the attestation of the generated forms in the selected corpora. This is done on an automatic basis through the comparison of the forms provided by the lemmatizer with an indexed version of both corpora. The index of the DOEC (Healey et al., 2004) has been obtained from the concorded version of the corpus filed in the knowledge base *The Grid* (Martín Arista, 2013). As for the YCOE (Taylor et al., 2003), those forms having a verbal POS –part of speech– tag have been extracted (see Appendix 1 for an exhaustive listing of POS labels and their meaning).

To summarise, three lists of words have now been compiled; the MG set, the DOEC index and the YCOE group of verbal forms. Each of these sets is filed in a separate database with different field structures. The MG database includes a field for the generated form (*Inflectional form*), a field for the lemma from which the form has been generated (*Class V Lemma*), a field to check attestation in the DOEC (*DOEC attestation*), a field to check attestation in the YCOE (*YCOE attestation*), and a field for the YCOE POS (*YCOE_verb_tags*) if the form is attested in the YCOE. The DOEC database displays a field for the indexed form in the DOEC (*ConcTerm*), a field for the text before the concorded term (*Prefield*), and a field for the text following the concorded term (*Postfield*). The YCOE database shows a field for the inflectional form in the YCOE (*YCOE_verbal_form*) and up to five fields for the POS tags (*YCOE_verbal_tag1…5*). Figure 7 offers an overview of the three databases.

Inflectional form	licgan			Class V Lemma	licgan
DOEC_attestation	☒ YES ☐ NO				
YCOE_attestation	☒ YES ☐ NO				
YCOE_Verb_Tags	VB VBPS				

Prefield	Conc_Term	Postfield
beam þe bið on ðinum eagan? Ne mæg se langa beam	licgan	on þinum eagan, ac se beam getacnað þa teonfullan
gecyrran to Egipta lande, þæt we on ðam lande ne	licgan	ofslagene, ne ure wif and cild ne wurðon gehergod
e eard þe eow swa mislicað, and eowre lic sceolon	licgan	bæftan him on ðysum westene for eower unrihtwisny
i inn ænne his burðena. And se afunde his hlaford	licgan	heafodleasne and he þa mid wanunge wende ut
et hine arisan and hine gereordian. He geseah þær	licgan	ænne snawhwitne focan, and an lytel fæt ful wæter
wura næs forod; and he forlet hi sona swa samcuce	licgan	, forþamþe witan cwædon þæt nan cwellere ne sceold
ewylde swa eaðelice. Hwæt wille we leng don buton	licgan	ealle æt his arwurðum cneowum, and eadmodlice bid

YCOE_verbal…	YCOE_verbal_tag 1	YCOE_verbal_tag 2	YCOE_verbal_tag 3	YCOE_verbal_tag 4	YCOE_verbal_tag 5
licgan	VB	VBPS			
licganne	VB^D				
licgað	VBPI				
licgaþ	VBPI				
licge	VBP	VBPS			
licgean	VB				
licgeað	VBPI				
licgendan	VAG^D	VAG^G	VAG^N		

Figure 7. A view of the form *licgan* in the MG, DOEC and YCOE databases.

The fields *Inflectional form, ConcTerm,* and *YCOE_verbal_form*, are related to one another, thus accounting for the automatic attestation of the generated forms. If there is a spelling coincidence between the generated

form filed in *Inflectional form* and a corpus occurrences filed either in *ConcTerm* and/or in *YCOE_verbal_form*, the corresponding *DOEC_attestation* and/or *YCOE_attestation* fields (YES) are activated. Figure 8 shows these relationships.
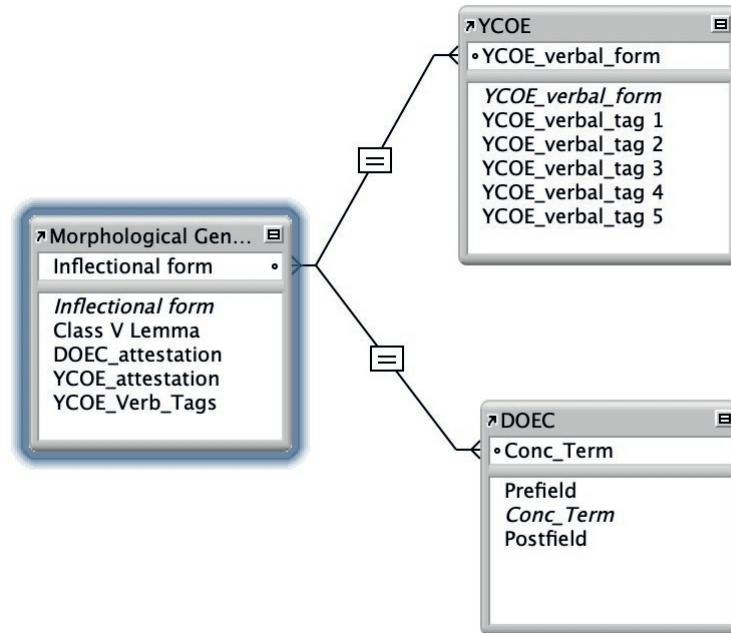

Figure 8. Relational interface of the MG, DOEC, and YCOE databases.

Section 4 below offers a general view of the results obtained and discusses the advances and limitations of the proposed method.

4. RESULTS AND DISCUSSION

The application of the methodological procedure described above provides the following results. The lemmatizer generates some 703,873 forms of which 1,098 are attested in the corpora. These forms belong in 119 lemmas. By corpus, the data shows the following distribution: 575 forms are attested only in the DOEC, corresponding to 106 different lemmas; 514 forms are attested in both corpora assigned 63 distinct lemmas; finally, 9 forms belonging in 6 lemmas have been attested only in the YCOE. Table 1 summarizes this data.

Table 1. Attested forms and assigned lemmas in the corpora.

Corpus	Attestations	Lemmas
DOEC	575 (52.36%)	106
DOEC & YCOE	514 (46.81)	63
YCOE	9 (0.81%)	6
Total	1,098	119*

*The figure shows different lemmas only.

As seen in Table 1, more than half of the attestations are found in the DOEC alone, which calls for further analysis to confirm the verbal nature of each occurrence of the attested form. However such a token-based analysis lies out of the scope of this work.

The complete list of the attested inflected forms by corpus and lemma is given in Appendix 2. The list of lemmas assigned by the lemmatizer is offered in (5).

(5)

gelesan, gelicgan, gemetan, genesan, gerecan, gerepan, gescrepan, gesittan, gesprecan, getredan, geþicgan, gewefan, gewegan, gewrecan, lesan, licgan, medlesan, medmetan, mednesan, medrecan, medsittan, medsprecan, medstecan, medwegan, metan, mislesan, misrecan, missprecan, misþicgan, nesan, oferlicgan, ofermetan, ofersittan, ofersprecan, ofertredan, oferþicgan, oferwegan, oferwrecan, oflesan, oflicgan, ofnesan, ofsittan, oftredan, olesan, onesan, onlesan, onlicgan, onsittan, onsprecan, onþicgan, onwegan, onwrecan, orepan, orlesan, orlicgan, ormetan, ornesan, orwegan, osittan, owefan, oþþicgan, recan, repan, rihtlicgan, rihtnesan, rihtrecan, rihtwegan, screpan, sinnesan, sintredan, sittan, sprecan, swefan, tolesan, tolicgan, tometan, tosittan, tosprecan, towegan, towrecan, tredan, twisprecan, twiwegan, þicgan, þrinesan, þurhwrecan, underlesan, underlicgan, undermetan, undernesan, undersittan, unlesan, unlicgan, unmetan, unnesan, unsprecan, unwegan, unwrecan, uplicgan, uprecan, upwegan, utlesan, utlicgan, utrecan, utsittan, uþmetan, wanwegan, wefan, wegan, wiþermetan, wiþersprecan, wiþlicgan, wiþmetan, wiþsittan, wiþsprecan, wrecan, ymblicgan, ymbsittan and *ymbsprecan.*

Table 2 displays a quantitative account of the inflectional forms assigned to each of the lemmas in the different corpora.

Table 2. Attested forms per lemma and corpus.

Lemma	DOEC	DOEC & YCOE	YCOE	Total	Lemma	DOEC	DOEC & YCOE	YCOE	Total
gelesan	11	5	0	16	*recan*	11	20	1	32
gelicgan	5	9	0	14	*repan*	10	4	0	14
gemetan	12	17	0	29	*rihtlicgan*	2	0	0	2
genesan	8	5	1	14	*rihtnesan*	1	0	0	1
gerecan	12	15	0	27	*rihtrecan*	1	0	0	1
gerepan	4	3	0	7	*rihtwegan*	1	0	0	1
gescrepan	2	0	0	2	*screpan*	3	1	0	4
gesittan	27	26	0	53	*sinnesan*	1	0	0	1
gesprecan	22	14	0	36	*sintredan*	1	0	0	1
getredan	3	1	0	4	*sittan*	20	33	0	53
geþicgan	1	11	2	14	*sprecan*	26	64	0	90
gewefan	6	2	0	8	*swefan*	5	9	0	14
gewegan	8	6	0	14	*tolesan*	8	4	0	12
gewrecan	2	12	0	14	*tolicgan*	3	3	0	6
lesan	21	11	0	32	*tometan*	1	0	0	1
licgan	14	27	0	41	*tosittan*	6	1	0	7
medlesan	3	0	0	3	*tosprecan*	5	3	0	8
medmetan	1	0	0	1	*towegan*	0	1	0	1
mednesan	3	0	0	3	*towrecan*	0	1	0	1
medrecan	1	0	0	1	*tredan*	17	12	0	29
medsittan	1	1	0	2	*twisprecan*	6	0	0	6
medsprecan	2	0	0	2	*twiwegan*	2	0	0	2
medstecan	0	1	0	1	*þicgan*	28	19	2	49
medwegan	2	0	0	2	*þrinesan*	4	0	0	4
metan	20	16	0	36	*þurhwrecan*	0	2	2	4
mislesan	0	1	0	1	*underlesan*	1	0	0	1
misrecan	0	2	0	2	*underlicgan*	2	3	0	5
missprecan	1	0	0	1	*undermetan*	2	0	0	2
misþicgan	0	1	0	1	*undernesan*	1	0	0	1
nesan	15	8	0	23	*undersittan*	2	0	0	2
oferlicgan	1	3	0	4	*unlesan*	1	1	0	2
ofermetan	6	0	0	6	*unlicgan*	3	0	0	3
ofersittan	4	8	0	12	*unmetan*	4	0	0	4
ofersprecan	4	0	0	4	*unnesan*	1	0	0	1
ofertredan	3	1	0	4	*unsprecan*	3	0	0	3
oferþicgan	0	3	1	4	*unwegan*	1	0	0	1
oferwegan	3	0	0	3	*unwrecan*	1	0	0	1
oferwrecan	2	0	0	2	*uplicgan*	2	0	0	2
oflesan	0	1	0	1	*uprecan*	1	0	0	1
oflicgan	2	3	0	5	*upwegan*	2	1	0	3
ofnesan	2	0	0	2	*utlesan*	1	0	0	1
ofsittan	5	14	0	19	*utlicgan*	4	0	0	4
oftredan	2	5	0	7	*utrecan*	0	1	0	1

Lemma	DOEC	DOEC & YCOE	YCOE	Total
olesan	1	0	0	1
onesan	1	0	0	1
onlesan	4	3	0	7
onlicgan	8	4	0	12
onsittan	9	11	0	20
onsprecan	5	1	0	6
onþicgan	0	3	0	3
onwegan	3	0	0	3
onwrecan	0	1	0	1
orepan	1	0	0	1
orlesan	1	0	0	1
orlicgan	2	0	0	2
ormetan	8	0	0	8
ornesan	1	0	0	1
orwegan	1	0	0	1
ositton	1	0	0	1
owefan	1	0	0	1

Lemma	DOEC	DOEC & YCOE	YCOE	Total
utsittan	1	0	0	1
uþmetan	1	0	0	1
wanwegan	2	0	0	2
wefan	13	5	0	18
wegan	24	17	0	41
wiþermetan	3	0	0	3
wiþersprecan	2	0	0	2
wiþlicgan	1	1	0	2
wiþmetan	6	5	0	11
wiþsittan	3	1	0	4
wiþsprecan	4	5	0	9
wrecan	17	19	0	36
ymblicgan	1	1	0	2
ymbsittan	10	20	0	30
ymbsprecan	5	2	0	7
Total	575	514	9	1,098

As can be seen in Table 2, the lemmas with the biggest number of attested forms are *sprecan* 'speak' (90 forms), *gesittan* 'sit' (53 forms) and *þicgan* 'take, receive' (49 forms). By corpus, the same verbs show the highest form-lemma ratio in those occurrences only attested in the DOEC, with *þicgan* displaying 28 forms, *gesittan* 27, and *sprecan* 26. As regards the forms attested in both the DOEC and the YCOE, *sprecan* outstands clearly above the rest with 64 inflectional forms, followed by *sittan* 'sit' and *licgan* 'lie' with 33 and 27 forms, respectively. As for those forms identified only in the YCOE, *þicgan* and *geþicgan*, both meaning, 'take, receive' show 2 occurrences each.

On the qualitative side, each of the 1,098 forms have been assigned a distinct class V lemma, which implies that the lemmatizer has a 100% accuracy as regards form-lemma association. As was shown in (6), homographic forms have been generated from the inflection of *swefan* 'sleep' and *wefan* 'device' and the attachment of the different forms of the prefixes *med-* and *mis-*. Nevertheless, none of these formally ambiguous forms has been attested in the corpora and consequently, no cases of lemma competition have arisen. Although promising, this data must be handled with care, given the scope of the research, limited to a group of verbs within a single class. It will be necessary to check the attested forms with forms generated in other strong verb classes. In the remainder of this section, I shall discuss the accuracy of the lemmatizer as regards both correct lemma assignment and the attestations of forms.

As for the former topic, the spelling inconsistency of OE may result in an increase of homographs like the ones just described. The overlapping can arise both intercategorially and intracategorially. In such cases, an automatic type-based lemmatizing process is not enough, and contextual, token-based lemmatization is needed for disambiguation. Consider the cases in (6), given with the automatically assigned lemma and potentially competing lemmas.

(6)
gales (gelesan ~ gal)
næs (nesan ~ nesan)

The form in (6a) may correspond, in principle, to an imperative form of the derived verb *gelesan* or to the singular genitive of the neuter noun *gal* 'lust, folly'. Two occurrences are attested in the corpus, presented in (7).

(7)

a. [ChristA,B,C 029000 (1032)]
*Hafað eall on him þæs þe he on foldan in fyrndagum, godes oþþe **gales**, on his gæste gehlod, geara gongum, hafað ætgædre bu, lic ond sawle.*
They shall have all upon them, which they once weighted upon their soul in bygone days, all of the good and the **folly**, over the course of the year—they shall hold both together, body and soul. (Hostetter, n.d.a)

b. [Rec 10.6.2 (Dickins-Earle) 000400 (1.8)]
*Osbern Hod Pilegrim Ialebriht Gesfrei se coc & Pierres se niulier Ailric, & **Gales**.*
Osbern Hod pilgrim; Ialebriht Gesfrei, the cook; & Pierres, the wafer-baker; Ailric; & Gales (Dickins 1950:367).

As can be seen in (7) none of the attested forms corresponds to the lemma *gelesan* assigned by the lemmatizer. (7a) shows the expected genitive of *gal* 'lust, folly', while (7b) corresponds to a proper name listed in an account of witnesses. Similarly, the form *næs* might be a 1st/3rd person singular of the preterit of *nesan* 'be saved from' or of the contracted negative verb *nesan* (*ne wesan*) 'not be'. There are 910 occurrences of this form in the DOEC that must be contextually disambiguated. The YCOE only provides the labels (BED) and (BEDI), which means that all the attestations in that corpus correspond to the contracted verb. Although the process of disambiguation falls out of scope here, it is worth mentioning at this point that, even if none of the occurrences corresponds to the proposed lemma, the inflection of the form *næs* proves necessary to generate the inflected form *genæs* for the lemma *genesan* 'be saved from' attested in (8).

(8)
[GenA,B 062100 (2018)]
*Him þa secg hraðe gewat siðian, an gara laf, se ða guðe **genæs**, Abraham secan.*
Then a man, one survivor of the spear, **escaped** from the battle, departed journeying hastily and seeking Abraham (Hostetter, n.d.b)

Regarding the identification of forms, let us compare the attested forms provided by the lemmatizer with a review of secondary sources indexing OE forms. The exhaustive analysis of the secondary sources stored in the database *Freya* retrieve the following fifty-nine forms under the lemma *sprecan* 'speak': *spæc, spæcan, spæcð, spæcon, spec, specað, specan, specð, spece, specende, specenne, specon, specst, spræc, spræcan, spræce, spræcen, spræcon, sprec, sprecað, sprecan, sprecanne, sprecaþ, sprecð, sprece, spreceð, sprecen, sprecende, sprecene, sprecenne, spreceþ, sprecon, sprecst, sprecu, sprecun, sprecþ, spreocað, spreocan, spreocende, spreocendra, spreocu, spricð, spriceð, spricest, spricþ, sprycð, sprycst, spycð, spæken, spæky, specce, speke, speken, spekinde, sprace, spracon, spreiced, sprice,* and *spricst.* Against this background, the lemmatizer generates and assesses the identification of the following ninety forms: *spæc, spæcan, spæcð, spæce, spæcen, spæcende, spæcenne, spæcon, spec, specað, specæn, specan, specaþ, specð, spece, speceð, specen, specende, specene, specenne, specon, specst, specu, spicð, spicþ, spræc, spræcað, spræcan, spræcaþ, spræcð, spræce, spræcen, spræcend, spræcende, spræco, spræcon, spræcst, spræcu, spræcun, sprec, sprecað, specæ, specæn, sprecænne, sprecan, sprecande, sprecanne, sprecaþ, sprecð, sprece, spreceð, sprecen, sprecenan, sprecend, sprecende, sprecendes, sprecendra, sprecendum, sprecene, sprecenne, sprecest, spreceþ, spreco, sprecon, spreconne, sprecst, sprecu, sprecun, sprecþ, spreocað, spreocan, spreocanne, spreocaþ, spreocende, spreocendra, spreoco, spreocu, spricð, spriceð, spricest, spriceþ, spricþ, sprycð, sprycest, spryceþ, sprycst, sprycþ, spycð, spycst,* and *spycþ.* Some inconsistencies arise, as the lemmatizer is not able to generate eleven of the forms accounted for in the literature, while it generates forty-two attested forms not provided in the sources consulted. Figure 9 summarizes these findings.

Forms not provided by the lemmatizer	Forms not found in the literature
spæken, spæky, specce, speke, speken, spekinde, sprace, spracon, spreiced, sprice, spricst	*spæce, spæcen, spæcende, spæcenne, specæn, specaþ, speceð, specen, specene, specu, spicð, spicþ, spræcað, spræcaþ, spræcð, spræcend, spræcende, spræco, spræcst, spræcu, spræcun, specæ, specæn, sprecænne, sprecande, sprecenan, sprecend, sprecendes, sprecendra, sprecendum, sprecest, spreco, spreconne, spreocanne, spreocaþ, spreoco, spriceþ, sprycest, spryceþ, sprycþ, spycst, spycþ*

Figure 9. Comparison with secondary sources.

With respect to those forms not provided by the lemmatizer, there are several reasons to justify their absence. The forms *spæken, spæky, specce, speke, speken, spekinde* present consonantal spellings not implemented in the set of rules of the lemmatizer, namely the spelling <k> and the geminated <cc> for <c>. Likewise, the forms *sprace, spracon* display an unpredicted preterit 2 vowel <a>. While Rule#11 accounts for the breaking of <æ> into <ea>, no rule has been developed to account for the retraction of <æ> into <a>. For its part, *sprice* displays an extension of the *i*-mutated vowel of the present 2nd and 3rd person singular into the 1st person singular. In view of this data, it might well be worth considering the implementation of these rules into the lemmatizer to account for these phenomena in other verb forms. *Spreiced* and *spricst* are different cases altogether. The former constitutes a completely unexpected spelling variant, with a dental suffix proper of the inflection of weak verbs, therefore falling out of the scope of this research. As for *spricst*, the form has been generated by the lemmatizer, but it has not been attested in the corpora, as shown by Figure 10.

Figure 10. The unattested generation of *spricst*

With regard to those generated forms not accounted for in the literature, the adequacy of the assigned lemma is to be proved. Take the examples in (9) as illustration of the accuracy of automatic lemma assignment.

(9)

a.   [Bede 5 011900 (6.402.13)]
*Ða fregn he mec, hwæðer ic wiste hwa ðæt wære se ðe to mec **spræcende** wæs.*
Then he asked me, whether I knew who it was who was **speaking** to me (Miller, 1999:179-180).

b.   [Æ HomM 1 (Bel 9) 010000 (191)]
*All swa bi <gehwylce> þinge þe heo ær cuðe oðer ne cuðe; heo mæg on hire mode sceawiæn þonne heo hereð bi þam **specæn**; & swa styriende is þe sawle þæt heo forþam on slepe ne stilð.*
Even so in all matters which it knew or did not know of before; when it hears them **spoken** about, it can look on them in its mind; and so active is the soul that it does not even rest in sleep (Belfour, 1962:89).

Both attested forms present irregular spellings, either in the stem, like (9a), or in the inflectional ending, like (9b). The inflectional rules implemented in the database allow for the identification of these non-standard forms and their correct inclusion in the paradigm of *sprecan*, thus accounting for the effectiveness of the overall MG automatic lemmatization process.

4.2 Conclusions

This article has designed and implemented an automatic lemmatizer of OE of class V strong verbs based on the MG and checked its accuracy. The following conclusions can be drawn.

While in its current state of description OE is not a language suitable for NLP studies, advances can be made in the direction of speeding and automating lemmatization processes. A lemmatized corpus provides the foundation upon which NLP may operate. While the features of OE –fragmentary data and spelling irregularity– constitute a major impediment for a completely automatic lemmatization of the corpus, this article shows that small, target-oriented applications can be developed to lemmatize specific sub-categories.

Considering the automatic generation of inflected forms, issues arise that call for a lexicographer's revision of the contextual occurrences and analysis of lemma assignment. However, even if the form-lemma association is not always successful, the research proves that type-based lemmatization can largely be automatized. As for validation, this article proposes a method that allows the automatic identification of the generated forms in the major corpora of OE, which is per se a remarkable advance for the discipline. As regards accuracy, no lemma competition has arisen intracategorially, although some generated forms may be in conflict with forms generated in other lexical classes. The development of class or sub-class specific sets of rules will contribute to highlight these cases and to reduce manual revision.

Thus, several lines of research have been opened. First, the completion of the analysis of class V strong verbs L-Y with A-I prefixes. Second, the study of recursively prefixed verbs is yet to be completed. Verbs like *upawegan* 'to lift up, support' have been left out of this research. Third, the study of the verbs A-I, whose results may be compared with the attested forms provided by DOE. Finally, the analysis might be extended to other strong verb classes and ultimately to other open lexical classes.

ACKNOWLEDGMENTS

REFERENCES

Adamczyk, E. (2002). "Reduplication and the Old English Strong Verbs Class VII", *Studia Anglica Posnaniensia, 38*, 23-34.

Belfour, A.O. (1962) [1909]. *Twelfth-Century Homilies in MS. Bodley 343*. London: Oxford University Press

Bosworth, J. & Toller T.N. (1973) [1898]. *An Anglo-Saxon Dictionary*. Oxford: Oxford University Press.

Campbell, A. (1987) [1959]. *Old English Grammar*. Oxford: Oxford University Press.

Clark Hall, J.R. (1996). *A Concise Anglo-Saxon Dictionary*. *Supplement by Herbert D. Merritt*. Toronto: University of Toronto Press.

Craven, M. & Kumlien, J. (1999). "Constructing biological knowledge bases by extracting information from text sources", in *Proceedings of the Seventh Inter- national Conference on Intelligent Systems for Molecular Biology*, *77-86.*

Dahlgren, K. (1988). *Naïve Semantics for Natural Language Understanding.* Boston/Dordrecht/London: Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4613-1075-4

Dickins, B. (1950). "The Beheaded Manumission in the Exeter Book", in C. Fox & B. Dickins (eds.) *The Early Cultures of North-West Europe* Cambridge: Cambridge University Press, 361-368.

Ferrés, D., AbuRa'ed, A. & Saggion, H. (2017). "Spanish Morphological Generation with Wide-Coverage Lexicons and Decision Trees", *Procesamiento del Lenguaje Natural, 58*, 109-116.

García Fernández, L. (2020). *Lemmatising Old English on a Relational Database. Preterite-Present, Contracted, Anomalous and Strong VII Verbs*. Munich: Utzverlag.

Hamdoun Bghiyel, Y. (2020). "The Lemmatization of Old English Comparative Adverbs", *Ræl–Revista electrónica de lingüística aplicada, 19*/2, 1-23. https://doi.org/10.18172/jes.4525

Healey, A. (ed.), Price, J., & Xiang, X. 2004. *The Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Healey, A. (ed.). 2018. Th*e Dictionary of Old English: A to I*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Hedderich, M.A., et al. (2021). "A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios". In K. Toutanova et al. (eds.) *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* ACL, 2545–2568. https://doi.org/10.18653/v1/2021.naacl-main.201

Hoffman, R., Zhang, C., Ling, X., Zettlemoyer, L., & Weld, D.S. (2011). "Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations", in D. Lin, Y. Matsumoto & R. Mihalcea (eds.) *Proccedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 541-550.

Hostetter, A.K. (n.d.a). "The Final Judgment (Christ III)", in A.K. Hostetter (ed.) *The Old English Narrative Poetry Project*. Available at https://oldenglishpoetry.camden.rutgers.edu/christ-iii/

Hostetter, A.K., (n.d.b). "Genesis A & B", in A.K. Hostetter (ed.) *The Old English Narrative Poetry Project*. Available at https://oldenglishpoetry.camden.rutgers.edu/genesis-ab/

Krygier, M. (1994). *The Disintegration of the English Strong Verb System*. Frankfurt: Peter Lang.

Laing, M., & Lass, R. (2010). "Raiders of the Lost Archetype: *eo* in the Strong Verbs of Classes IV and V", *Transactions of the Philological Society, 108*/2, 145-163. https://doi.org/10.1111/j.1467-968X.2010.01238.x

Levin, R. (1964). "A Reclassification of the Old English Strong Verbs", *Language, 40*, 156-161. https://doi.org/10.2307/411574

Li, S., Graça, J., & Taskar, B. (2012). "Wiki-ly Supervised Part-Of-Speech Tagging", in J.I. Sujii, J. Henderson, & M. Pasca (eds.) *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. ACL. 1389–1398.

Kastovsky, D. (1992). "Semantics and Vocabulary", in R. Hogg (ed.), *The Cambridge History of the English Language I: The Beginnings to 1066*. Cambridge: Cambridge University Press, 290-408. https://doi.org/10.1017/CHOL9780521264747.006

Kuhn, S.M. (ed.). (1965). *The Vespasian Psalter*. Ann Arbor: University of Michigan Press.

Liu, X., Gao, J., He, X., Deng, L., Duh, K., & Wang, Y.Y. (2015). "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval", in R. Mihalcea, J. Chai, & A. Sarkar (eds.) *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, 912-921. https://doi.org/10.3115/v1/N15-1092

Mailhammer, R. (2007). *The Germanic Strong Verbs: Foundation and Development of a New System. Trends in Linguistics*. Studies and Monographs 183. Berlin/New York:de Gruyter. https://doi.org/10.1515/9783110198782

Martín Arista, J. (2012). "The Old English Prefix ge-: A panchronic reappraisal", *Australian Journal of Linguistics* 32/4, 411-433. https://doi.org/10.1080/07268602.2012.744264

Martín Arista, J. 2013. *Nerthus. Lexical Database of Old English: From Word-Formation to Meaning Construction*. Research Seminar, School of English, University of Sheffield.

Martín Arista, J. & Ojanguren López A.E. (2018). "Doing Electronic Lexicography of Old English with a Knowledge-Base." *Workshop delivered at the Consolidated Library of Anglo-Saxon Poetry (CLASP) Project* (University of Oxford).

Martín Arista, J., Domínguez Barragán,S., García Fernández, L.,Ruíz Narbona, E., Torre Alonso, R., & Vea Escarza, R. (comp.). (2021). *ParCorOEv2. An Open Access Annotated Parallel Corpus Old English-English*. Nerthus Project, Universidad de La Rioja, www.nerthusproject.com.

Mateo Mendaza, R. (2016). "The Old English exponent for the semantic prime MOVE", *Australian Journal of Linguistics, 34*/4, 542-559. https://doi.org/10.1080/07268602.2016.1169976

Metola Rodríguez, D. (2015). *Lemmatisation of Old English Strong Verbs on a Lexical Database*. Ph.D. dissertation, University of La Rioja.

Metola Rodríguez, D. (2017). "Strong Verb Lemmas from a Corpus of Old English. Advances and Issues", *Revista de Lingüística y Lenguas Aplicadas, 12*, 65-76. https://doi.org/10.4995/rlyla.2017.7023

Miller, T. (1999). *The Old English Version of Bede's Ecclesiastical History of the English People*. Cambridge, Ontario: In parenthesis Publications.

Mueller, E.T. (2014). *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-801416-5.00002-4

Novo Urraca, C. (2016). "Morphological relatedness and the typology of adjectival formation in Old English2, *Studia Neophilologica, 88*/1, 43-55. https://doi.org/10.1080/00393274.2016.1150788

Novo Urraca, C., & Ojanguren López, A.E. (2018). "Lemmatising Treebanks. Corpus Annotation with Knowledge Bases", *RAEL-Revista electrónica de Lingüística Aplicada, 17*/1, 99-120.

Pintzuk, S., & Plug, L. (2001). *The York-Helsinki Parsed Corpus of Old English Poetry.* Department of Language and Linguistic Science, University of York.

Rissanen, M. et al., (comp). (1991). *The Helsinki Corpus of English Texts*. Department of English, University of Helsinki.

Reiter, E., & Dale, R. (1997). "Building Applied Natural Language Generation Systems", *Natural Language Engineering, 3*/1, 57-87. https://doi.org/10.1017/S1351324997001502

Schank, R.C. (1972). "Conceptual Dependency: A Theory of Natural Language Understanding", *Cognitive Psychology, 3*/4, 552-631. https://doi.org/10.1016/0010-0285(72)90022-9

Taylor, A. et al*.* (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York: University of York.

Tichy, O. & Rocek, M. (2019). Online edition of *An Anglo-Saxon Dictionary*. Faculty of Arts, Charles University Prague.

Tío Sáenz, M. (2019). *The Lemmatisation of Old English Weak Verbs of a Relational Database*. Ph.D. dissertation, University of La Rioja.

Vázquez González, J.G. & Barðdal, J. (2019). "Reconstructing the Ditransitive Construction for Proto-Germanic: Gothic, Old English and Old Norse-Icelandic", *Folia Linguistica Historica, 40*/2, 555-620. https://doi.org/10.1515/flih-2019-0021

Vea Escarza, R. (2018). "Las funciones y categorías de los nombres y adjetivos afijados del inglés antiguo", *Onomázein, 41*, 208-226. https://doi.org/10.7764/onomazein.41.07

von Mengden, F. (2011). "Ablaut or Transfixation? On the Old English Strong Verbs", in R. Bauer & U. Krischke (eds.) *More than Words: English Lexicography and Lexicology Past and Present*. Frankfurt: Lang, 123-139.

APPENDIX 1. YCOE POS VERB TAGS

POS TAG	CATEGORY	POS TAG	CATEGORY
AX	Infinitive	HVI	Have, imperative
AXD	Past, ambiguous form	HVN	Have, past participle (vb. or adj.)
AXDI	Past, unambiguous indicative	HVN^N	Have, past participle (vb. or adj.), nominative
AXDS	Past, unambiguous subjunctive	HVP	Have, present, ambiguous form
AXG	Present participle	HVPI	Have, present, unambiguous indicative
AXI	Imperative	HVPS	Have, present, unambiguous subjunctive
AXN	Past participle (verbal or adjectival)	MD	Modal, infinitive
AXP	Present, ambiguous form	MD^D	Modal, infinitive, inflected
AXPI	Present, unambiguous indicative	MDD	Modal, past, ambiguous form
AXPS	Present, unambiguous subjunctive	MDDI	Modal, past, unambiguous indicative
BAG	Present participle	MDDS	Modal, past, unambiguous subjunctive
BAG^N	Present participle, nominative	MDI	Modal, imperative
BE	Be, infinitive	MDP	Modal, present, ambiguous form
BE^D	Be, infinitive, dative	MDPI	Modal, present, unambiguous indicative
BED	Be, past, ambiguous form	MDPS	Modal, present, unambiguous subjunctive
BEDI	Be, past, unambiguous indicative	VAG	Present participle
BEDS	Be, past, unambiguous subjunctive	VAG^A	Present participle, accusative
BEI	Be, imperative	VAG^D	Present participle, dative
BEN	Be, past participle	VAG^G	Present participle, genitive
BEN^A	Be, past participle, accusative	VAG^I	Present participle, instrumental
BEN^D	Be, past participle, dative	VAG^N	Present participle, nominative
BEN^G	Be, past participle, genitive	VB	Infinitive
BEN^N	Be, past participle, nominative	VB^D	Infinitive, inflected
BEP	Be, present, ambiguous form	VBD	Past, ambiguous form
BEPH	Be, present, ambiguous imp./subj.	VBDI	Past, unambiguous indicative
BEPI	Be, present, unambiguous indicative	VBDS	Past, unambiguous subjunctive
BEPS	Be, present, unambiguous subjunctive	VBI	Imperative
HAG	Have, present participle	VBN	Past participle (verbal or adjectival)
HAG^A	Have, present participle, accusative	VBN^A	Past participle (verbal or adjectival), accusative
HAG^D	Have, present participle, dative	VBN^D	Past participle (verbal or adjectival), dative
HAG^N	Have, present participle, nominative	VBN^G	Past participle (verbal or adjectival), genitive
HAG^G	Have, present participle, genitive	VBN^I	Past participle (verbal or adjectival), instrumental
HV	Have, infinitive	VBN^N	Past participle (verbal or adjectival), nominative
HV^D	Have, infinitive, inflected	VBP	Present, ambiguous form
HVD	Have, past, ambiguous form	VBPH	Ambiguous imperative/subjunctive
HVDI	Have, past, unambiguous indicative	VBPI	Present, unambiguous indicative
HVDS	Have, past, unambiguous subjunctive	VBPS	Present, unambiguous subjunctive

RLyLA Vol. 17 (2022), 143-161 | **157**

APPENDIX 2. ATTESTED FORMS BY LEMMA

    a.   **Attestations in the DOEC**

gelesan: *gales, geleasan, geleoso, gelese, gelesen, gilese, glæs, glæse, glæsen, glæsenne, glesð;* **gelicgan:** *geleagan, gelicgað, gelicgan, gelicgende, gilæg;* **gemetan:** *gemæte, gemeotu, gemest, gemetænne, gemetanne, gemetend, gemetenne, gemeto, gemetu, gemyt, gimest, gimett;* **genesan:** *gænes, geneosaþ, geneoseð, geneosende, geneoseþ, genes, genesan, gernes;* **gerecan:** *geræcað, geræcest, gerec, gerecaþ, gerecenes, gerecest, gerecet, gerecþ, giræc, giræce, girec, grec;* **gerepan:** *gedrep, grep, grepe, gripð;* **gescrepan:** *gescræpe, gescrepe;* **gesittan:** *gesætun, gesetan, gesetena, gesetenes, geseton, gesettet, gesetteþ, gesetun, gesist, gesit, gesittaþ, gesitteð, gesitten, gesitteþ, gisæt, gisætt, gisete, gisettan, gisette, giseton, gisett, gisetteð, gisetun, gisettun, gisist, gisitte, gisittende;* **gesprecan:** *gespæc, gespæce, gespec, gespræcen, gespræcend, gespræcu, gesprec, gesprecæ, gesprecæn, gesprecaþ, gesprece, gesprecend, gesprecendum, gespreco, gesprecu, gespreocu, gispræc, gisprece, gispreceð, gisprecen, gisprecon, gisprecun;* **getredan:** *getred, getrede, getredene;* **geþicgan:** *geþogenne;* **gewefan:** *gewef, gewefe, gewefene, gewefenum, gewoefen, giwefen;* **gewegan:** *gewægan, gewæge, gewæh, gewoege, gewoegen, giwege, giwegen, giwigeð;* **gewrecan:** *gewræcon, gewrec;* **lesan:** *læsan, læse, læsest, læsu, leasan, lease, leasest, leaso, leason, least, leosan, leose, leosende, les, lesað, lese, lesend, lesende, leso, lisit, lorene;* **licgan:** *læh, lagun, leage, lecgaþ, lecgen, legan, legen, legena, legene, legon, licgeð, licgenda;* **medlesan:** *meles, midles, miles;* **medmetan:** *memet;* **mednesan:** *midnæs, minæs, mines;* **medrecan:** *merece;* **medsittan:** *misit;* **medsprecan:** *midspecan, midsprecan;* **medwegan:** *midwæge, midwege;* **metan:** *mæst, mætan, mæteð, mæten, mætu, meast, meat, meoto, meotu, metæ, metena, metend, metest, meteþ, meto, mett, mist, mit, myst, myt;* **missprecan:** *missprecon;* **nesan:** *næsan, næse, næso, næst, nearan, neosað, neosæ, neose, neosendes, neren, nerin, nerun, nesan, nese, nest;* **oferlicgan:** *oferlecgað;* **ofermetan:** *ofermætan, ofermæte, ofermæto, ofermete, ofermett, ofyrmæte;* **ofersittan:** *oferseton, ofersetton, ofersetun, ofersettun, ofyrsæton;* **ofersprecan:** *oferspræc, oferspræcan, oferspræce, ofersprecendes;* **ofertredan:** *ofertræd, ofertredan, ofertret, ofyrsette, ofyrsetton;* **oferwegan:** *oferwege, oferwigeð, ouerwyhð;* **oferwrecan:** *oferwrecð, oferwrycð;* **oflicgan:** *afleah, oflicge;* **ofnesan:** *æfnes, ofnes;* **ofsittan:** *ofsæton, ofsete, ofsettet, ofsetton, ofsettun;* **oftredan:** *oftræde, oftredene;* **olesan:** *oles;* **onesan:** *ones;* **onlesan:** *onlesað, onlese, onleseð, onlesend;* **onlicgan:** *onlæg, onleah, onlecge, onlege, onlegen, onlegena, onlegene, onlegenum;* **onsittan:** *onsæte, onset, onsettan, onsetteð, onsettun, onsittaþ, onsittend, onsittendan, onsittendre;* **onsprecan:** *onspæce, onspece, onspræce, onspræcon, onsprece;* **onwegan:** *onwæg, onweg, onwege;* **orepan:** *orep;* **orlesan:** *orlease;* **orlicgan:** *orlæg, orlege;* **ormetan:** *ormæt, ormætan, ormæte, ormæten, ormæton, ormætu, ormete, ormeten;* **ornesan:** *ornest;* **orwegan:** *orweg;* **osittan:** *oset;* **owefan:** *owef;* **recan:** *ræc, ræcað, ræcende, ræceþ, ræco, recend, recende, recene, recon, reconne, reocende;* **repan:** *ræpan, reopað, reopan, rep, repað, repeð, repo, ripeð, rypð, rypeð;* **rihtlicgan:** *rihtlagan, rihtlage;* **rihtnesan:** *rehtnes;* **rihtrecan:** *rihtræce;* **rihtwegan:** *rihtwege;* **screpan:** *screope, scripið, scripit;* **sinnesan:** *sinnes;* **sintredan:** *sintredende;* **sittan:** *sat, satan, seate, setð, seten, setena, setene, setin, seton, setteð, setteþ, setun, sitteð, sittenda, sittendes, sittendra, sittendu, sittest, sittet, sitteþ;* **sprecan:** *spæcð, spæcende, spæcenne, specæn, specaþ, speceð, specen, specene, specu, spræcð, spræcend, spræco, spræcu, sprecæ, sprecæn, sprecend, sprecendes, sprecendra, sprecu, spreocanne, spreocaþ, spreocendra, spreoco, spreocu, spriceþ, spryceþ;* **swefan:** *swefen, swefena, swefene, swefenu, swefenum;* **tolesan:** *toles, tolesan, tolesendes, toleseþ, tolist, tolysð, tolyst, tolysþ;* **tolicgan:** *tolæg, tolecge, tolicgaþ;* **tometan:** *tomete;* **tosittan:** *tosett, tosetteð, tosetteþ, tosettan, tosetten, tosetton;* **tosprecan:** *tospræcan, tospræce, tosprece, tosprecene, tospycst;* **tredan:** *trædan, træden, tred, tredanne, tredeð, tredend, tredendra, tredene, tredst, tredun, tret, tretst, trideþ, trydest, trydst, trytt;* **twisprecan:** *twigspræc, twigspræce,*

twispæce, twispræce, twisprece, twyspræce; **twiwegan:** *twiwæge, twiwege;* **þicgan:** *ðæg, ðæge, ðægen, ðæh, ðech, ðegen, ðegena, ðegenan, ðegene, ðegenes, ðegenum, ðegin, ðicgendum, þæg, þæge, þægen, þæh, þag, þage, þagen, þegan, þegen, þegena, þegene, þegenes, þegenum, þicgeð, þoegenes;* **þrinesan:** *ðrines, þrienes, þrines, þrynes;* **underlesan:** *underlist;* **underlicgan:** *underlæg, underlecge;* **undermetan:** *undermete, undernmete;* **undernesan:** *undernes;* **undersittan:** *underset, undersitte;* **unlesan:** *unlease;* **unlicgan:** *unlag, unlage, unlagon;* **unmetan:** *unmætan, unmæte, unmetan, unmete;* **unnesan:** *unnes;* **unsprecan:** *unsprecendan, unsprecende, unsprecendra;* **unwegan:** *unwegen;* **unwrecan:** *unwrecen;* **uplicgan:** *uplegen, uplegene;* **uprecan:** *upreceð;* **upwegan:** *upweg, upwegen;* **utlesan:** *utlæse;* **utlicgan:** *utlagan, utlage, utlagen, utlah;* **utsittan:** *utset;* **uþmetan:** *uþmæte;* **wanwegan:** *wanwægendum, wanwegendum;* **wefan:** *wæfan, wæfð, wæfendum, wæfenne, wef, wefanne, wefð, wefe, wefen, wefendum, weofendan, wifeð, wyfð;* **wegan:** *wægæ, wægeð, wægendes, wægendre, wægest, wægeþ, wæh, wæxt, weg, wegæ, wegð, wegena, wegend, wegende, wegon, weh, wigið, wigþ, wihð, wixt, woeg, woege, woegena, wyhst;* **wiþermetan:** *wiðermet, wiðermeten, wiþermet;* **wiþersprecan:** *wiðerspece, wiðersprecend;* **wiþlicgan:** *wiðlagan;* **wiþmetan:** *wiðmet, wiðmeteþ, wiþmeten, wiþmetendra, wiþmetenes, wiþmeteþ;* **wiþsittan:** *wiðset, wiþsette, wiðsettun;* **wiþsprecan:** *wiðspræcan, wiðspræcon, wiðsprecendes, wiðsprecst;* **wrecan:** *wræcð, wræcen, wræcenda, wræco, wrecæ, wrecæn, wrecend, wrecenda, wrecendan, wrecenne, wreco, wreocan, wreocende, wreocu, wricest, wrocena, wrocene;* **ymblicgan:** *ymbelicgað;* **ymbsittan:** *ymbesette, ymbsete, ymbsett, ymbsette, ymbsetteþ, ymbsetton, ymbsettun, ymbsittendum, ymsitt, ymsittendum;* **ymbsprecan:** *embspræc, ymbespæce, ymbespræce, ymbspræce, ymbsprecon.*

b. Attestations in the DOEC and the YCOE

gelesan: *gelæran-(VB), gelæron-(VB), gelæst-(VBN)-(VBPI), gelest-(VBN), gelyst-(VBPI);* **gelicgan:** *gelæg-(VBDI), gelæge-(VBDS), gelah-(VBDI), geleah-(VBDI), gelecge-(VBPS), gelege-(VBI), gelegen-(VBN), gelicgaþ-(VBPI), gelicge-(VBPS);* **gemetan:** *gemæst-(VBN), gemet-(VBN)-(VBPI), gemetað-(VBPI), gemetæ-(VBPS), gemetan-(VB)-(VBPS), gemetaþ-(VBPI), gemete-(VBD)-(VBP)-(VBPS), gemeteð-(VBPI), gemeten-(VB)-(VBN)-(VBPS), gemetende-(VAG), gemetest-(VBPI), gemeteþ-(VBPI), gemeton-(VB)-(VBPS), gemetst-(VBPI), gemett-(VBN)-(VBPI), gemetun-(VBDI), gymest-(VBPI);* **genesan:** *genæs-(VBDI), genæson-(VBDI), geneosað-(VBPI), genesen-(VBN), geneseþ-(VBPI);* **gerecan:** *gedrecð-(VBPI), geræc-(VBDI), geræcan-(VB), geræcð-(VBPI), geræce-(VBP)-(VBPS), geræceð-(VBPI), geræcen-(VB), geræcon-(VB), gerecan-(VB), gerecanne-(VB^D), gerecð-(VBPI), gerece-(VBI)-(VBPS), gereceð-(VBPI), gereceþ-(VBPI), gereocan-(VB);* **gerepan:** *gedreopan-(VB), gedrepen-(VBN), gripeð-(VBPI);* **gesittan:** *gesæt-(VBDI)-(VBN)-(VBPI), gesætan-(VBDI), gesæte-(VBDS), gesæton-(VBDI), gesæt-(VBDI), gesætte-(VBD)-(VBN^N), gesætton-(VBDI), geset-(VBDI)-(VBN)-(VBN^N)-(VBPI), gesete-(VBI)-(VBPS), geseten-(VBN), gesetene-(VBN^A)-(VBN^N), gesetenne-(VBN^A), gesetenum-(VBN^D), gesett-(VBN)-(VBN^N)-(VBPI), gesettan-(VB)-(VBDI)-(VBN)-(VBN^A)-(VBN^D), gesette-(VBD)-(VBDS)-(VBN)-(VBN^A)-(VBN^N), gesetten-(VBD), gesetteð-(VBPI), gesetton-(VB)-(VBDI)-(VBN^D)-(VBPS), gesetton-(VBDI), gesitst-(VBPI), gesittað-(VBPI), gesittan-(VB), gesitte-(VBP)-(VBPS), gesittende-(VAG), gesittenne-(VB^D);* **gesprecan:** *gespecan-(VBN), gespecen-(VBN), gespræc-(VBDI), gespræcan-(VB)-(VBDI)-(VBN), gespræce-(VBDS), gespræcon-(VBDI), gesprecað-(VBPI), gesprecan-(VB)-(VBN), gesprecen-(VB)-(VBN), gesprecende-(VAG), gesprecene-(VBN^N), gesprecenra-(VBN^G), gesprecenum-(VBN^D), gesprecon-(VBN)-(VBPS);* **getredan:** *getreden-(VBN);* **geþicgan:** *geðah-(VBDI), geðeah-(VBDI), geðicgan-(VB), geðicge-(VBPS), geðogen-(VBN)-(VBN^N), geþægon-(VBDI), geþah-(VBDI), geþeah-(VBDI), geþicge-(VBPS), geþicgenne-(VB^D), geþogen-(VBN)-(VBN^N), geþogene-(VBN^N);* **gewefan:** *gewæf-(VBDI), gewefen-*

(VBN)-(VBN^N); **gewegan:** *gewæg-*(VBDI), *gewegan-*(VB), *gewege-*(VBPS), *gewegen-*(VBPS), *gewigeð-*(VBPI), *gewihð-*(VBPI); **gewrecan:** *gewræc-*(VBDI), *gewræcan-*(VBDI), *gewræce-*(VBDS), *gewræcen-*(VBDS)-(VBPS), *gewrecan-*(VB)-(VBN)-(VBPS), *gewrecð-*(VBPI), *gewrece-*(VBP), *gewrecen-*(VB)-(VBN), *gewrecene-*(VBN^N), *gewrecon-*(VBPS), *gewrecst-*(VBPI), *gewrecþ-*(VBPI); **lesan:** *læran-*(VB)-(VBPS), *læron-*(VB)-(VBPS), *læs-*(VBDI), *læson-*(VBDI), *læst-*(VBI), *lesan-*(VB), *lest-*(VBPI), *list-*(VBPI), *lisð-*(VBPI), *lysð-*(VBPI), *lyst-*(VBPI); **licgan:** *læg-*(VBDI), *lægan-*(VBDI), *læge-*(VBD)-(VBDS), *lægen-*(VBD)-(VBDS), *lægon-*(VBDI), *lægun-*(VBDI), *lag-*(VBDI), *lagan-*(VBDI), *lage-*(VBD)-(VBDS), *lagen-*(VBD), *lagon-*(VBDI), *lah-*(VBDI), *leag-*(VBDI), *leah-*(VBDI)-(VBI), *lecgað-*(VBI)-(VBPI), *lecge-*(VBP)-(VBPS), *lecgende-*(VAG^A)-(VAG^N), *lege-*(VBI)-(VBPS), *licgað-*(VBPI), *licgan-*(VB)-(VBPS), *licganne-*(VB^D), *licgaþ-*(VBPI), *licge-*(VBP)-(VBPS), *licgendan-*(VAG^D)-(VAG^G)-(VAG^N), *licgende-*(VAG)-(VAG^A)-(VAG^N), *licgendre-*(VAG^D)-(VAG^N), *licgendum-*(VAG^D)-(VAG^N), *licgenne-*(VB^D); **medsittan:** *midsittendum-*(VAG^D)-(VAG^N); **medstecan:** *mistæcað-*(VBPI); **metan:** *mæt-*(VBDI)-(VBPI), *mæte-*(VBPS), *mæton-*(VBDI), *mest-*(VBPI), *met-*(VBPI), *metað-*(VBPI), *metan-*(VB), *metanne-*(VB^D), *metaþ-*(VBPI), *mete-*(VBPS), *meteð-*(VBPI), *meten-*(VBN)-(VBPS), *metende-*(VAG^N), *metenne-*(VB^D), *metst-*(VBPI), *moten-*(MDP)-(MDPS); **mislesan:** *mislæran-*(VB); **misrecan:** *misræce-*(VBPS), *misræceð-*(VBPI); **misþicgan:** *misþah-*(VBDI); **nesan:** *næran-*(BEDI), *næron-*(BED)-(BEDI), *nærun-*(BEDI), *næs-*(BED)-(BEDI), *neosan-*(VB), *neron-*(BEDI), *nes-*(BEDI), *nesen-*(VBDS); **oferlicgan:** *oferlecge-*(VBPS), *oferlege-*(VBI), *oferlicge-*(VBPS); **ofersittan:** *oferseten-*(VBN), *ofersette-*(VBD), *ofersettan-*(VBN^A), *ofersittan-*(VB), *ofersitte-*(VBP); **ofertredan:** *ofertrit-*(VBPI); **oferþicgan:** *oferðeah-*(VBDI), *oferþeah-*(VBDI), *oferþogen-*(VBN); **oflesan:** *oflyst-*(VBN); **oflicgan:** *aflogen-*(VBN), *aflogene-*(VBN^N), *oflege-*(VBI); **ofsittan:** *ofsæt-*(VBDI), *ofset-*(VBN), *ofseten-*(VBN), *ofsetenan-*(VBN^A)-(VBN^D), *ofsetene-*(VBN^N), *ofsetenum-*(VBN^D), *ofsett-*(VBN), *ofsettan-*(VBDI)-(VBN^A)-(VBN^D), *ofsette-*(VBD)-(VBN^N), *ofsit-*(VBPI), *ofsitt-*(VBPI), *ofsittað-*(VBPI), *ofsittan-*(VB)-(VBPS), *ofsitte-*(VBPS); **oftredan:** *oftræd-*(VBDI), *oftrædan-*(VBDI), *oftredan-*(VB), *oftreden-*(VBN), *oftret-*(VBPI); **onlesan:** *onlyseð-*(VBPI), *onlyseþ-*(VBPI), *onlyst-*(VBPI); **onlicgan:** *onlag-*(VBDI), *onlagon-*(VBDI), *onlah-*(VBDI), *onlecgende-*(VAG^A)-(VAG^N); **onsittan:** *onsæt-*(VBDI), *onsæton-*(VBDI), *onsett-*(VBPI), *onsette-*(VBD)-(VBPS), *onsetton-*(VBDI), *onsit-*(VBPI), *onsitt-*(VBPI), *onsittað-*(VBPI), *onsittan-*(VB), *onsitte-*(VBP)-(VBPS), *onsittende-*(VAG); **onsprecan:** *onspræc-*(VBDI); **onþicgan:** *onðah-*(VBDI), *onþah-*(VBDI), *oðþah-*(VBDI); **onwrecan:** *onwrecen-*(VBN); **recan:** *ræcan-*(VB), *ræcð-*(VBPI), *ræce-*(VBP), *ræceð-*(VBPI), *rec-*(VBI), *recð-*(VBPI), *rece-*(VBI), *receð-*(VBPI), *recen-*(VBPS), *recenne-*(VB^D), *recest-*(VBPI), *receþ-*(VBPI), *recst-*(VBPI), *recþ-*(VBPI), *reocað-*(VBPI), *reocan-*(VB), *reocendan-*(VAG^D)-(VAG^G)-(VAG^N), *reocendes-*(VAG^G)-(VAG^N), *reocendum-*(VAG^D)-(VAG^N), *rycþ-*(VBPI); **repan:** *ræpon-*(VBDI), *ripð-*(VBPI), *ripst-*(VBPI), *ripþ-*(VBPI); **screpan:** *screp-*(VBI); **sittan:** *sæt-*(VBD)-(VBDI), *sætan-*(VBDI), *sæte-*(VBDS), *sæten-*(VBD), *sæton-*(VBDI), *sætt-*(VBDI), *sættan-*(VB); *sætte-*(VBD), *sætun-*(VBDI), *set-*(VBDI)-(VBPI), *setan-*(VBDI), *sete-*(VBDS)-(VBI), *sett-*(VBPI), *settan-*(VB)-(VBDI)-(VBPS), *settæn-*(VB), *sette-*(VBD)-(VBP)-(VBPS), *setten-*(VB)-(VBD)-(VBPS), *setton-*(VB)-(VBDI)-(VBPS), *settun-*(VBDI), *sit-*(VBPI), *sitst-*(VBPI), *sitt-*(VBPI), *sittan-*(VB)-(VBPS), *sittande-*(VAG), *sittanne-*(VB^D), *sittaþ-*(VBPI), *sittað-*(VBI)-(VBPI), *sitte-*(VBP)-(VBPS), *sitten-*(VB)-(VBPS), *sittendan-*(VAG^A)-(VAG^N), *sittende-*(VAG)-(VAG^A)-(VAG^N), *sittendne-*(VAG^A)-(VAG^N), *sittendum-*(VAG^D)-(VAG^N), *sittenne-*(VB^D); **sprecan:** *spæc-*(VBDI), *spæcan-*(VB)-(VBDI), *spæce-*(VBD)-(VBDS), *spæcen-*(VBD), *spæcon-*(VB)-(VBDI), *spec-*(VBDI)-(VBI), *specað-*(VBI)-(VBPI), *specan-*(VB)-(VBPS), *specð-*(VBPI), *spece-*(VBP)-(VBPS), *specende-*(VAG), *specenne-*(VB^D), *specon-*(VBDI)-(VBPS), *specst-*(VBPI), *spicð-*(VBPI), *spicþ-*(VBPI), *spræc-*(VBDI)-(VBI), *spræcað-*(VBPI), *spræcan-*(VB)-(VBDI), *spræcaþ-*(VBPI), *spræce-*(VBD)-(VBDS)-(VBP)-(VBPS), *spræcen-*(VBPS), *spræcende-*(VAG), *spræcon-*(VBDI)-(VBPS), *spræcst-*(VBPI), *spræcun-*(VBDI), *sprec-*(VBDI)-(VBI), *sprecað-*(VBDI)-(VBI)-(VBPI), *sprecænne-*(VB^D), *sprecan-*(VB)-(VBDI)-(VBPS), *sprecande-*(VAG^A)-(VAG^N), *sprecanne-*(VB^D), *sprecaþ-*(VBDI)-(VBPI), *sprecð-*(VBPI), *sprece-*(VB)-(VBD)-(VBP)-(VBPS),

spreceð-(VBPI), *sprecen*-(VBN)-(VBPS), *sprecenan*-(VBN^D), *sprecende*-(VAG)-(VAG^N), *sprecendum*-(VAG^D)-(VAG^N)-(VBN^D), *sprecene*-(VB^D), *sprecenne*-(VB^D), *sprecest*-(VBPI), *spreceþ*-(VBPI), *spreco*-(VBP), *sprecon*-(VB)-(VBDI)-(VBN)-(VBPS), *spreconne*-(VB^D), *sprecst*-(VBPI), *sprecun*-(VBPS), *sprecþ*-(VBPI), *spreocað*-(VBPI), *spreocan*-(VB), *spreocende*-(VAG), *spricð*-(VBPI), *spriceð*-(VBPI), *spricest*-(VBPI), *spricþ*-(VBPI), *sprycð*-(VBPI), *sprycest*-(VBPI), *sprycst*-(VBDI)-(VBPI), *sprycþ*-(VBPI), *spycð*-(VBPI), *spycst*-(VBDI)-(VBPI), *spycþ*-(VBPI); **swefan:** *swæf*-(VBDI), *swæfeð*-(VBPI), *swæfon*-(VBDI), *swæfun*-(VBDI), *swefað*-(VBPI), *swefan*-(VB), *swefeð*-(VBPI), *swefeþ*-(VBPI), *swifeð*-(VBPI); **tolesan:** *toleseð*-(VBPI), *tolesen*-(VBN), *tolyseð*-(VBPI), *tolyseþ*-(VBPI); **tolicgan:** *tolegena*-(VBN^N), *tolicgað*-(VBPI), **tosittan:** *tosetene*-(VBN^N); *tosette*-(VBN^N); **tosprecan:** *tospræc*-(VBDI), *tospræcon*-(VBDI), *tosprecende*-(VAG); **towegan:** *towegen*-(VBN); **towrecan:** *towrecene*-(VBN^N); **tredan:** *træd*-(VBDI), *træde*-(VBD), *trædon*-(VBDI), *trædun*-(VBDI), *tredað*-(VBPI), *tredan*-(VB), *trede*-(VBP), *treden*-(VBN), *tredenne*-(VB^D), *tredon*-(VBPS), *trit*-(VBPI), *trydeð*-(VBPI); **underlicgan:** *underlæge*-(VBDS), *underlicge*-(VBPS), *underlicgen*-(VBPS); **unlesan:** *unlesan*-(VB); **upwegan:** *upwegað*-(VBPI); **utrecan:** *utræcan*-(VB); **þicgan:** *ðah*-(VBDI), *ðeah*-(VBDI), *ðegon*-(VBDI), *ðicgan*-(VB)-(VBPS), *ðicganne*-(VB)-(VB^D), *ðicgað*-(VBI)-(VBPI), *ðicge*-(VBP)-(VBPI)-(VBPS), *ðicgenne*-(VB^D), *þah*-(VBDI), *þeah*-(VBDI), *þege*-(VBDS)-(VBPS), *þegon*-(VBDI), *þegun*-(VBDI), *þicgað*-(VBPI), *þicgan*-(VB), *þicganne*-(VB^D), *þicge*-(VBPS), *þicgen*-(VB)-(VBPS), *þicgenne*-(VB^D); **þurhwrecan:** *þurhwræcon*-(VBDI), *þurhwrecen*-(VBN^A); **wefan:** *wæf*-(VBDI), *wæfon*-(VBDI), *wefan*-(VB), *weofað*-(VBPI), *wyfst*-(VBPI); **wegan:** *wæg*-(VBDI), *wægan*-(VB), *wægð*-(VBPI), *wæge*-(VBP), *wægon*-(VBDI), *wægun*-(VBDI), *wegan*-(VB)-(VBDI), *wegað*-(VBI)-(VBPI), *wege*-(VBP)-(VBPS), *wegeð*-(VBPI), *wegen*-(VBN), *wegendum*-(VAG^D)-(VAG^N), *wegenne*-(VB^D), *wehst*-(VBPI), *wigeð*-(VBPI), *wihst*-(VBPI), *wyxt*-(VBPI); **wiþlicgan:** *wiðlæg*-(VBDI); **wiþmetan:** *wiðmæten*-(VBN), *wiðmete*-(VBP), *wiðmeten*-(VBN), *wiðmetene*-(VB^D)-(VBN^N), *wiðmetenne*-(VB^D); **wiþsittan:** *wiðsette*-(VBD); **wiþsprecan:** *wiðspæc*-(VBDI), *wiðspræce*-(VBDS), *wiðsprecan*-(VB), *wiðsprecð*-(VBPI), *wiðsprece*-(VBP); **wrecan:** *wræc*-(VBDI), *wræce*-(VBDS)-(VBPS), *wræcon*-(VBDI), *wrec*-(VBI), *wrecað*-(VBI)-(VBPI), *wrecan*-(VB), *wrecanne*-(VB^D), *wrecð*-(VBPI), *wrece*-(VBP)-(VBPS), *wreceð*-(VBPI), *wrecen*-(VB)-(VBN)-(VBN^A)-(VBN^N)-(VBPS, *wrecende*-(VAG), *wrecendum*-(VAG^D)-(VAG^N), *wrecene*-(VB^D), *wreceþ*-(VBPI), *wrecon*-(VB)-(VBPS), *wricð*-(VBPI), *wriceð*-(VBPI), *wrycð*-(VBPI); **ymblicgan:** *ymbelæg*-(VBDI); **ymbsittan:** *embsæt*-(VBDI), *embsette*-(VBD), *imbsæton*-(VBDI), *ymbesætan*-(VBDI), *ymbesæton*-(VBDI), *ymbesittendra*-(VAG^G), *ymbsæt*-(VBDI), *ymbsætan*-(VBDI), *ymbsæton*-(VBDI), *ymbset*-(VBN), *ymbseten*-(VBN), *ymbsetenan*-(VBN^A), *ymbsette*-(VBN^A)-(VBN^N), *ymbsittan*-(VB), *ymbsittað*-(VBI)-(VBPI), *ymbsitte*-(VBPS), *ymbsittend*-(VAG), *ymbsittenda*-(VAG^A), *ymbsittendan*-(VAG^A)-(VAG^N), *ymbsittendra*-(VAG^G)-(VAG^N); **ymbsprecan:** *embespæc*-(VBDI), *ymbspræcon*-(VBDI).

c.   Attestations in the YCOE

genesan: *geneseð* (VBPI), **geþicgan:** *geðicgenne* (VB^D), *geðogene* (VBN^N); **oferþicgan:** *oferðogen* (VBN); **recan:** *rycð* (VBPI); **þicgan:** *ðege* (VBDS) (VBPS), *ðicgen* (VB) (VBPS); **þurhwrecan:** *ðurhwræcon* (VBDI), *ðurhwrecen* (VBN).