AUTOMATIC LEMMATIZATION OF OLD ENGLISH CLASS III STRONG VERBS (L-Y) WITH ALOEV3

Roberto Torre Alonso
Universidad de La Rioja
roberto.torre@unrioja.es

ABSTRACT. This article presents ALOEV3, a lemmatizer based on Morphological Generation that allows for the type-based automatic lemmatization of Old English Class III strong verbs beginning with the letters L–Y. The lemmatizer operates on the basis of the inflectional, derivational and morpho-phonological alternation rules characteristic of this class. The generated form-types are checked against the two most reputed Old English corpora, namely the *Dictionary of Old English Corpus* and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* to validate their attestations and assign the corresponding lemma. Results show that 97 percent of the validated forms are successfully assigned a single lemma. The remaining inflectional forms (38 out of 1,256) show competition between two lemmas, which implies that despite the high level of accuracy of the lemmatizer, contextual, token-based analysis is still needed for disambiguation. However, the research shows that competition only occurs in a limited set of lemma pairs and their derivatives. Although the research focuses on but one strong verb class, it confirms that exploring the avenues of automatic lemmatization will contribute to the field of Old English lexicography by either lemmatizing attested inflectional form types or by highlighting areas for manual revision.

Keywords: Old English, automatic lemmatization, strong verb, morphology, Natural Language Generation, Morphological Generation.

LEMATIZACIÓN AUTOMÁTICA DE LOS VERBOS FUERTES DE CLASE III (L-Y) DEL INGLÉS ANTIGUO CON ALOEV3

RESUMEN. Este artículo presenta ALOEV3, un lematizador basado en la Generación Morfológica que permite la lematización automática basada en tipos de los verbos fuertes de la clase III del inglés antiguo que comienzan por las letras L-Y. El lematizador opera sobre la base de las reglas de alternancia flexiva, derivativa y morfofonológica propias de esta clase. Las formas generadas se contrastan con los dos corpus de referencia del inglés antiguo, a saber, el *Dictionary of Old English Corpus* y el *York-Toronto-Helsinki Parsed Corpus of Old English Prose*, a fin de validar su atestiguación y asignarles el lema correspondiente. Los resultados muestran que al 97% de las formas generadas se les asigna un único lema. El resto de formas flexivas generadas (38 de 1.256) muestran competencia entre dos lemas. Esto implica que a pesar del algo grado de precisión del lematizador, aún es necesaria la desambiguación contextual basada en tokens. No obstante, la competición de lemas está restringida a un número limitado de pares de lemas y sus derivados. Aunque centrada en sólo una clase de verbos, esta investigación confirma que explorar procesos de lematización automática contribuye al campo de la lexicografía del inglés antiguo, bien mediante la lematización automática de las formas atestiguadas o la identificación de zonas grises que requieren revisión manual.

Palabras clave: inglés antiguo, lematización automática, verbo fuerte, morfología, Generación del Lenguaje Natural, Generación Morfológica.

1. AIMS AND RELEVANCE

This article explores the limits and potentials of the automation of the lemmatization process for Old English (OE) Class III strong verbs beginning with the letters L–Y. The aim of this study has been to develop ALOEV3, a tool that performs the morphological generation (MG) of the paradigms of the selected verbs and assesses the attestation of the generated forms in the major OE corpora. Four partial goals have been targeted. To wit, (i) to develop a morphological generator that can turn out OE inflectional forms consistent with the morphological paradigms described in the grammars of this historical language; (ii) to implement ALOEV3 with sets of rules that guide the MG of Class III strong verb forms (both simplex and complex) taking morpho-phonological and spelling alternation variants into consideration and (iii) to automatically assess the attestation of the generated forms in the authoritative OE corpora, Healey *et al.*'s (2009) *The Dictionary of Old English Corpus* (DOEC) and Taylor e*t al.*'s (2003) *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (YCOE) and (iv) to link the attested forms with their corresponding lemma.

Biber *et al.* (1998: 29) define lemma as the base form of a word. Consequently, lemmatization is the mapping of the different tokens in a corpus onto their corresponding lemma, as in DEAL (deal, deals, dealing, dealt…). Automatic

lemmatization has been one of the targets of the Natural Language Generation (NLG) theory, defined by Reiter and Dale (1997: 1) as "[…] the subfield of artificial intelligence and computational linguistics that is concerned with the construction of computer systems that can produce understandable texts in English or other human languages from some underlying non-linguistic representation of information". Indeed, according to Manjavacas *et al.* (2019: 1493) "lemmatization is considered to be solved for analytic and resource-rich languaguages such as Englihs". To just mention a few, NLG has been applied to different modern languages, including Spanish (Forcada *et al.* 2011), Arabic (Khemakhem *et al.* 2015), Turkish (Oflazer and Saraçlar 2018) or Thai (Tapsai *et al.* 2021).

However, lemmatization is still a pending task for low-resource languages (including historical languages like OE) for a number of reasons, including overlaps of graphic forms resulting from morphological and phonological developments as well as from diatopic variation, on the one hand, and the relatively limited amount of textual data, on the other. The surviving OE word stock is reduced to a few million words stored in defective, fragmentary corpora. The major corpora compiling OE written records are the DOEC, comprising 3,000,000 words and the YCOE which includes about 1,500,000 words. Further, several smaller corpora including Rissanen *et al.*'s (1991) *The Helsinki Corpus of English texts* (300,000 words) and Pintzuk and Plug's (2001) *The York-Helsinki Parsed Corpus of Old English Poetry* (70,000 words) are at the researcher's disposal. However, these corpora have not yet been lemmatized, although the YCOE and the York-Helsinki corpus do provide linguistic metadata, including morphological tagging and syntactic parsing. Martín Arista *et al.*'s (2021) *An open access annotated parallel corpus Old English-English* constitutes the most recent incorporation to the repository of OE corpora. It currently files 110,000 word tokens, which, unlike the corpora cited above, are provided with glosses and translations and are fully lemmatized.

Most importantly, the lack of a lemmatized corpus prevents lemma-form associations to be established. Thus, the development of automatized text taggers, which need abundant linguistic (meta)data for machine training, is heavily restricted. As Manjavacas *et al.* (2019: 1493) explain, "for languages with higher surface variation, lemmatization plays a crucial role as a preprocessing step for downstream tasks such as topic modelling, stylometry and information retrieval."

Even if some editions of the texts include glossaries, there is considerable divergence in their structure and lemma selections both internally and intratextually. This is also the case with dictionaries. Bosworth and Toller's (1973[1898]) *An Anglo-Saxon Dictionary*, or Clark Hall's (1996) *A Concise Anglo-Saxon Dictionary*, show several inconsistencies and differences in headword or lemma selections. *The Dictionary of Old English,* by Healey *et al.* (2018), does offer a rather systematic lemma selection process in addition to including lists of attested forms linked to each lemma. However, it has only been completed up to the letter I at this point

Given all these limmitations, this article explores the potentials of the MG of a set of OE Class III strong verbs as a means to contribute to the lemmatization of OE

corpora. Ferrés *et al.* (2017: 110) define MG as 'the task of producing the appropriate inflected form of a lemma in a given textual context and according to some morphological features.' Figure 1 shows an example approach of MG with regard to the OE verb *swelgan* 'to swallow', which, inflected for *person* (third), *number* (singular), *tense* (present), *mode* (indicative), generates the form *swilhþ* 'he/she swallows.'
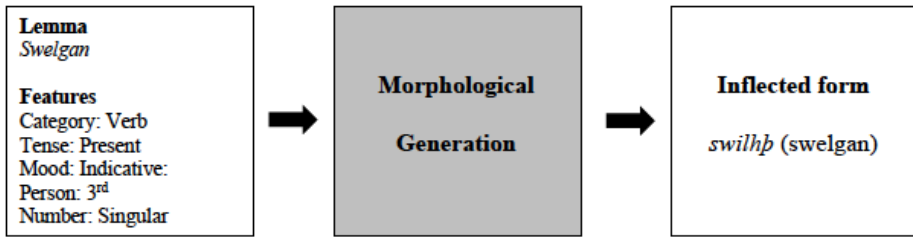

Figure 1. MG in the OE strong verb paradigm of *swelgan*.

With regard to the methodological approach adopted, this article is type-based rather than token-based, which entails the methodological decision of lemmatizing in two steps. Firstly, a type, that is to say, an abstraction from all the attestations of an inflection, is lemmatized. Then, in a second step, that necessarily involves disambiguation, all the attestations are lemmatized in their respective contexts. This article takes the first of these steps.

Finally, this article may be seen as a contribution to the linguistic analysis of OE with lexical databases carried out, among others, by García García (2019), Martín Arista (2012a; 2012b; 2013a; 2013b; 2016; 2017a; 2017b; 2018; 2019; 2020a; 2020b; 2021a; 2021b), Vea Escarza (2013; 2016a; 2016b; 2018), Mateo Mendaza (2014; 2015; 2016a; 2016b), Novo Urraca (2015; 2016a; 2016b), Martín Arista and Ojanguren López (2018), Ojanguren López (2020), and Garcia García and Ruiz Narbona (2021).

The outline of the article is as follows. Section 2 describes previous research on the automatic lemmatization of the OE corpora. Section 3 discusses the subclassification of Class III verbs and limits the scope of the research. Section 4 describes the methodological underpinnings. Section 5 discusses the results of the research and elaborates on the limits of automation. Finally, Section 6 summarizes the most relevant findings, presents some conclusions, and offers paths of research yet to be explored.

2. PREVIOUS RESEARCH ON OE (SEMI)-AUTOMATIC LEMMATIZATION

Several works have recently been published which engage in the semi-automatic lemmatization of OE corpora. Novo Urraca and Ojanguren López (2018) have

successfully incorporated lemma assignment to the YCOE syntactic treebanks. Metola Rodríguez (2015; 2017) has tackled the lemmatisation of strong verbs; Tío Sáenz (2019) has dealt with weak verbs; and García Fernández (2020) has engaged in the identification of lemmas for preterite-present, anomalous, and contracted verbs. These authors set their research on the knowledge base *The Grid* (Martín Arista 2013b), while adopting different approaches and scopes that will be discussed in turn. Among other data, *The Grid* includes a dictionary of Old English (*Nerthus*), a secondary source indexing dictionary (*Freya*) and indexed versions of the DOEC and the YCOE.

As for semi-automatic verb lemmatization, Metola Rodríguez (2015) and Tío Sáenz (2019) follow a similar method, and design sets of specific query strings (QS), guided by the specificities of the strong and weak verbal morphology. Their goal is to detect text strings in the DOEC and YCOE indexes compatible with the hypothetical inflectional forms of the corresponding verb classes. However, their research is limited by the possibilities of the search tools included in the storing database program. For example, some QS make use of wildcard symbols to replace strings of texts –e.g. (*) replaces an unspecified number of characters–, which results in the retrieval of a considerable number of undesired results This produces the need on the researchers' part to develop filters, and also increases the amount of manual revision needed to rule them out. To mention just a few, undesired results obtained from searches of the paradigm of *beadan* 'to command' include the forms *beada* 'counselor', and *beadas* 'tables'.

While they adopt similar search strategies, Metola Rodríguez (2015) and Tío Sáenz (2019) differ in the scope of their research in two ways. For one, Metola Rodriguez (2015; 2017) does not take participial forms into consideration. However, Tío Sáenz (2019) acknowledges the verbal nature of past and present participial adjectives and develops a particular QS to detect them. For another, Metola Rodríguez (2015; 2017) compares his data with the DOE while Tío Sáenz (2019) additionally checks her results against the YCOE. This allows her to lemmatize 6,300 inflectional forms of verbs I-Y not included in the DOE. At the same time, Tío Sáenz (2019: 544) admits the need for further validation of these forms with the corresponding fragments from the DOEC. As for accuracy, Metola Rodriguez (2017: 73) claims that, before manual revision, his method allows the validation of 80% of the proposed forms when compared to the inflectional forms provided by the DOE.

By contrast, García Fernández (2020) develops a system closer in nature to MG. She compiles a list of inflectional forms of simplex verbs based upon and attested in various grammars including Brunner (1965), Campbell (1987[1959]), and Hogg and Fulk (2011). To each of those forms, García Fernández (2020) attaches the different spelling variants of the prefixes listed by Kastovsky (1992) to develop complex inflectional forms.[1]

---

[1] See García Fernandez's (2020: 213-214) for an exhaustive list of inflectional forms of contracted verb derivatives.

The question that arises from this approach is the direct relation that García Fernández (2020) assumes between the simplex and the complex forms of the verbs. If complex forms are derived on the basis of the spelling of the attested simplexes, it is the case that the lack of a given simple form or a change to the spelling prevents the identification in the corpora of the corresponding derivative. To illustrate this question, consider the case of *beþearfst* 'you have need', which cannot be identified nor lemmatized in the paradigm of *beþurfan* because a form *þearfst* is not listed as an inflectional form of the verb *þurfan* 'to need'. The attested forms of *þurfan* include *ðearf, ðearfende, ðearft, ðorfeð, ðorfað, ðorfæð, ðorfte, ðorfton, ðuran, ðurfe, ðurfon, ðurfu, ðyrfe, þearf, þearfende, þearft, þorfende, þorfonde, þorfte, þorfton, þurfan, þurfe, þurfende, þurfon, þurfu, þurfun,* and *þyrfe* (García Fernández 2020: 133). However, the form is attested in the DOEC as shown in (1).

(1) [PsGlC (Wildhagen) 015600 (15.2)]

Ic sæde drihtne god min eart þu forþon goda minra þu ne <u>beþearfst</u>

'I have said to the Lord, thou art my God, for thou *hast* no *need* of my goods.' (Douay-Rheims, 1971[1899]: 586)

The reviewed works constitute, to the best of my knowledge, the only available research addressed at the automatic verb lemmatization of the OE corpora. Against the described scenario, Section 3 limits the scope of this research.

3. SCOPE

Considering previous research and the current state of publication of the DOE the scope of this study is limited to the letters L–Y, whose inflections have not been provided yet. This section delves into the features of Class III that allow the identification of form patterns upon which automatic MG processes can be developed.

The classification of OE verbs into seven classes based on the ablaut patterns of the verbal paradigms constitutes, according to von Mengden (2011), an almost undiscussed description of the grammar of this language. As for Class III, the taxonomy given in Figure 2 is generally accepted.

	Infinitive	Preterite 1	Preterite 2	Past Participle
IIIa	*bindan* 'bind'	*band*	*bundon*	*gebunden*
IIIb	*weorpan* 'cast'	*wearp*	*wurpon*	*geworpen*
IIIc	*stregdan* 'strew'	*strægd*	*strugdon*	*gestrogden*

Figure 2. OE Class III sub-classes.

Nevertheless, the state of affairs displayed in Figure 2 constitutes an oversimplified representation of the various types of OE class III verbs. To pave the ground for later discussion, it is necessary to offer a closer description of Class III subtypes.

Levin (1964), and Campbell (1987[1959]) distinguish five Class III subclasses, as seen in Figure 3.

	Ablaut pattern	Infinitive	Preterite 1	Preterite 2	Past Participle	
a	i-a-u-u	*bindan*	*band*	*bundon*	*(ge)bunden*	'bind'
b	e-ea-u-o	*helpan*	*healp*	*hulpon*	*(ge)holpen*	'help'
c	eo-ea-u-o	*weorpan*	*wearp*	*wurpon*	*(ge)worpen*	'throw'
d	u-ea-u-o	*spurnan*	*spearn*	*spurnon*	*(ge)spornen*	'spurn'
e	e-æ-u-o	*stregdan*	*strægd*	*strugdon*	*(ge)strogden*	'strew'

Figure 3. Levin's (1964) Classification of OE Class III Strong Verbs.

Class III verbs are characterized by presenting a single vowel and a consonant cluster in the stem. The phonological differences in the first consonant of the cluster justifies the first three subclasses given in Figure 3. Thus, Class IIIa verbs have a nasal sound after the vowel (closed to <i> from Proto-Germanic <e>); Class IIIb verbs present a liquid <l> after the vowel, which remains unchanged, and Class IIIc verbs display a liquid <r> which triggers breaking in the stem vowel. Class IIId includes the aorist present verbs *spurnan* 'spurn' and *murnan* 'mourn', and Class IIIe, for a group of verbs which were originally Class V. Class IIIe verbs display a syllabic pattern <e+C> (where C is neither a nasal nor a liquid). Their infinitives were extended with a dental suffix, as is the case with *stregdan* 'strew' and *feohtan* 'fight' (Campbell, 1987[1959]: 303).

Considering that the ablaut series described by Levin (1964) are identical for subclasses b, c, and d, but for the infinitive vowel, Marcin Krygier (1994) advances a 3 sub-class taxonomy. He postulates that Class IIIa verbs have a nasal sound after the vowel and display the *ablaut* pattern iNC – a/oNC – uNC – uNC; Class IIIb verbs display a liquid consonant <l, r> and the *ablaut* pattern elC/eorC/urC – eaLC – uLC – oLC; and Class IIIc includes those originally class V verbs described by Campbell (1987[1959]).

For the purpose of this research, a fine-grained distinction as the one proposed by Levin (1964) and Campbell (1987[1959]) is to be dismissed as the development and implementation of rules into the lemmatizer depend primarily on the existence of generalized patterns. Consequently, I shall follow Krygier's (1994) model for class assignment. With the alphabetical limitation of the study and class assignment thus established, this article analyses the following verbs: *limpan* (IIIa), *linnan* (IIIa), *melcan* (IIIb), *meltan* (IIIb), *murnan* (IIIb), *rinnan* (IIIa), *scelfan* (IIIb), *scellan* (IIIb), *sceorfan* (IIIb), *sceorpan* (IIIb), *scrimman* (IIIa), *scrincan* (IIIa), *seorpan* (IIIb), *sincan* (IIIa), *singan* (IIIa), *sinnan* (IIIa), *slincan* (IIIa), *slingan* (IIIa), *smeortan* (IIIb),

sneorcan (IIIb), *spinnan* (IIIa), *springan* (IIIa), *sprintan* (IIIa), *spurnan* (IIIb), *steorfan* (IIIb), *stincan* (IIIa), *stingan* (IIIa), *stregdan* (IIIc), *swelgan* (IIIb), *swellan* (IIIb), *sweltan* (IIIb), *sweorcan* (IIIb), *sweorfan* (IIIb), *swimman* (IIIa), *swincan* (IIIa), *swindan* (IIIa), *swingan* (IIIa), *teldan* (IIIb), *tingan* (IIIa), *tringan* (IIIa), *twingan* (IIIa), *þerscan* (IIIc), *þindan* (IIIa), *þingan* (IIIa), *þringan* (IIIa), *þrintan* (IIIa), *weorpan* (IIIb), *weorþan* (IIIb), *windan* (IIIa), *winnan* (IIIa), and *wringan* (IIIa).[2]

4. METHOD

This section unfolds the methodological underpinnings of this research. There are two basic processes, namely, the generation of plausible inflectional forms for each of the selected verb paradigms and the automatic identification of the generated forms in the authoritative corpora. Given the limitations in the available data described in Section 1, ALOEV3 generates a remarkable amount of forms that cannot be attested in the corpora.

To start with, ALOEV3 generates complete verb paradigms on the basis of attested inflectional patterns. The paradigm of *sneorcan* 'shrivel' is given in (2).

(2)

Infinitive	sneorcan	Pres. subj. (sg.)	sneorce
Inflected Infinitive	sneorcenne	Pres. subj. (pl.)	sneorcen
Pres. ind. (1ˢᵗ sg)	sneorce	Pret. subj. (sg.)	snurce
Pres. ind. (2ⁿᵈ sg.)	sneorcest	Pret. subj. (pl.)	snurcen
Pres. ind. (3ʳᵈ sg)	sneorceþ/sneorceð	Pres. part.	sneorcend
Pres. ind. (pl.)	sneorcaþ/sneorcað	Past part.	snorcen
Pret. ind. (1ˢᵗ/3ʳᵈ sg.)	snearc	Imperative (sg.)	sneorc
Pret. ind. (2ⁿᵈ sg.)	snurce	Imperative (pl.)	sneorcaþ/sneorcað
Pret. ind. (pl.)	snurcon		

The inflectional paradigm in (2) does not show well attested changes involving vocalic alternation, syncopation, assimilations, Verner's law, of the simplification of consonant clusters. It provides, however, a basic schema upon which MG rules can be implemented to account for variation in (i) the inflectional endings, including the assimilation and simplification of consonants, the weakening of vowels in unaccented syllables as well as diatopic spelling varieties and (ii) the stem, where *i*-mutation and Verner's law operate.

Given the high degree of formal variation in OE, the MG of inflectional forms has been addressed in three stages, namely inflection (4.1), mutation (4.2) and derivation (4.3). To round off, section (4.4) describes the automatized attestation process.

---

[2] The grapheme <þ> rather than <ð> has been selected for lemma representation.

4.1. MG OF CLASS III VERBS: INFLECTIONAL FORMS

ALOEV3 has been implemented with MG rules to generate inflectional endings subject to formal variation. For the development of the rules given in Figure 4 below, I draw on Campbell (1987: 299–300).

Rule No[3]	Rule description	Inflectional form	Rule No	Rule description	Inflectional form
I_#1	-e >-æ	Pres. ind. (1st sg)	I_#17	-tt > -t	Pres. ind. (3rd sg)
I_#2	-e >-o	Pres. ind. (1st sg)	I_#18α	-gþ > hþ	Pres. ind. (3rd sg)
I_#3	-e >-u	Pres. ind. (1st sg)	I_#19α	-ngþ > -ncþ	Pres. ind. (3rd sg)
I_#4*	-est > -st	Pres. ind. (2nd sg.)	I_#20α	-sþ > -st	Pres. ind. (3rd sg)
I_#5*	-dst > -tst	Pres. ind. (2nd sg.)	I_#21α	-þþ > t	Pres. ind. (3rd sg)
I_#6	-tst > -st	Pres. ind. (2nd sg.)	I_#22α*	-eþ > -iþ	Pres. ind. (3rd sg)
I_#7α*	-þst > -sst	Pres. ind. (2nd sg.)	I_#23α	-eþ > -it	Pres. ind. (3rd sg)
I_#8	-sst > -st	Pres. ind. (2nd sg.)	I_#24α	-iþ > -it	Pres. ind. (3rd sg)
I_#9	-ngst > -ncst	Pres. ind. (2nd sg.)	I_#25	-on >-an	Pret. ind. (pl.)
I_#10*	-gst > -hst	Pres. ind. (2nd sg.)	I_#26	-on >-un	Pret. ind. (pl.)
I_#11	-hst > -xt	Pres. ind. (2nd sg.)	I_#27	-enne > -anne	Inflected Inf.
I_#12	-est > -ist	Pres. ind. (2nd sg.)	I_#28	-enne > -onne	Inflected Inf.
I_#13α	-þs > -ts	Pres. ind. (2nd sg.)	I_#29*	-end > -and	Pres. participle
I_#14α*	-eþ > -þ	Pres. ind. (2nd sg.)	I_#30*	-end > -ind	Pres. participle.
I_#15α*	-tþ- > tt	Pres. ind. (3rd sg)	I_#32*	-en > -in	Past participle.
I_#16α	dþ > tt	Pres. ind. (3rd sg)	I_#31*	-en > -æn	Past participle.

Figure 4. MG rules for Class III alternative inflections.

Each of the rules in Figure 4 addresses a specific process of formal variation. Several processes may operate upon the same inflectional form. Thus, rules I_#1, I_#2 and I_#3 apply to the 1st person singular of the present indicative forms. MG processes can be recurrent or non-recurrent. A non-recurrent MG process implies the application of a MG rule which puts an end to the MG process. In recurrent MG processes, a form generated by a specific MG rule re-enters the system for further rules to operate upon it.[4] More precisely, the forms generated by I_#4 are processed again and the ALOEV3 checks whether I_#5, I_#7, I_#9 or I_#10 need to be applied.

Rules I_#29 through I_#31 deserve further comment. These rules generate present and past participle forms which, in OE, can be inflected following the weak and strong adjectival paradigms. In line with Campbell (1987[1959]: 266–272), rules I_#32-I_#75 have been generated for ALOEV 3 to attach the endings ø; *-ne*; *-es*; *-um*; *-e*; *-ra*; *-u*; *-re*; *-a*; *-an*; and *-ena* to the generated participial forms.

---

[3] ALOEV 3 is implemented with pairs of rules (α, ß) to account for the conventional <þ, ð> variation. For the sake of clarity, only α-rules are displayed in this article.

[4] Rules starting a recurrent MG process are marked (*). The notation system is consistent in Figures 5, 6 and 7 below.

4.2. MG OF CLASS III VERBS: STEM MUTATION

Three sets of rules have been designed to account for vocalic and consonantal changes taking place in the verb stem. Such changes include *i*-mutation, Verner's law, and the simplification of consonantal clusters among others. Evidence supporting these rules is provided by Krygier (1994: 44–49) and Campbell (1987[1959]: 310–312). These rule sets are class-specific and only operate on the MG inflectional forms belonging in the corresponding sub-class.[5] Figures 5, 6, and 7 display the rules designed for Class IIIa, IIIb and IIIc mutations respectively.

Rule No	Rule description	Rule operation
M_#1a	-mm > -m	Preterite singular forms
M_#2a	-nn > -n	Preterite singular forms
M_#3a	-a- > -o-	Preterite 1 forms
M_#4a	-u > -i	Extension of present stem vowel -*i*- to preterite plural
M_#5a	-o > -i	Extension of present stem vowel -*i*- to past participle
M_#6a*	-u- > -iu-	Preterite stem vowel after initial *sc-*
M_#7a	-iu- > -y-	Preterite stem vowel after initial *sc-*
M_#8a*	ri- > ir-	Methathesis of -*r- rinnan*
M_#9a	ir>ier	Stem vowel variation
M_#10a	ir> yr	Stem vowel variation
M_#11a	ir> eor	Stem vowel variation
M_#12a	ir> ior	Stem vowel variation
M_#13a	-u- > -a-	Extension of preterite 1 -*a*- to preterite 2 forms
M_#14a	-u- > -o-	Extension of preterite 1 -*o*- to preterite 2 forms
M_#15a	sw- > su-	Consonant cluster simplification

Figure 5. MG rules for Class IIIa mutations.

Rule No	Rule description	Rule operation
M_#1b	-ea- >-*eo*	Analogical extension to preterite 1 forms
M_#2b	-ea +l- > -a+l-	Diatopic variation
M_#3b*	*-eo- > -e-*	Before front vowels
M_#4b	*-e- > -i-*	*i*-mutation
M_#5b	*-eo- > -io*	*i*-mutation
M_#6b	*-u- > -y-*	*i*-mutation
M_#7b*	*-e- > -ie-*	Palatization after inicial *sc-*
M_#8b	*-ie. > -i-*	*i*-mutation
M_#9b	*-ie- > -y-*	*i*-mutation

Figure 6. Inflectional rules for Class IIIb mutations.

---

[5] While phonological distinctions like vowel quantity are crucial to explain the diachronic evolution of English, they are not relevant to the identification of forms in the corpora. Consequently, such differences have been disregarded in the configuration of stem mutation rules.

Rule No	Rule description	Rule operation
M_#1c	-Vowel+*r*- > -*r*+Vowel-	Reverse metathesis
M_#2c	-sc- > -cs-	Metathesis
M_#3c	-sc- > x	Consonant cluster simplification
M_#4c	-g- > ø	elision before *þ*, *d*, and *n*.

Figure 7. Inflectional rules for Class IIIc mutations.

4.3. MG OF CLASS III VERBS: DERIVATION

The application of the inflection and mutation rule sets described in sections 4.1 and 4.2 allows for the generation of full paradigms of the simplexes of the verbs under analysis. This section addresses the generation of their complex counterparts.

Complex forms of the selected verb lemmas and of the MG inflectional forms are generated by attaching the set of L–Y prefixes described by Metola Rodríguez (2015) and García Fernández (2020) to each of the MG forms. In spite of lying out of the alphabetical scope of this research, the prefix *ge-* has also been included in the inventory, given its salient role in the formation of past participle forms.[6] OE prefixes are also subject to spelling variation, which is the reason why a rule has been designed for each prefix formal variant. As the aim of this article is to advance in the lemmatization of OE, these prefixes have been classified in canonical (lemma) and non-canonical forms. (3) shows the canonical form and non-canonical forms (in brackets) of the selected prefixes.

(3)   ge-(cg-, g-, ga-, gæ-, gæn-, gær-, gad-, gan-, gar-, ged-, gen-, gem-, ger-, gi-, gif-, gim-, gy-); med-(me-, met-, mi-, mid-, mið-, miþ-, mod-); mis-(miss-, mus-); niþer-(neoþer-, nioþer-, nyþer-, nieþer-, niþor-, niðer-, neoðer-, nioðer-, nyðer-, nieðer-, niðor-); o-; of-(æf-, af-, off-); ofer-(eofer-, eofor-, ofær-, ofern-, ofor-, of'-, ofyr-, ouer-, ouyr-); on-; or-; oþ-(oeþ-, oð-, oeð-); onweg-(anweg-, aweg-, unweg-); riht-(reht-, reoht-, rieht-, ryht-); sam-; sin-; sub-; to-; twi-(twig-, twy-); þri-(þry-, þrie-, ðri-, ðry-, ðrie-); þurh-(þorh-, ðurh-, ðorh-); 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-).

Summarizing, ALOEV3 has been designed to generate standardized Class III verb paradigms. Rules have been implemented in the system to account for formal variation and word formation processes. As a result, a pool of potential OE class III verb forms is developed. Each of the forms is provided with a lemma tag, as shown by Figure 1 above.

---

[6] The automation of the process precludes the development of target specific rules that guide the attachment of prefixes to particular sets of forms. Consequently, the application of the derivational MG rules results in the MG of complete *ge-* prefixed paradigms.

Before moving onto the next stage, MG forms must be reduced to types. In the MG process, homographic, formally ambiguous forms have been generated both within and across paradigms. Consequently, intra-paradigmatic homographs must be ruled out. To do so, ALOEV3 searches each paradigm for duplicated MG forms and reduces them to one single occurrence, as shown by Figure 8.
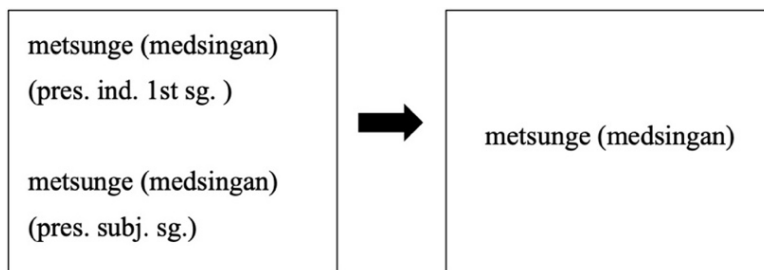

Figure 8. ALOEV3's 1-type-per-paradigm reduction model.

At the same time, the process described in Figure 8 allows for the presence of formally ambiguous forms in different paradigms. Thus, the forms *metsunge* (medsingan) and *metsunge* (medswingan). The identification in the corpora of these ambiguous forms will give raise to instances of lemma competition. The implications hereby derived will be discussed in section 5.

4.4. ASSESSING THE ATTESTATION OF FORMS IN THE CORPORA

The final part of the method tackles the identification of the MG forms in the selected corpora. Such identification is automatically carried out by means of three interrelated databases. To wit, (i) ALOEV3_DB, a database filing ALOEV3's generated form types; (ii) DOEC_DB, which stores the indexed version of the DOEC (Healey *et al*. 2009) retrieved from Martín Arista's (2013b) *The Grid,* and (iii) YCOE_DB, a database filing those words that are labelled with a verbal POS –part of speech– tag in the YCOE (Taylor *et al*. 2003). A list of POS tags and their meaning is given in Appendix 1.

ALOEV3_DB includes a field for the MG form (Inflectional form), a field for the lemma from which the form has been generated (Class III 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 (Tag summary) for forms attested in the YCOE. DOEC_DB incorporates 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). YCOE_DB present a field for the inflectional form in the YCOE (YCOE_verbal_form) and a field for the different verbal POS tags assigned to such form (YCOE_verbal_tag). Figure 9 offers a view of *limpe* in each of the databases.

Inflectional form	limpe				Class III Lemma	limpan	

DOEC_attestation [X] YES [ ] NO

YCOE_attestation [X] YES [ ] NO

Tag summary | VBPS

Prefield	Conc_Term	Postfield
ænig word cweþe, butan hit swa	limpe	, þæt hwylc cuma unmyndluncga cume and se abbod f
ce scylen. Gif þe wyrt drenc ne	limpe	sele strangne. Leohte mettas þicge & geseaw broþu
f gelimpat. werpinge. forlæten.	limpe	. ster þu. þær of. fordeme þu. wedlaces. þancfulre

YCOE_verbal...	YCOE_verbal_tag 1	YCOE_verbal_tag 2	YCOE_verbal_tag 3	YCOE_verbal_tag 4	YCOE_verbal_tag 5
limpe	VBPS				
limpende	VAG				
limpeð	VBPI				
limpeþ	VBPI				
limpð	VBPI				

Figure 9. A view of *limpe* in ALOEV3_DB, DOEC_DB and YCOE_DB.

The three databases are related to one another through the fields *Inflectional form*, *ConcTerm,* and YCOE_verbal_form. Consequently, if there is a spelling coincidence between the form in Inflectional form and ConcTerm and/or the YCOE_verbal_form, the corresponding DOEC_attestation and/or YCOE_attestation fields (YES) are activated. Such is the case of *limpe* in Figure 9, whose attestation is confirmed in both corpora. Figure 10 shows these relationships.
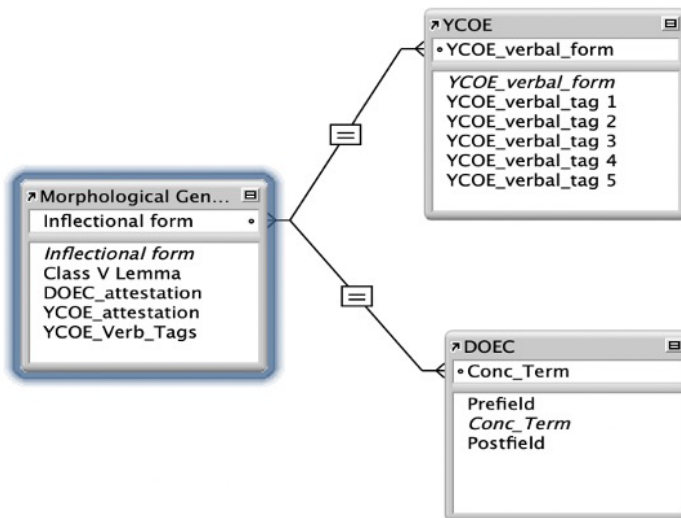

Figure 10. Relational interface of the ALOEV3_DB, DOEC_DB and YCOE_DB.

5. RESULTS AND DISCUSSION

ALOEV3 has generated 1,101,555 form types, which, after being filed in ALOEV3_DB, are compared with the inflectional forms filed in DOEC_DB and YCOE_DB. On the quantitative side, such comparison turns out the identification of 571 types in the DOEC belonging to 150 distinct lemmas (Appendix 2a); 653 in both the DOEC and the YCOE corresponding to 121 different lemmas (Appendix 2b), and 31 types from 14 lemmas which are only attested in the YCOE (Appendix 2c). The identified types are grouped by lemma, which may or may not be itself attested. When the generated form is identified in the YCOE, the POS tag is provided alongside. (4) offers an account of the forms of *þerscan* 'to strike, beat' identified in the corpora.

(4) DOEC: *þerscenne, þrescen, þrexe, þreax, ðurscon, ðerscende*.

DOEC and YCOE: *þerscan* (VB), *þirsceð* (VBPI), *þærsc* (VBDI), *ðerscað* (VBI), *ðærsc* (VBDI).

YCOE: *ðerscan* (VB).

All in all, 1,256 form types have been identified in the corpora, corresponding to 192 distinct lemmas. On the qualitative side, 1,218 forms are assigned a distinct lemma. This means that the MG lemmatizer reaches 96,97% efficiency when assigning a single lemma to Class III strong verb inflectional forms, a figure which constitutes a 17% percent improvement with respect to Metola Rodriguez's (2017) approach. The remaining 38 forms are duplications which show competition among lemmas. The competing lemmas and inflectional forms are given in (5).

(5) **gesingan ~ geswingan**: *gesungæn, gesunge, gesungen, gesungena, gesungene, gesungenne, gesungenre, gesungenum, gisunge*; **medrinnan ~ murnan**: *mearn*; **medsingan ~ medswingan**: *metsunge*; **stincan ~ stingan**: *stincð, stincþ*; **singan ~ swingan**: *sunge*, sungen, sungene; sungenne; **swincan ~ swingan**: *swincst, swincð*.

As seen in (5), most of the lemma competition occurrences involve the lemas *singan* 'to sing' and *swingan* 'to beat, strike' and their derivatives. Overlapping cases mostly occur in the preterite and past participle forms. Campbell (1987[1959]: 310) attests the occurrence of a past participle *sungen* as well as a form *swungen* in the Martyrology, which justifies the implementation of the rule *-w- > -u- > ø* when followed by *-u-* to the lemmatizer. The other cases in conflict involve the devoicing of /g/ into /k/ through assimilation processes, thus giving way to homophonic and homographic forms for the verbs *stincan~stingan* and *swincan~swingan*. The last case in competition corresponds to the inflectional form *mearn,* which is assigned to both the prefixed lemma *medrinnan* (through the syncopated form of the prefix; metathesis of *rinnan* into *irnan* and the retracted form of the preterite instead of

the most common *-ea-* diphthong) and to the aorist present *murnan*. The complexity of the processes justifying the association of *mearn* to the lemma *medrinnan* suggests it is not a correct association. Indeed, a quick search in the DOEC reveals that there are only six occurrences of this inflectional form, four of which appear in Beowulf. In all cases, the verb at stake is *murnan* 'to care, be anxious, mourn'. Take (6) as illustration.

(6) [Beo 041200 (1441)]

Gyrede hine Beowulf eorlgewædum, nalles for ealdre mearn.

'Beowulf made himself ready with noble armor, he didn't *mourn* for his life' (Hostetter, n.d.)

This and other examples justify the need for additional manual revisions of the results. A fully automated method cannot be put forward for two main reasons. The limited scope of the research and the type-based approach adopted.

As for the scope of the research, only Class III strong verbs are being analysed in this article. OE morphological features and spelling variation lead to the development of formally ambiguous forms, not only within a single paradigm of a given lexical class, as shown in (2) above, but also across paradigms across lexical classes. As an example, consider the form *gelimpe*, which ALOEV3 generates in the paradigm of *gelimpan*. Such form, however, can also be associated to the nominal paradigm of the neuter noun *gelimp* 'an event, accident'. In such cases, disambiguation and lemma assignment must come from token-based lemmatization.

The second reason is that type-based lemmatization is context independent and, consequently, semantic features and syntagmatic relationships cannot be taken into account. ALOEV3 assigns a lemma to a form-type, but disregards the amount of occurrences as well as the context(s) in which such form is attested. However, type-based lemmatization is useful when it comes to identifying forms in conflict and areas of formal overlapping.

In the case of *gelimpe*, there are 58 occurrences attested in the DOEC. The contextual analysis of these forms will prove whether the form-lemma association established by ALOEV3 between *gelimpe* and *gelimpan* holds true in each context. As illustration, (7) shows a confirmed association.

(7) [BenR 027200 (11.36.1)]

[...], *butan hit færlice swa* gelimpe, [...]

[...] unless it *happen* by chance [...] (Riyeff 2017: 67)

Once the association is confirmed, the lemma as well as any other relevant inflectional information can be incorporated by the lexicographer into a lemmatized corpus such as ParCorOEv2. The combination of both tools represents a remarkable advance in the process of lemmatizing the existing OE corpora. ALOEV3 establishes a wide net of relations between forms and lemmas, while the contextual analysis provided by ParCorOEv2 can confirm the accuracy of the association. When multiple lemmas are associated to a given form type, the lexicographer is provided with a closed set of lemmas to choose from, which guides the manual revision and lemma selection processes, thus enhancing the overall efficiency of the method.

6. CONCLUSIONS

NLG greatly depends on the existence of large, lemmatized corpora. Given that the extant OE corpora are not lemmatized, there is a great gap to bridge before considering OE as a candidate language to conduct NLG research on an extensive basis. This article has presented a way to fill this gap, by exploring and checking the possibilities to carry out a systematic and automatized type-based lemmatization process of the OE corpora. A semi-automatic process of lemmatization has been proposed, rooted in the theoretical framework of MG. The research has been guided by three main aims, namely automation, validation, and accuracy. Regarding automation, several issues have arisen, especially regarding the spelling and morphological features of OE. Nevertheless, the article has proven that type-based lemmatization can be largely automatized, thus speeding and systematizing the lemmatization of OE. As for validation, the proposed method generates plausible OE word forms and compares them with the major OE corpora. This constitutes a step forward regarding previous research on OE morphology based on dictionaries. As for accuracy, the inferential approach followed here results in the identification of specific forms in the corpora which are mostly assigned a single lemma. Thus, the amount of manual revision is greatly reduced.

In this exploration of non-manual lemmatization, the limits of automation have also been reached in the current state of affairs. While type-based lemmatization can be automatized to a large extent, the spelling and morphological properties of OE prevent the MG analysis from systematizing and implementing rules to account for all the spelling variants found in the corpora and assign a distinctive lemma to all the validated forms. However, as research advances through the strong verb classes, light will be shed upon this matter, in such a way that the disambiguation of homographic forms will be possible or, at least, by highlighting those instances that require contextual analysis.

This article opens two avenues of research. The first is the completion of the analysis of Class III strong verbs. There are two major groups of verbs that have been disregarded. First, only the letters L-Y have been considered. Second, this article has dealt with simplex and single prefixed verbs. This means that cases of recursive prefixation like *inbestingan* 'to penetrate' or *onawinnan* 'to fight against'

have been left out. The second avenue involves extending this method to other strong verb classes and later to other variable lexical categories.

REFERENCES

1971 (1899). *The Holy Bible Translated from the Latin Vulgate* (Douay Rheims Version). London: Tan books and publishers.

Biber, D., S. Conrad and R. Reppen 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press

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

Brunner, K. 1965. *Altenglische Grammatik, nach der Angelsachsischen Grammatik von Eduard Sievers*. Berlin: Max Niemeyer.

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.

Ferrés, D., A. AbuRa'ed and H. Saggion. 2017. "Spanish morphological generation with wide-coverage lexicons and decision trees". *Procesamiento del Lenguaje Natural* 58: 109-116.

Forcada, M.L., M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J.A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez and F. Tyers. 2011. "Apertium: A free/open-source platform for translation". *Machine Translation* 25 (2): 127-144.

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

García García, L. 2019. "The basic valency orientation of Old English and the causative *ja*- formation: A synchronic and diachronic approach". *English Language and Linguistics* 24 (1): 153-177. http://doi.org/10.1017/S1360674318000345

García García, L. and E. Ruiz Narbona. 2021. "Lability in Old English verbs: Chronological and textual distribution". *Anglia: Journal of English Philology* 139 (2): 283-326. http://doi.org/10.1515/ang-2021-0022

Healey, A., ed., with J. Price and X. Xiang. 2009. *The Dictionary of Old English Web Corpus*. Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

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

Hogg, R. and R.D. Fulk. 2011. *A Grammar of Old English*. Oxford: Wiley-Blackwell.

Hostetter, A.K. n.d. *The Old English Narrative Poetry Project*: *Beowulf*. Accessed September 2. https: //oldenglishpoetry.camden.rutgers.edu/beowulf/

Kastovsky, D. 1992. "Semantics and vocabulary". *The Cambridge History of the English Language I: The Beginnings to 1066*. Ed. R. Hogg. Cambridge: Cambridge University Press. 290-408.

Khemakhem, A., B. Gargouri, A. Ben Hamadou and G. Francopoulo. 2015. "ISO standard modeling of a large Arabic dictionary". *Journal of Natural Language Engineering* 22 (6): 849-879.

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

Levin, S.R. 1964. "A reclassification of the Old English strong verbs". *Language* 40 (2): 156-161.

Manjavacas, E., A. Kádár and M. Kestemont. 2019. "Improving lemmatization of non-standard languages with joint". *Proceedings of NAACL-HLT 2019*. Ed. J. Burstein, Doran, Christy and T. Solorio. Minneapolis: ACL. 1493-1503.

Martín Arista, J. 2012a. "Lexical database, derivational map and 3D representation". *RESLA-Revista Española de Lingüística Aplicada* (*Extra 1*): 119-144.

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

Martín Arista, J. 2013a. "Recursivity, derivational depth and the search for Old English lexical primes". *Studia Neophilologica* 85(1): 1-21. http://doi.org/10.1080/00393274.2013.771829.

Martín Arista, J. 2013b. Nerthus. Lexical database of Old English: From word-formation to meaning construction. Research Seminar, School of English, University of Sheffield.

Martín Arista, J., ed. 2016. *NerthusV3. Online Lexical Database of Old English*. Nerthus Project, Universidad de La Rioja. www.nerthusproject.com.

Martín Arista, J. 2017a. "El paradigma derivativo del inglés antiguo". *Onomázein* 37: 144-169.

Martín Arista, J. 2017b. The design and implementation of a pilot parallel corpus of Old English. Paper presented at the SHELL Session of the 2017 International Medieval Conference, University of Leeds, July 4.

Martín Arista, J. 2018. "The semantic poles of Old English. Toward the 3D representation of complex polysemy". *Digital Scholarship in the Humanities* 33 (1): 96-111.

Martín Arista, J. 2019. "Another look at Old English zero derivation and alternations". *Atlantis* 41 (1): 163-182.

Martín Arista, J. 2020a. "Old English rejoice verbs. Derivation, grammatical behaviour and class membership". *Poetica* 93: 133-153.

Martin Arista, J. 2020b. "Further remarks on the deflexion and grammaticalization of the Old English past participle with *habban*". *International Journal of English Studies* 20 (1): 51-71.

Martín Arista, J. 2021a. "The syntax and semantics of the Old English predicative construction"., *Language Change and Linguistic Theory in the 21st Century*. Eds. N. Lavidas and K. Nikiforidou. Leiden: Brill. Forthcoming.

Martín Arista, J. 2021b. "Word alignment in a parallel corpus of Old English prose. From asymmetry to inter-syntactic annotation". *Corpora in Translation Research: Recent Advances and Applications*. Eds. J. Lavid-López, C. Maíz-Arévalo and J.R. Zamorano. Amsterdam: John Benjamins. 76-100.

Martín Arista, J. and A.E. Ojanguren López. 2018. Doing electronic lexicography of Old English with a knowledge-base. Workshop delivered at the Consolidated Library of Anglo-Saxon Poetry (CLASP) Project, Faculty of English Language and Literature of the University of Oxford.

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

Mateo Mendaza, R. 2014. "The Old English adjectival affixes *ful-* and *-ful*: a text-based account on productivity". *NOWELE-North-Western European Language Evolution* 67 (1): 77-94. http://doi.org/10.1075/nowele.67.1

Mateo Mendaza, R. 2015. "Matching productivity indexes and diachronic evolution. The Old English affixes *ful-*, *-isc*, *-cund* and *-ful*". *Canadian Journal of Linguistics* 60 (1): 1-24.

Mateo Mendaza, R. 2016a. "The search for Old English semantic primes: The case of HAPPEN". *Nordic Journal of English Studies* 15: 71-99.

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

von Mengden, F. 2011. "Ablaut or transfixation? On the Old English strong verbs". *More than Words: English Lexicography and Lexicology Past and Present*. Eds. R. Bauer and U. Krischke. Frankfurt am Main: Peter Lang. 123-139.

Metola Rodríguez, D. 2015. *Lemmatisation of Old English strong verbs on a lexical database*. Unpublished Ph. D. thesis. University of La Rioja, Spain.

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.

Novo Urraca, C. 2015. "Old English deadjectival paradigms. Productivity and recursivity". *NOWELE-North-Western European Language Evolution* 68 (1): 61-80.

Novo Urraca, C. 2016a. "Old English suffixation. Content and transposition". *English Studies* 97: 638-655.

Novo Urraca, C. 2016b. "Morphological relatedness and the typology of adjectival formation in Old English". *Studia Neophilologica* 88 (1): 43-55.

Novo Urraca, C. and A. E. Ojanguren López. 2018. "Lemmatising treebanks. Corpus annotation with knowledge bases". *RAEL: Revista electrónica de Lingüística Aplicada* 17 (1): 99-120.

Oflazer, K. and M. Saraçlar, eds. 2018. *Turkish Natural Language Processing*. Cham: Springer. http://doi.org/10.1007/978-3-319-90165-7

Ojanguren López, A. E. 2020. "The semantics and syntax of Old English end verbs". *Atlantis* 42 (1): 163-188.

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

Reiter, E. and R. Dale. 1997. "Building applied natural language generation systems". *Natural Language Engineering* 3 (1): 57-87. http://doi.org/10.1017/S1351324997001502

Rissanen, M., M. Kytö, L. Kahlas-Tarkka, M. Kilpiö, S. Nevanlinna, I. Taavitsainen, T. Nevalainen, and H. Raumolin-Brunberg, comp. 1991. *The Helsinki Corpus of English Texts*. Department of English, University of Helsinki.

Riyeff, J. 2017. *The Old English rule of Saint Benedict with related Old English texts*. Collegeville: Liturgical press.

Taylor, A., A. Warner, S. Pintzuk and F. Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York: University of York.

Tapsai, C., H. Unger and P. Meesad. 2021. *Thai Natural Language Processing*. Cham: Springer.

Tío Sáenz, M. 2019. *The lemmatisation of Old English Weak verbs of a relational database*. Unpublished Ph. D. thesis. University of La Rioja: Spain.

Vea Escarza, R. 2013. "Old English adjectival affixation. Structure and function". *Studia Anglica Posnaniensia* 48 (2-3): 5-25.

Vea Escarza, R. 2016a. "Recursivity and inheritance in the formation of Old English nouns and adjectives". *Studia Neophilologica* 88: 1-23. http://doi.org/10.1080/00393274.2015.1049830

Vea Escarza, R. 2016b. "Old English affixation. A structural-functional analysis". *Nordic Journal of English Studies* 15 (1): 101-119.

Vea Escarza, R. 2018. "Las funciones y categorías de los nombres y adjetivos afijados del inglés antiguo". *Onomázein* 41: 208-226.

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

APPENDIX 2. ATTESTED FORMS BY LEMMA

 a) Attestations in the DOEC

gelimpan: *gelumpene, gelimp, gilimpe, gilamp;* **gemeltan**: *gemilteþ, gemolten, gemeltende;* **gerinnan**: *gerunnenes, gerunnene, gerunnon, geornne, grin, garan, garon, gæran, gæron, ganran, genran, genarn, giarn, gifran, gyran;* **gesceorfan**: *gesceorf;* **gescrincan**: *gescruncenan, gescriuncan, gescryncan;* **gesincan**: *gesonc;* **gesingan**: *gesungenne, gesungenum, gesungenre, gesungena, gesungæn, gesingen, gesingað, gesunge, gesungan, gesungun, gesingon, gisingeð, gisunge, gisang;* **gesinnan**: *gesan;* **gesneorcan**: *gesnerc;* **gespinnan**: *gespunnen, gespann;* **gespurnan**: *gespurnan, gespearn;* **gestincan**: *gestince, gestanc;* **gestregdan**: *gestread;* **geswelgan**: *geswelge;* **geswellan**: *geswelle, geswell, geswollene;* **gesweltan**: *giswelte;* **gesweorcan**: *gesweorceð, gesweorc, geswerc, geswarc, gesworcen;* **gesweorfan**: *geswyrfeþ, gesweorf;* **geswimman**: *giswom;* **geswincan**: *geswincendra, geswincan, geswincu, geswinc, geswanc, geswuncon;* **geswingan**: *gesungenne, gesungenum, gesungenre, gesungena, gesungæn, giswungen, geswing, geswang, gesunge, gesungæn, gisunge;* **geteldan**: *getelde, geteld;* **geþerscan**: *geþersce, geðurscon, geðorscen, giðorscen, giðorscenne;* **geþindan**: *geþind;* **geþingan**: *geþungenne, geþungenum, geþungene, geþungenra, geþungenre, geðungenes, geðungenre, geðungænra, giðingende, giðingande, geþinge, geþingo, geþingu, geþincð, geþange, geþang, geðinge, geðincð, geðang;* **geþringan**: *geðrungen, geþring, geþrange, geþrong, geþrungon, geðringað, geðring, giðringað, giðring, giðrungun;* **geweorpan**: *geweorpe, gewerpð, geweorp, gewarp, gewurpon, geworpun, gewurpen, geworpen, geworpene, giwarp, giwurpon, giwurpun, giworpen;* **geweorþan**: *geweorþæ, geweorþen, gewerþen, geweorþ, gewerþ, gewurþen, geworþen, gewordenne, gewordena, geweorðaþ, geweorðen, geweorð, gewerð, gewurðen, geworðen, geworðene, giwurdun, giworden, giwarð;* **gewindan**: *gewundene, gewinde, gewind, gewande, gewundon, gewundun, giwundun;* **gewinnan**: *gewunnenan, gewinnene, gewinne, gewinnæ, gewinnu, gewinn, garwan, giwinne;* **gewringan**: *gewringe;* **limpan**: *limpan, limpað;* **linnan**: *linnenne, linne, linnen, lan, lon;* **medlinnan**: *melan, milan;* **medrinnan**: *meran;* **medsingan**: *midsingend, metsunge;* **medsinnan**: *misan;* **medswingan**: *metsunge;* **medwinnan**: *medwan;* **melcan**: *melcð, milciþ, milcet, milcit, melc;* **meltan**: *melteþ, milteþ, milt, melteð, malt;* **mislimpan**: *mislimp;* **mislinnan**: *muslan;* **miswinnan**: *miswunne;* **murnan**: *murneð, murnð, murnon;* **niþerrinnan**: *niðerran, neoðerran, nioðerran, niþerran;* **oferþingan**: *oferþunge;* **oferweorpan**: *oferweorpan, oferweorp;* **oferwinnan**: *oferwunnenum, oferwinnendum, oferwinnende, oferwinnaþ, oferwinn, ouerwunnan;* **oflinnan**: *oflan;* **ofrinnan**: *ofran, afran;* **ofsinnan**: *æfsan;* **ofspringan**: *ofspringe, ofspring, ofsprunge;* **ofspurnan**: *ofspurne;* **ofsweorcan**: *æfsweorc;* **onlinnan**: *onlan, onlon;* **onrinnan**: *onran;* **onsingan**: *onsong;* **onspurnan**: *onspurnan, onspurneð, onspurnað;* **onstingan**: *onstinge, onstingð, onsting;* **onstregdan**: *onstregdst, onstridest;* **onþingan**: *onþungan;* **onþringan**: *onþrong, onðrungun;* **onweorpan**: *onwearp;* **onwindan**: *onwindeð;* **onwinnan**: *onwunnen, onwinnendes, onwinnendre, onwinneð, onwunnon, onwunnan, onwunnun;* **orinnan**: *oran;* **orweorþan**: *orwurðe;* **orþingan**: *orþunge, orðunge;*

oþrinnan: *oþron, oðron, oðeorn;* oþþringan: *oðþringan;* oþwindan: *oþwinde;* rihtrinnan: *rihtran, ryhtran;* rinnan: *runnen, rinnende, iernenne, yrnnende, rinnan, rinne, rann, ran, irnn, urnne, yrn;* sammeltan: *sammeltan;* samsinnan: *samson;* samsteorfan: *samstorfenne;* samwinnan: *samwinnendan;* sceorfan: *scurfe;* sceorpan: *scerpe, scerpest, scerpeþ, scerþð, sceorp, scearp, scarp;* scielfan: *scylfan, scylfe, scelf, scylf;* sciellan: *scelle, scille, scylle, sciell, scill, scyll, sculle, scullon, scullen, scyllende, scyllendre;* scrimman: *scrimme;* scrincan: *scrincen, scrince, scrinceð, scrincon;* seolcan: *seolce, seolcen, seolc, solcenan;* sincan: *sincende, since, sanc;* singan: *sungenne, singæn, singendne, singendre, singendan, singande, singæ, singu, singþ, sange, song;* sinnan: *sunnen, sunnæn, sinne, sinnæ, sinn, sunne, sann, san, son, sunnon, sunnan;* slincan: *slincendes, slincendum, slincende, slincendra, slincendu;* slingan: *slingð, slang;* spinnan: *spinne, spinnaþ, spunne, spanne, spunnun;* springan: *sprungene, springendum, springende, springan, spring;* spurnan: *spurneð, spurnen, spurnende;* stincan: *stincend, stincendre, stincþ, stunce;* stingan: *stingendum, stingendre, stingeð, stincþ, stingað;* stregdan: *stregdan, strigdeð, strugde, stregdende, strede, stredest, stret, stridit, stræd;* swelgan: *swelgaþ, swelgað, swelgen, swelhð swulgon, swelgend, swelgendes, swelgende, swelgande;* swellan: *swellað, swellendum, swellendu, swellendre;* sweltan: *sweltu, swilteþ, swilteð;* sweorcan: *swearc, sweorcende;* sweorfan: *swyrfþ;* swimman: *swimmende, swamm;* swincan: *swincu, swinceþ, swinceð;* swindan: *swindan, swindeþ, swindeð, swand, swindende;* swingan: *swungene, swungæn, sungenne, swingon;* teldan: *teldan, telde, tildest, teldað, teld;* tolinnan: *tolan;* torinnan: *torinnan;* tospringan: *tospringe;* tostregdan: *tostregdo, tostrigdeð, tostregden, tostregd, tostrugde, tostrogdenra, tostrede, tostredeþ, tostredeð, tostred;* toswellan: *toswelle, toswellaþ;* tosweorcan: *tosworcene;* toþindan: *toþundenne, toþundenes, toþindendum, toþindende, toðinden, toþint, toþand, toþundon, toðinde, toðindst, toðindaþ;* toweorpan: *towerpe, towyrpest, towyrþþ, towerþð, toweorp, towurpen, toworpenum, towerpende;* toweorþan: *towordene;* twislingan: *twislunge;* twispinnan: *twispunnenum;* þerscan: *þerscenne, þrescen, þrexe, þreax, ðurscon, ðerscende;* þindan: *þunden, þundenne, þindan, þindeð, þand;* þingan: *þingen, þingan, þinge, þingæ, þingeð, þunge, þang, þingon, þingun, ðingan, ðinge, ðingæ, ðingo, ðingu, ðunge, ðang, ðingon;* þringan: *þringeð, þrungun, ðringe;* þrintan: *þrinteð;* þurhrinnan: *þurharn;* þurhstingan: *ðurhstingþ;* undersincan: *undernsanc;* undersingan: *undersinge, undersange, undernsange, undernsang;* unlimpan: *unlimpe;* unmurnan: *unmurne;* unsceorpan: *unscearp;* unweorþan: *unweorþe, unweorþest, unweorþ, unweorðe, unweorð, unwurðe;* unwindan: *unwunden;* unþingan: *unþang;* upspringan: *upsprungenne, uppsprungen, uppsprungene, uppsprungenre, uppspringende, upspringe, upsprincð, upspring, upsprang, uppspringan;* utrinnan: *utran, uttran;* utsingan: *utsange, utsang, utsangan;* utweorpan: *utweorpeþ, utweorpeð, utweorpaþ, utwearp;* weorpan: *weorpenne, werpe, wyrpþ, werþð, wyrpeð, weorpað, werp, worpen, worpenne;* weorþan: *werþe, wirst, wert, wirt, wyrþeþ, werþeð, wyrþeð, weorþ, warþ, wurþon, wurþen, wyrðest, werðen, worðenne;* windan: *wundæn, windenne, windende, winde, windæ, windeþ, windað, wind, wunde, wande, wundun, windon;* winnan: *winnend, wanne, wannan;* wiþerweorþan: *wiþerwearð;* wiþerwinnan: *wiþerwinnena,*

wiþerwinnende, wiðerwinnen, wiðerwinnena, wiðerwinnæn, wiþerwinnan, wiðerwinne, wiðerwinnæ; **wiþspurnan***: wiðspurne;* **wiþwindan***: wiþwinde;* **wiþwinnan***: wiðwinnend;* **wringan***: wringene, wrange, wrong, wrangan;* **ymblinnan***: ymlan;* **ymbrinnan***: embrin;* **ymbwindan***: ymbwundon;* **ymbþringan***: ymbþrungæn, ymbþringen, embþrungen, ymbþrungon, ymbðringað, ymbðrungon, ymþrungon, ymbeþrungon.*

b) Attestations in the DOEC and the YCOE

gelimpan: *gelumpen* (VBN), *gelimpan* (VB), *gelimpe* (VBPS), *gelimpeþ* (VBDI), *gelimpþ* (VBPI), *gelimpeð* (VBDI), *gelimpð* (VBPI), *gelimpaþ* (VBPI), *gelimpað* (VBPI), *gelumpe* (VBDS), *gelamp* (VBDI), *gelomp* (VBDI), *gelumpon* (VBDI), *gelumpan* (VBDI), *gelumpun* (VBDI); **gelinnan:** *galan* (VB); **gemelcan:** *gemolcen* (VBN); **gemeltan:** *gemeltan* (VB), *gemelte* (VBD), *gemilt* (VBN), *gemelteð* (VBPI), *gemeltað* (VBPI), *gemelt* (VBI), *gemealt* (VBDI), *gemulton* (VBDI); **gerinnan:** *gerunnen* (VBN), *giernen* (VBPS), *geran* (VBDI), *gearn* (VBDI), *georn* (VBDI), *gyrn* (VBI); **gescrincan:** *gescruncene* (VBN^N), *gescrince* (VBPS), *gescrincað* (VBPI), *gescruncan* (VBDI); **gesincan:** *gesuncen* (VBN), *gesincan* (VB); **gesingan:** *gesungen* (VBN), *gesungene* (VBN^N), *gesingan* (VB), *gesinge* (VBPS), *gesingð* (VBPI), *gesing* (VBI), *gesang* (VBDI), *gesungon* (VBDI); **gespringan:** *gespringeð* (VBPI), *gesprang* (VBDI), *gesprong* (VBDI), *gesprungon* (VBDI); **gespurnan:** *gespurnun* (VBDI); **gestincan:** *gestincen* (VBPS), *gestincan* (VB), *gestincanne* (VB^D), *gestincað* (VBPI), *gestunce* (VBDS), *gestuncon* (VBDI), *gestuncan* (VBDI); **gestingan:** *gestungen* (VBN); **gestregdan:** *gestred* (VBN), *gestrude* (VBD), *gestroden* (VBN); **geswelgan:** *geswealh* (VBDI); **geswellan:** *geswollen* (VBN), *geswollenum* (VBN^D), *geswollena* (VBN^N); **gesweorcan:** *gesweorce* (VBPS), *geswearc* (VBDI), *gesworcene* (VBN^N); **geswimman:** *geswummen* (VBN); **geswincan:** *geswince* (VBPS), *geswincað* (VBPI), *geswunce* (VBD); **geswingan:** *geswungen* (VBN), *geswungene* (VBN^N), *gesungen* (VBN), *gesungene* (VBN^N), *geswinge* (VBPS), *geswungon* (VBDI); **geteldan:** *geteald* (VBN), *getald* (VBN); **geþingan:** *geþungen* (VBN), *geþungenan* (VBN^A), *geðungen* (VBN^N), *geðungenum* (VBN^D), *geðungene* (VBN^N), *geðungenra* (VBN^G), *geþingeð* (VBPI), *geþingað* (VBPI), *geþungon* (VBDI); **geþringan:** *geþrungen* (VBN), *geþringan* (VB), *geþrang* (VBDI), *geðringan* (VB); **geweorpan:** *geweorpan* (VB), *gewyrpð* (VBPI), *gewearp* (VBDI); **geweorþan:** *geweorþan* (BE), *geweorþe* (BEPS), *geweorþeþ* (BEPI), *geweorþeð* (BEPI), *geweorþaþ* (BEPI), *geweorþað* (BEPI), *gewurþe* (VBPS), *gewearþ* (BEDI), *gewurþon* (BEPS), *gewurdon* (BEDI), *gewurdun* (BEDI), *gewurden* (BEN), *geworden* (BEI), *gewordenum* (BEN), *gewordene* (BEN), *gewordenu* (BEN^N), *gewordenre* (BEN^D), *gewordenan* (BEN^D), *geweorðan* (BE), *geweorðe* (BEPS), *geweorðest* (BEPI), *geweorðeþ* (BEPI), *geweorðeð* (BEP), *gewyrðeð* (BEPI), *gewiorðeð* (BEPI), *geweorðað* (BEPI), *gewurðe* (BEPS), *gewearð* (BE), *gewarð* (BEPI), *gewurðon* (BEDI), *gewurðun* (BEPS); **gewindan:** *gewindan* (VB), *gewint* (VBPI), *gewand* (VBDI), *gewundan* (VBDI); **gewinnan:** *gewunnen* (VB), *gewunnenne* (VBN^A), *gewunnenum* (VBN^D), *gewunnene* (VBN^A), *gewinnen* (VB), *gewinnenne* (VB^D), *gewinnan* (VB), *gewinnanne* (VB^D), *gewinnað* (VBPI), *gewunne* (VBD), *gewann* (VBDI), *gewonn* (VBDI), *gewan* (VBDI), *gewon* (VBDI), *gewunnon* (VBDI), *gewunnan* (VBDI), *gewinnon* (VBPS), *gedwan* (VBDI), *gerwan* (VB); **gewringan:** *gewrungen* (VBN^N), *gewrungene* (VBN), *gewringan* (VB), *gewring* (VBI), *gewrungan* (VBN); **limpan:** *limpende* (VAG), *limpe* (VBPS), *limpeþ* (VBPI), *limpeð* (VBPI), *limpð* (VBPI), *lumpe* (VBDS), *lamp* (VBDI), *lomp* (VBDI); **linnan:** *linnan* (VB), *linneð* (VBPI), *linnið* (VBPI), *lin* (VBPS), *lunnon* (VBDI); **medrinnan:** *mearn* (VBDI); **medsinnan:** *mesan* (VB); **melcan:** *melcan* (VB), *melce* (VBN^N), *mealc* (VBDI), *molcen* (VBN^N); **meltan:** *meltan* (VB), *melte* (VBPS), *meltað* (VBPI), *melten* (VBPS), *melt*

(VBPI), *mealt* (VBDI), *multon* (VBDI), *meltende* (VAG^A); **mislimpan:** *mislimpe* (VBPS), *mislamp* (VBDI); **misþingan:** *misþingð* (VBPI); **murnan:** *murnan* (VB), *murne* (VBPS), *myrnst* (VBPI), *myrneð* (VBPI), *myrnð* (VBPI), *murnað* (VBPI), *murn* (VBI), *mearn* (VBDI), *murnende* (VAG); **oferrinnan:** *oferarn* (VBDI), *oferorn* (VBDI); **ofersingan:** *ofersingon* (VBPS); **oferswingan:** *oferswunge* (VBDS); **oferteldan:** *ofertolden* (VBN); **oferþingan:** *oferþungen* (VBN), *oferðungen* (VBN); **oflinnan:** *oflinnan* (VB); **ofrinnan:** *ofarn* (VBDI); **oferweorpan:** *oferweorpe* (VBPS), *oferwearp* (VBDI), *oferworpen* (VBN); **oferwinnan:** *oferwunnen* (VBDS), *oferwunnene* (VBN^N), *oferwinnen* (VBN), *oferwinnenne* (VB^D), *oferwinnan* (VB), *oferwinnanne* (VB^D), *oferwinne* (VBPS), *oferwinnað* (VBPI), *oferwunne* (VBDS), *oferwann* (VBDI), *oferwonn* (VBDI), *oferwan* (VBDI), *oferwon* (VBDI), *oferwunnon* (VBDI), *oferwunnan* (VBDI); **ofstingan:** *ofstungen* (VBN), *ofstingan* (VB), *ofstinge* (VBPS), *ofstunge* (VBDS), *ofstang* (VBDI), *ofstong* (VBDI); **ofswelgan:** *ofswelgð* (VBPI); **ofswingan:** *ofswungen* (VBN), *ofswungon* (VBDI); **ofþingan:** *ofþingð* (VBPI), *ofþincð* (VBPI), *ofðingþ* (VBPI), *ofðingð* (VBPI), *ofðincð* (VBPI); **ofþringan:** *ofþrungen* (VBN), *ofþrungon* (VBDI), *ofðrincð* (VBPI); **ofweorpan:** *ofwearp* (VBDI), *ofwurpon* (VBDI), *ofworpen* (VBN); **onrinnan:** *onarn* (VBDI); **onspringan:** *onspringað* (VBPI), *onsprang* (VBDI), *onsprungon* (VBDI); **onþringan:** *onþrang* (VBDI); **onweorpan:** *onwurpe* (VBDS), *onweorpendum* (VAG^D); **onwindan:** *onwand* (VBDI); **onwinnan:** *onwinnendne* (VAG^A), *onwinnendum* (VBN^D), *onwinnende* (VAG), onwinnenda (VBN^N), onwinnendan (VAG), onwinnan (VB), *onwinnað* (VBPI), *onwann* (VBDI); **oþrinnan:** *oþran* (VBDI), *oðran* (VBDI); **oþspurnan:** *oþspurne* (VBDS); **oþþringan:** *oþþringan* (VB), *oþþringeð* (VBPI), *oðþringeð* (VBPI), *oðþrong* (VBDI); **oþwindan:** *oðwand* (VBDI); **rinnan:** *iernen* (VBPS), *yrnnendum* (VAG^D), *arn* (VBDI), *earn* (VBDI), *orn* (VBDI), *iern* (VBI), *eorn* (VBI); **sceorfan:** *sceorfe* (VBPS), *sceorfendum* (VAG^D), *sceorfende* (VAG), *sceorfendan* (VAG^A); **sceorpan:** *sceorpe* (VBPS); **scielfan:** *scilfð* (VBPI), *scylfð* (VBPI); **sciellan:** *scell* (MDPI), *sceall* (MDP), *scall* (MDPI); **scrincan:** *scrincaþ* (VBPI), *scrincað* (VBPI), *scranc* (VBDI); **sincan:** *sincendum* (VAG^D), *sincan* (VB), *sincað* (VBPI), *sinc* (VBI), *suncon* (VBDI); **singan:** *sungen* (VBDS), *sungene* (VBN^N), *singen* (VBPS), *singenne* (VB^D), *singend* (VAG^N), *singendum* (VAG^D), *singende* (VAG), *singendra* (VAG^G), *singinde* (VAG^N), *singan* (VB), *singanne* (VB^D), *singe* (VBP), *singeþ* (VBPI), *singeð* (VBPI), *singð* (VBPI), *singaþ* (VBPI), *singað* (VBPI), *sing* (VBI), *sunge* (VBD), *sang* (VBDI), *sungon* (VBDI), *sungan* (VBDI), *sungun* (VBDI), *singon* (VBPS); **sinnan:** *sinniþ* (VBPI); **slincan:** *slincan* (VB); **spinnan:** *spinnende* (VAG^N), *spinnað* (VBPI), *spunnon* (VBDI), *spannan* (VB); **springan:** *springe* (VBPS), *sprincð* (VBPI), *springaþ* (VBPI), *springað* (VBPI), *sprang* (VBDI), *sprong* (VBDI), *sprungon* (VBDI); **spurnan:** *spurne* (VBPS), *spyrneð* (VBPI), *spyrnð* (VBPI), *spearn* (VBDI), *spurnon* (VBDI); **steorfan:** *steorfan* (VBPS), *steorfende* (VAG^N); **stincan:** *stincen* (VBPS), *stincendne* (VAG^A), *stincendum* (VAG^D), *stincende* (VAG^A), *stincendan* (VAG^A), *stincan* (VB), *stince* (VBPS), *stinceð* (VBPI), *stincð* (VBPI), *stincað* (VBPI), *stanc* (VBDI), stonc (VBDI), *stuncon* (VBDI); **stingan:** *stingende* (VAG^N), *stingan* (VB), *stinge* (VBPS), *stingð* (VBPI), *stincð* (VBPI), *sting* (VBI), *stunge* (VBDS), *stang* (VBDI), *stungon* (VBDI); **stregdan:** *stregde* (VBD), *stregdað* (VBPI), *stregd* (VBI), *strægd* (VBDI),

strogden (VBN), *stredan* (VB), *stredað* (VBPI), *stred* (VBI), *strude* (VBD), *strad* (VBDI), *strudon* (VBDI); **swelgan:** *swelgan* (VB), *swelge* (VBPS), *swelgeþ* (VBPI), *swelgeð* (VBPI), *swilgeð* (VBPI), *swilgð* (VBPI), *swulge* (VBDS), *swealg* (VBDI), *swealh* (VBDI); **swellan:** *swellan* (VB), *swelle* (VBPS), *swellende* (VAG); **sweltan:** *sweltan* (VB), *sweltenne* (VB^D), *sweltanne* (VB^D), *swelte* (VBP), *sweltest* (VBPI), *swelteþ* (VBPI), *swilt* (VBPI), *swelteð* (VBPI), *sweltaþ* (VBPI), *sweltað* (VBPI), *swelten* (VB), *swelt* (VBI), *swulte* (VBDS), *swealt* (VBDI), *swalt* (VBDI), *swulton* (VBDI), *swulten* (VBDS), *sweltendne* (VAG^A), *sweltendum* (VAG^D), *sweltende* (VAG), *sweltendra* (VAG^G), *sweltenda* (VAG^N), *sweltendan* (VAG^A); **sweorcan:** *sweorceð* (VBPI), *swyrceð* (VBPI); **sweorfan:** *sworfen* (VBN); **swimman:** *swimman* (VB), *swimme* (VBPS), *swimmaþ* (VBPI), *swimmað* (VBPI), *swam* (VBDI), *swom* (VBDI), *swumman* (VBDI); **swincan:** *swuncen* (VBDS); : *swincen* (VBPS), *swincenne* (VB^D), *swincene* (VB^D), *swincendum* (VAG^D), *swincende* (VAG), *swincan* (VB), *swince* (VBP), *swincst* (VBPI), *swincð* (VBPI), *swincaþ* (VBPI), *swincað* (VBPI), *swinc* (VBI), *swunce* (VBDS), *swanc* (VBDI), *swuncon* (VBDI), *swincon* (VBPS); **swingan:** *swungen* (VBN), *swingen* (VBPS), *swingenne* (VB^D), *swingæn* (VB), *sungen* (VBDS), *sungene* (VBN^N), *swingende* (VAG^N), *swingan* (VB), *swinganne* (VB^D), *swinge* (VBP), *swingest* (VBPI), *swincst* (VBPI), swingeð (VBPI), swingð (VBPI), *swincð* (VBPI), *swingaþ* (VBPI), *swingað* (VBI), *swing* (VBI), *swunge* (VBDS), *swang* (VBDI), *swong* (VBDI), *swungon* (VBDI), *swungan* (VBDI), *sunge* (VBD); **teldan:** *telst* (VBPI), *tald* (VBN); **torinnan:** *torinne* (VBPS), *toarn* (VBDI); **tospringan:** *tosprang* (VBDI); **tostincan:** *tostincað* (VBPI); **tostingan:** *tostinge* (VBPS), *tosting* (VBI); **tostregdan:** *tostregde* (VBN^N), *tostrægd* (VBN), *tostrogden* (VBN), *tostredan* (VBPS), *tostret* (VBPI), *tostredað* (VBPI); **toswellan:** *toswollen* (VBN), *toswollenne* (VBN^A), *toswollenum* (VBN^D), *toswollene* (VBN^A); **toþerscan:** *toþærsc* (VBDI), *toþræsc* (VBDI); **toþindan:** *toþunden* (VBN), *toþundenum* (VBN^D), *toþundene* (VBN^N), *toþundenre* (VBN^D), *toðunden* (VBN), *toðundene* (VBN^A), *toþindað* (VBPI), *toðint* (VBPI); **toþringan:** *toþringe* (VBP); **toweorpan:** *toweorpan* (VB), *towerpan* (VB), *toweorpe* (VBP), *toweorpest* (VBPI), *towyrpst* (VBPI), *toweorpeð* (VBPI), *toweorpð* (VBPI), *towirpð* (VBPI), *towyrpð* (VBPI), *toweorpað* (VBI), *toweorpen* (VBPS), *towerp* (VBDI), *towurpe* (VBD), *towearp* (VBDI), *towurpon* (VB), *towurpun* (VBDI), *toworpen* (VBN), *toworpene* (VBN^A), *toworpenu* (VBN^N), toworpena (VBN^N); **þerscan:** *þerscan* (VB), *þirsceð* (VBPI), *þærsc* (VBDI), *ðerscað* (VBI), *ðærsc* (VBDI); **þindan:** *þindende* (VAG), *ðindenne* (VB^D), *þinde* (VBPS), þindeþ (VBPI), *þint* (VBPI), *þindaþ* (VBPI); **þingan:** *ðungen* (VBI), *þingþ* (VBPI), *þingð* (VBPI), *þincþ* (VBPI), *þincð* (VBPI), *þingað* (VBPI), *þing* (VBPI), *þungon* (VBDI), *ðingþ* (VBPI), *ðingð* (VBPI), *ðincþ* (VBDI), *ðincð* (VBDI), *ðingað* (VBPI), *ðing* (VBPI); **þringan:** *ðringende* (MDPS), *þringan* (VB), *þringað* (VBPI), *þrunge* (VBDS), *þrang* (VBDI), *þrong* (VBDI), *þrungon* (VBD), *ðringð* (VBPI), *ðringað* (VBDI), *ðrang* (VBDI); **þurhrinnan:** *ðurharn* (VBDI); **þurhstingan:** *þurhstungen* (VBN), *þurhstingan* (VB), *þurhsting* (VBI), *þurhstong* (VBDI), *ðurhstinge* (VBPS).**understingan:** *understungen* (VBN); **understregdan:** *understregd* (VBN); **unwindan:** *unwindan* (VB); **upspringan:** *upsprungen* (VBN), *upsprungenre* (VBN^D), *upspringende* (VAG^N), *upsprungon* (VBDI); **upweorpan:** *upwearp* (VBDI); **utweorpan:** *utworpen* (VBN); **weorpan:**

weorpan (VB), *weorpe* (BEPS), *weorþeþ* (VBPI), *weorþeð* (VBPI), *weorþð* (VBPI), *wyrþð* (VBPI), *weorpaþ* (VBPI), *weorpen* (VBPS), *weorp* (VBI), *wurpe* (VBD), *wearp* (VBDI), *wurpon* (VBDI), *wurpun* (VB), *wurpen* (VBPS), *worpene* (VBN^N), *werpende* (VAG); **weorþan:** *weorpan* (BE), *weorþe* (BEP), *weorþeþ* (BEPI), *wyrt* (VBPI), *weorþeð* (BEPI), *weorþaþ* (BEPI), *weorþað* (BEDI), *weorþen* (BEPS), *werþ* (BEDI), *wurþe* (BEDS), *wearþ* (BEDI), *wurdon* (BEDI), *wurdun* (BEDI), *wurden* (BED), *worden* (BEN), *wordene* (BEN^N), *weorðan* (BE), *weorðe* (BEDS), *weorðest* (BEPI), *wiorðest* (BEPI), *weorðeþ* (BEPI), *wyrðeþ* (BEPI), *weorðeð* (BEPI), *wyrðeð* (BEPI), *wiorðeð* (BEPI), *weorðaþ* (BEDS), *weorðað* (BEDI), *weorðen* (BEDS), *weorð* (BEPI), *werð* (BEDI), *wurðe* (BEDS), *wearð* (BEDI), *warð* (BEDI), *wurðon* (BEDI), *wurðen* (BE), *weorðendan* (BAG); **windan:** *wunden* (VBN), *windendum* (VAG^A), *windan* (VB), *winst* (VBPI), *wint* (VBPI), *wand* (VBDI), *wond* (VBDI), *wundon* (VBDI), *wundan* (VBDI); **winnan:** *wunnen* (VBDS), *winnen* (VBPS), *winnenne* (VB^D), *winnendum* (VAG^D), *winnende* (VAG), *winnendan* (VAG^A), *winnan* (VB), *winnanne* (VB^D), *winne* (VB), *winneð* (VBPI), *winnð* (VBPI), *winnaþ* (VBPI), *winnað* (VBPI), *winn* (VBI), *wunne* (VBD), *wann* (VBDI), *wonn* (VBDI), *wan* (VBDI), *won* (VBDI), *wunnon* (VBDI), *wunnan* (VBDI), *wunnun* (VBDI), *winnon* (VB); **wiþerwinnan:** *wiðerwinnan* (VB); **wiþweorpan:** *wiðwurpon* (VBDI); **wiþwinnan:** *wiðwinnende* (VAG), *wiðwinnan* (VB), *wiðwinnað* (VBPI), *wiðwunnun* (VBDI); **wringan:** *wringan* (VB), *wringe* (VBPS), *wringð* (VBPI), *wring* (VBI), *wrang* (VBDI), *wrungon* (VBDI); **ymbspinnan:** *ymbspannan* (VB); **ymbweorpan:** *ymbwurpon* (VBDI), *ymbworpenu* (VBN^N); **ymbwindan:** *ymbwinde* (VBP); **ymbþringan:** *embþrungon* (VBDI).

c) Attestations in the YCOE

geþingan: *geðungenan* (VBN^A), *geðingað* (VBPI), *geðungon* (VBDI); **geþringan:** *geðrang* (VBDI); **misþingan:** *misðingð* (VBPI); **ofþringan:** *ofðrungen* (VBN), *ofðrungon* (VBDI); **oþþringan:** *oð ðringan* (VB); **oþspurnan:** *oðspurne* (VBDS); **toþerscan:** *toðærsc* (VBDI), *toðræsc* (VBDI); **toþindan:** *toðundenum* (VBN^D), *toðundenre* (VBN^D), *toðindað* (VBPI); **ymbþringan:** *embðrungon* (VBDI); **þerscan:** *ðerscan* (VB); **þindan:** *ðindende* (VAG), *ðinde* (VBPS), *ðindeð* (VBPI), *ðint* (VBPI), *ðindað* (VBPI); **þingan:** *ðungon* (VBDI); **þringan:** *ðringan* (VB), *ðrunge* (VBDS), *ðrong* (VBDI), *ðrungon* (VBDI); **þurhstingan:** *ðurhstungen* (VBN), *ðurhstingan* (VB), *ðurhsting* (VBI), *ðurhstong* (VBDI).