RAEL: Revista Electrónica de Lingüística Aplicada
Vol./Núm.: 1/14. Año 2015
Páginas: 1-12
Artículo recibido: 30/09/2015
Artículo aceptado: 01/01/2016

ADVANCED AND ISSUES IN THE LEMMATIZATION OF OLF ENGLISH VERBS ON A LEXICAL DATABASE[1]

Marta Tío

Abstract. The aim of this article is to discuss the advances already made as well as the issues that have arisen in the process of lemmatization of Old English weak verbs on a lexical database. A list of lemmas of the second class weak verbs of Old English is compiled by using the latest version of the lexical database *Nerthus*, which incorporates the texts of the *Dictionary of Old English Corpus.* A number of issues are discussed, mainly related to queries and spelling. The conclusion insists of the ways in which the queries as defined so far should be refined.

Keywords: *Old English, lemmatization, normalization, lexical database Nerthus, weak verbs*

1. Aims and relevance of research

This article deals with the lemmatization of Old English and, more specifically, with the lemmas of verbs of the second weak class. Its aim is to discuss the advances as well as the issues of lemmatization on a lexical database.

Very briefly, the process of lemmatization can be described as follows. The different inflectional forms as they appear in the texts have to be related to an abstract form or lemma inflected for a conventional form: in the case of verbs, the infinitive. For instance, given a textual attestation like *hopiað,* it is associated with the infinitive *hopian* 'to hope' by means of the process of lemmatization. Quite often, it is necessary to regularize the forms by means of a process of normalization. For example, when we come across a form like *healsie* we relate it to an infinitive like *hālsian* 'to heal'.

The data have been retrieved from the lexical database of Old English *Nerthus* (www.nerthusproject.com), which incorporates the texts of the *Dictionary of Old English Corpus* (DOEC), with a total of approximately 3,000 texts and 3 million words.

This article can be seen as a contribution in two directions. On the lexicographical side, gathering a list of verbal lemmata and filing them into a database is relevant because the standard dictionaries of Old English, including *An Anglo-Saxon Dictionary, A Concise Anglo-Saxon Dictionary* and *The student's Dictionary of Anglo-Saxon*, do not offer an exhaustive inventory of the inflective forms of each headword and *The Dictionary of Old English*, which does, is still in progress (the letter G was published in 2008). On the theoretical side, this work can be seen as a contribution to the research programme in the morphology and semantics of Old English represented by Martín Arista (2008, 2010a, 2010b, 2011a, 2011b, 2011c, 2012a, 2012b, 2012c, 2013a, 2013b, 2014), Martín Arista et al. (2011), Martín Arista and Mateo Mendaza (2013), Martín Arista and Cortés Rodríguez (2014) and Martín Arista and Vea Escarza (fc.).

[1] This research has been funded through the project FFI2011-29532.

The remainder of this article is organized as follows. Section 2 describes the foundation and organization of a lexical database of Old English. Section 3 reviews the relevant aspects of the morphology of the Old English weak verb classes. Section 4 focuses on the lemmatization of class 2 weak verbs and discusses the advances and issues of the process. Section 5 draws the main conclusions.

2. *Nerthus*. A lexical database of Old English

In its latest format, the lexical database of Old English *Nerthus,* called *The Grid,* consists of five relational layouts, including the dictionary database, the concordance by word to the DOEC, the concordance by fragment to the DOEC, the index to the DOEC (called *The Crib*) and the reversed index to the DOEC (called *The Mirror*). Due to copyright reasons, *Nerthus* is the only open access resource. It contains 30,000 files of lemmatized forms, based primarily on Clark Hall and secondarily on Bosworth-Toller and Sweet. The initial headword list has been compiled by Martín Arista et al. (2011) and the meaning definitions provided by the dictionaries of Old English mentioned above have been synthesized by Martín Arista and Mateo Mendaza (2013).

Martín Arista (2013) presented this new organization in a lecture delivered at the University of Sheffield. Its most salient feature is that the lexical database is no longer based on dictionary forms but on textual forms. In quantitative terms, this means that the number of files increases from 30,000 to 3,000,000. From the quantitative point of view, the new organization provides all textual occurrences of lemmas together with their context and, therefore, allows the researcher to carry out not only morphological and lexical analysis, as the previous version of the database, but also semantic and syntactic analysis. Moreover, all textual variants, frequencies and syntactic patterns can be linked to the dictionary files of the previous version of *Nerthus*.

To briefly illustrate the functionalities of the version of *Nerthus* reviewed in this section, it may be pointed out, in the first, place, that the database can turn out the number of textual occurrences of a lemma. For example, *lufian* appears 319 times in the texts. In the second place, the database can break down the occurrences by inflectional form. For instance, the verb *wunian* occurs in the inflectional forms presented in Figure 1:

Inflectional form	Occurrences	Weak verb 2
wunode	377	*wunian*
wunian	237	*wunian*
wuniað	202	*wunian*
wuniende	127	*wunian*
wunodon	77	*wunian*
wunast	18	*wunian*
wuniaþ	17	*wunian*
wunodest	6	*wunian*
wunianne	6	*wunian*

wunoden	6	*wunian*

Figure 1: Inflections and frequency of ***wunian****.*

Thirdly, the formalism used for representing the prefix *ge*- guarantees the direct link to the *ge*-prefixed counterparts of a given simplex verb such as *wunian*, in Figure 2.

Inflectional form	Occurrences	Weak verb 2
gewunode	93	*wunian(ge)*
gewunian	77	*wunian(ge)*
gewuniað	32	*wunian(ge)*
gewunige	19	*wunian(ge)*
gewunod	18	*wunian(ge)*
gewunodon	6	*wunian(ge)*
gewunie	3	*wunian(ge)*
gewuniaþ	120	*wunian(ge)*

Figure 2: The prefix ge- in ***wunian (ge)****.*

And, fourthly, a given inflectional form, such as *gewunige* appears in the following fragments in Figure 3, whose short titles are based on Mitchell, Ball and Cameron (1975).

[Abbo 000800 (104.6)]

[Abbo 000900 (104.9)]

[Alc (Warn 35) 014200 (286)]

[BenR 020700 (7.24.21)]

[Beo 000800 (20)]

[CollGl 22 (Liebermann-Ker) 003300 (33)]

[Conf 4 (Fowler) 012700 (33.450)]

[CP 022800 (10.61.20)]

[CP 155700 (43.317.17)]

[HomM 1 (Healey) 004900 (157)]

[JnGl (Li) 065800 (15.4)]

[LawVAtr 003900 (22)]

[LawVAtr 004800 (29)]

[LawVIAtr 004600 (27)]

[Lch II (2) 005100 (12.1.1)]

[Lch II (2) 038100 (46.2.4)]

[Lch II (3 Head) 003000 (30)]

[Lch II (3) 009700 (30.1.7)]

[LS 10 (Guth) 005300 (5.242)]

[Met 002000 (1.35)]

[PPs 098400 (108.7)]

[PsGlI (Lindelöf) 229600 (138.9)]

[ThCap 1 (Sauer) 020700 (39.389.8)]

[WHom 15 000900 (33)]

Figure 3: Textual witnesses to ***gewunige****.*

3. The inflection of the Old English weak verb

According to Pyles and Algeo (1982: 125), weak verbs "formed their preterites and past participles in the characteristically Germanic way, by the addition of a suffix containing *d* or immediately after consonants, *t*". Many weak verbs were originally causative verbs derived from other categories, such as nouns or adjectives, by means of the "addition of a suffix with an *i*-sound that mutated the stem vowel of the word" (Pyles and Algeo 1982: 125). In contrast to strong verbs, weak verbs do not change their stem. Mitchell and Robinson (1993: 46) stress that the stem vowel was normally the same throughout all the verbal forms of the paradigm, which reinforces the idea of regularity and that the inflectional endings of strong and weak verbs showed lots of similarities, although they underwent different evolutions. Hogg and Fulk (2011: 258) further remark that the most accepted theory is that weak verbs developed their preterite forms from a periphrasis.

Weak class 1 is one of the largest groups of verbs of all the verbal classes in Old English, among other reasons as process of causative stem formation above mentioned. Class 1 of weak verbs is subdivided into two classes, illustrated by the verbs verbs *fremman* 'to do' and *hīeran* 'to hear' paradigms of these weak verbs are presented in Figure 4, which is based on Mitchel and Robinson (1993: 46) and Hogg and Fulk (2011: 261):

Infinitive: subclass 1: *nerian* 'to save'; subclass 2: fēran 'to depart'

Inflected Infinitive: subclass 1: *tō nerienne*; subclass 2: *tō* fērenne

Present Participle: subclass 1*: neriende*; subclass 2: fērende

Past Participle: subclass 1: *(ge-)nered*; subclass 2*:* fēred

Present indicative			Present subjunctive		
Subclass 1	Subclass 2		Subclass 1	Subclass 2	
sg.	1. *nerie*	*fēre*	sg.	1. *nerie*	*fēre*
	2. *nerest*	*fērst*		2. *nerie*	*fēre*
	3. *nereþ*	*fērþ*		3. *nerie*	*fēre*
pl.	*neriaþ*	*fēraþ*	pl.	*nerien*	*fēren*
Preterite indicative			Preterite subjunctive		
Subclass 1	Subclass 2		Subclass 1	Subclass 2	
sg.	1. *nerede*	*fērde*	sg.	1. *nerede*	*fērde*

	2. *neredest*	*fērdest*		2. *nerede*	*fērde*
	3. *nerede*	*fērde*		3. *nerede*	*fērde*
pl.	*neredon*	*fērdon*	pl.	*nereden*	*fērden*

<u>Imperative</u>

	Subclass 1	Subclass 2
sg.	*nere*	*fēr*
pl.	*neriaþ*	*fēraþ*

Figure 4: The paradigm of class 1 weak verbs ***nerian*** *‘to save’ and fēran ‘to depart’.*

A number of weak verbs had no vowel *i* before the dental preterite suffix in Proto-Germanic, with the consequence that they lack umlaut in the Old English preterite and past participle. In addition, their stems all ended in *-l*, as presented in Figure 5, or velar consonant with the alternation of tʃ <cc> and x <h>, as shown in Figure 6 (Hogg and Fulk 2011: 274):

cwellan ‘to kill’	*cwealde*	*cweald*
dwellan ‘to mislead’	*dwealde*	*dweald*
stellan ‘to position’	*stealde*	*steald*

Figure 5: Stems in *-l*.

cwecc(e)an ‘to vibrate’	*cweahte*	*cweaht*
drecc(e)an ‘to afflict’	*dreahte*	*dreaht*
recc(e)an ‘to recount’	*reahte, rehte*	*reaht, reht*

Figure 6: Stems in velar consonant.

Campbell (1987: 300) remarks that the 2nd. and 3rd. person of the singular (present indicative) of class 1 weak verbs are subject to assimilation. The assimilations of consonants are presented in Figure 7, with an instance of each pattern.

-d-st > *-tst*	*fētst* (infinitive *fēdan* ‘to feed’) then *-tst* > *-st, fēst*
-þ-st >*tst*	*cȳþst, cȳtst* (infinitive *cȳþan* ‘to proclaim’)
-g-st > *-hst*	*bīhst* (infinitive *bīegan* ‘to bend’)
-ng-st > *-ncst*	*sprenst* (infinitive *sprengan* ‘to scatter’)
-t-þ, -d-þ > *-tt*	*mētt,* (infinitive *mētan* ‘to measure’)
-s-þ > *-st*	*alȳst* (infinitive *alīesan* ‘to free’)
-g-þ > *-hþ*	*bīhþ* (infinitive *bīegan* ‘to bend’)
-ng-þ > *ncþ*	*glencþ* (infinitive *glengan* ‘to decorate’)

Figure 7: Assimilation in the 2nd. and 3rd. person of the singular number.

Moving on to the characteristics of the next class, we find class 2 of weak verbs, the one on which this work focuses. Mitchell and Robinson (1993: 49) remark that this class of verbs "present few problems". As Hogg puts it (2011: 279), this was the only group of verbs which kept adding new verbs during the Old English period. The paradigms of the weak verbs *lufian* 'to love' (Mitchell and Robinson. 1993: 49-50), identified as 'subclass 1', and the verb *lofi(g)an* 'to praise' (Hogg and Fulk 2011: 279-280), identified as 'subclass 2', are presented in Figure 8 in order to compare their forms:

Infinitive: subclass 1: *lufian* 'love'; subclass 2: *lofian* 'praise'

Inflected infinitive: subclass 1: *tō lufienne*; subclass 2: *tō lofianne*

Present Participle: subclass 1: *lufiende*; subclass 2: *lofiende*

Past Participle: subclass 1: *(ge-)lufod*; subclass 2: *lofod*

Present indicative			Present subjunctive		
	Subclass 1	Subclass 2	Subclass 1		Subclass 2
sg.	1. *lufie*	*lofige*	sg.	1. *lufie*	*lofige*
	2. *lufast*	*lofast*		2. *lufie*	*lofige*
	3. *lufaþ*	*lofað*		3. *lufie*	*lofige*
pl.	*lufiaþ*	*lofiað*	pl.	*lufien*	*lofigen*

Preterite indicative			Preterite subjunctive		
	Subclass 1	Subclass 2	Subclass 1		Subclass 2
sg.	1. *lufode*	*lofode*	sg.	1. *lufode*	*lofode*
	2. *lufodest*	*lofodest*		2. *lufode*	*lofode*
	3. *lufode*	*lofode*		3. *lufode*	*lofode*
pl.	*lufodon*	*lofodon*	pl.	*lufoden*	*lofoden*

Imperative		
Subclass 1		Subclass 2
sg.	1. *lufa*	*lofa*
pl.	2. *lufiað*	*lofiað*

Figure 8: The paradigm of class 2 weak verbs ***lufian*** *'to love' and* ***lofian*** *'to praise'*

Although Hogg and Fulk (2011: 280) notice that "the inflexions of weak verbs of class 2 are, with the exceptions discussed below, the same for all stems, regardless of weight", these verbs also present some peculiarities, such as contracted forms. As a result of the loss of intervocalic *h*, there were two stems within paradigms like *smēagan* 'to consider': *smēag-* and *smēa-* (Campbell 1987: 334), illustrated in Figure 9.

Infinitive	*smēagan*
Pres. part.	*smēagende*
Pass part.	*smēad*

Present indicative		Present subjunctive	
sg.	1. *smēage*	sg.	1. *smēage*
	2. *smēast*		2. *smēage*
	3. *smēaþ*		3. *smēage*
pl.	*smēagaþ*	pl.	*smēagen*
Preterite indicative		Preterite subjunctive	
sg.	1. *smēade*	sg.	1. *smēade*
	2. *smēaest*		2. *smēade*
	3. *smēade*		3. *smēade*
pl.	*smēadon*	pl.	*smēaden*
Imperative			
sg.	*smēa*		
pl.	*smēagaþ*		

Figure 9: The contracted class 2 weak verb ***smēagan*** *'to consider'*

The last class of weak verbs is class 3. Hogg and Fulk (2011: 289) explain that "verbs of the third weak class in Germanic are in origin structurally parallel to those of the second weak class" and that the only reason why they became a different class is a vocalic alternation in the formation of the stem. There are just four verbs in class 3, *habban* 'to have', *libban* 'to live', *secg(e)an* 'to say' and *hycg(e)an* 'to think' (Campbell 1987: 337), whose paradigms can be seen in Figure 10.

Infinitive		*habban*	*libban*	*secgan*	*hycgan*
Pres. part.		*hæbbende*	*libbende*	*secgende*	*hycgende*
Past part.		*hæfd*	*lifd*	*sægd*	*hogd*
Present indicative					
	sg.	1. *hæbbe*	*libbe*	*secge*	*hycge*
		2. *hæfst*	*leofast*	*sægst*	*hygst*
		3. *hæfþ*	*leofaþ*	*sægþ*	*hygþ*
	pl.	*habbaþ*	*libbaþ*	*secgaþ*	*hycgaþ*
Present subjunctive					
	sg	*hæbbe*	*libbe*	*secge*	*hycge*
	pl.	*hæbben*	*libben*	*secgen*	*hycgen*
Preterite indicative					
	sg.	1. *hæfde*	*lifde*	*sægde*	*hogde*
		2. *hæfdest*	*lifdest*	*sægdest*	*hogdest*
		3. *hæfde*	*lifde*	*sægde*	*hogde*
	pl.	*hæfdon*	*lifdon*	*sægdon*	*hogdon*

Preterite subjunctive

sg.	*hæfde*	*lifde*	*sægde*	*hogde*
pl.	*hæfden*	*lifden*	*sægden*	*hogden*

Imperative

sg.	*hafa*	*leofa*	*sæge*	*hyge*
pl.	*habbaþ*	*libbaþ*	*secgaþ*	*hycgaþ*

Figure 10: The paradigms of class 3 weak verbs ***habban*** *'to have',* ***libban*** *'to live',* ***secg(e)an*** *'to say' and* ***hycg(e)an*** *'to think'.*

4. Finding and lemmatizing Old English weak verbs of the second class

For Burkhanov (1998: 122), "the term 'lemmatization' is used to refer to the reduction of inflectional word forms to their lemmata, i.e. basic forms, and the elimination of homography" (...) [i]n practice, lemmatization involves the assignment of a uniform heading under which elements of the corpora containing the word forms of same lexeme are represented." According to Burkhanov, in order to organize the corpus of a dictionary it is necessary to lemmatize the attestations or textual (inflected) forms that correspond to each dictionary headword. Thus, as Atkins and Rundell (2008: 325) point out, the headword "links all the information about one word together in one entry.

Lemmatizing requires the previous task of finding the relevant forms. So as to avoid ambiguity and overlapping with other paradigms, a set of formally distinctive forms of verbs of the second weak class have been selected that include: the infinitive (-*ian*), the inflected infinitive (-*ianne*), the present participle (-*iende*), the past participle (*ge*-*od*), the first person singular of the present indicative (-*ie/ge*-*ige*) the second person singular of the present indicative (-*ast*), the present indicative plural (-*iað/*-*iaþ*), the present subjunctive singular (-*ie/ge*-*ige*), the first/third person singular of the preterite indicative (-*ode*), the second person singular of the preterite indicative (-*odest*), the preterite indicative plural (-*odon*) and the preterite subjunctive plural (-*oden*).

The next step of the lemmatization process is to extract the attestations ending with these inflections from the DOEC. This has not been done by means of the search engine provided by the online corpus but on the lexical database of Old English *Nerthus*. The database format has a great advantage over the online corpus: it can search the results of previous searches. Thus, the process of lemma assignment advances on the basis of succesive searches that refine little by little the results. With query strings like ==*iað, ==*ode, ==*ian, ==*iaþ, ==*ast, ==*odon, ==*iende, ==*ianne, ==*odest, ==*od, ==*ige, ==*oden and ==*ie the database turns out verbal forms such as *eardiað, geeardode, eardian, eardiaþ, geeardast, eardodon, eardiende, eardianne, eardodest, geeardod*, *geeardige*, *eardoden* and *eardie* respectively. In the process of lemmatization, these inflectional forms are grouped under the basic form or lemma of *eardian(ge)* (17 occurrences).

A total of 187,000 inflectional forms have been searched for these endings and 1,064 lemmas of weak verbs from the second class have been found, of which 285 were not on the lexical database of Old English *Nerthus* before. Apart from proposing lemmas, this analysis has also helped to improve the information on some lemmas that already appear in dictionaries. This is the case with verbs to which, given the textual evidence, it is necessary to add the prefix *ge*-, as, for instance, *hegian, hȳrian, sīþian*, *sorgian* and *windwian*.

Whereas the lemmatization that has been carried out provides a more accurate knowledge of the relationship as regards the second class of weak verbs between Old English texts on the one hand and dictionaries and databases on the other, a number of issues have arisen that advise to make some changes to future research.

In spite of the advantages of the lexical database, when it comes to searching the corpus the search process is far from automatic. In the first place, many undesired results are turned out if the query segment is very short or unspecific. For instance, searching for the inflectional endings *-od* and *-ige* we not only obtain verbs but also adjectives and nouns, such as forestige 'vestibule' and forebod 'preaching'. The solution that has been adopted in this respect is that the inflectional endings *-ige* and *-od* have been searched only in combination with the prefix *ge-*, thus *ge-ige*, *ge-od*, as in *gehagige* and *gehyrod*. In the second place, the dictionaries have been necessary for assigning vowel length to lemmas because the DOEC does not mark vowel length. This is the case with the infinitive *āclian*, which displays the long vowel *ā.*

A major issue of the process of lemmatization has to do with the manual work needed to find forms that deviate from the paradigms provided by grammars, which tend to represent the West-Saxon dialect. This is to say, some sort of regularization is necessary that accomodates diachronic, dialectal or textual variants to the inflectional paradigms as presented by grammars. Normalization is, in fact, a part of the process of lemmatization and consists of the regularization of non-standard spellings. As Sweet (1976: xi) explains it, "it is often necessary to put the word where the user of the dictionary expects to find it. Therefore, when several spellings of a word appear in the texts, it is necessary to opt for one of them in a consistent way". For instance, inflected forms such as *hersumie* or *gehersumiað* are found under the lemma *hīersumian(ge*) (2 occurrences). *A Concise Anglo-Saxon Dictionary* provides an extensive list of the correspondences it uses for the normalization of Old English texts, but this list has not been used as such because it overnormalizes has many circularities. Instead, the only correspondences that have been selected are those idenfied by Stark (1982) and de la Cruz (1986) as constituting instances of dialectal or diachronic variation. Such instances of dialectal or diachronic variation include the simplification and gemination of consonants as well as the vocalic correspondences that can be seen in the following figure (≈ means 'is normalized as'; notice that normalization selects graphemes, indicated by < >).

Normalization based on intradialectal variation

< y > ≈ < ie >
< i > ≈ < ie >
< i > ≈ < y >
< e > ≈ < ea >
Geminación: VCC ≈ VC

Normalization based on interdialectal variation

< e > ≈ < æ >
< e > ≈ < ie >
< e > ≈ < ēa >
< e > ≈ < ea >
< e > ≈ < eo>
< æ > ≈ < ēa >

< æ > ≈ < ea >
< a > ≈ < ea >
< eo > ≈ < e >
< eo > ≈ < ie >
< io > ≈ < i >

Figure 11: Vocalic and consonantal correspondences in normalization.

To solve the problem of circularity, the process of normalization is unidirectional, so that it takes place from left to right, but not viceversa. Some instances of normalization based on these correspondences are the following:

<u>Intradialectal</u>

< y > ≈ < ie >: i.e. *gehyrsumast* ≈ *hiersumian(ge)*; *gyrwast* ≈ *gierwan(ge).*

< i > ≈ < ie >: i.e. *giddodest* ≈ *gieddian; gediglodon* ≈ *dīeglan(ge).*

< i > ≈ < y >: *drigast* ≈ *drȳgan; asindrodest* ≈ *āsyndran.*

< e > ≈ < ea >: *yrfewerdast* ≈ *yrfeweardian; berefodon* ≈ *berēafian.*

VCC ≈ VC: *gemicclodest* ≈ *miclian; geættrodon* ≈ *ættrian(ge).*

<u>Interdialectal</u>

< e > ≈ < æ >: *arefnodon >>> ārǣfnan.*

< e > ≈ < ie >: *gedeglodon* ≈ *dīeglan(ge); gehersumige* ≈ *hiersumian(ge).*

< e > ≈ < ēa >: *berefodon* ≈ *berēafian.*

< e > ≈ < ea >: *yrfewerdast* ≈ *yrfeweardian.*

< e > ≈ < eo>: *sweðerodon* ≈ *sweoðerian.*

< æ > ≈ < ēa >: *bescæwast* ≈ *bescēawian*; *forescæwodest* ≈ *forescēawian(ge).*

< æ > ≈ < ea >: *gegærwige* ≈ *gearwian(ge)*; *yrfwærdast* ≈ *yrfeweardian.*

< a > ≈ < ea >: *oferscadodest* ≈ *ofersceadian*; *gemonifaldod* ≈ *manigfealdian(ge).*

< eo > ≈ < e >: *streowodon* ≈ *strēwian(ge.)*

< eo > ≈ < ie >: *cleopodest* ≈ *cliepian.*

< io > ≈ < i >: *cliopodon* ≈ *clipian(ge).*

Figure 12: Illustration of normalization.

5. Conclusion

The first conclusion of this research is that, when it comes to lemmatizing Old Enligh verbs, the database format has clear advantages over online corpora. A database can be adapted to the specific needs of a particular research. It can be sorted and searched in ways that online corpora cannot. It facilitates the definition of relations between data that cannot be captured by online corpora. And, moreover, the database format allows us to use simultaneously the corpus, the concordance and the index of the language of analysis.

The second conclusion is in fact an outlook of future research. In spite the advances already made, this work leaves some aspects pending. The first is exhaustivity. Once a significant number of lemmas have been related to their corresponding textual forms, it is necessary to search the corpus form more lemmas and, above all, for more inflectional forms. To do this, it will be necessary to refine the searches in at least three ways. Firstly, more inflectional endings and more variants of such endings should be considered. Secondly, the variants of the verbal prefixes should be taken into account. And, thirdly, the prefix *ge-* in combination with all the endings should be searched for. It will also be necessary to widen the scope of the analysis. The other two classes of weak verbs may be included, especially the first class. It not only represents the most numerous class but also has points of contact with the second class.

References

Atkins, B. T. Sue and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

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

Burkhanov, I. 1998. *Lexicography. A Dictionary of basic terminology*. Wydawn: Wyższej Szkoły Pedagogicznej w Rzeszowie.

Campbell, A. 1987. *Old English Grammar*. Oxford: Oxford University Press.

Clark Hall, J. R. 1996 (1896). *A Concise Anglo-Saxon Dictionary*. Toronto: University of Toronto Press.

de la Cruz, J. 1982. *Iniciación Práctica al Inglés Antiguo*. Madrid: Editorial Alhambra.

Healey, A.d.P., ed. 2008: *The Dictionary of Old English in Electronic Form A-G*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Healey, A.d.P., J. Price-Wilkin and X.Xiang, eds. 2004: *The Dictionary of Old English Web Corpus*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto. [Available at http://www.doe.utoronto.ca/pages/pub/web-corpus.html]

Hogg, R. M. and R. Fulk. 2011. *A grammar of Old English. Volume 2. Morphology*. Oxford: Blackwell.

Jackson, H. 2002. *Lexicography. An Introduction*. London: Routledge.

Lass, R. and J. Anderson. 1975. *Old English Phonology*. Cambridge: Cambridge University Press.

Martín Arista, J. 2008. Unification and separation in a functional theory of morphology. In R. Van Valin (ed.) *Investigations of the Syntax-Semantics-Pragmatics Interface*. Amsterdam: John Benjamins, 119-45.

Martín Arista, J. 2010a. OE strong verbs derived from strong verbs. *SKASE Journal of Theoretical Linguistics* 7: 36-56.

Martín Arista, J. 2010b. Lexical negation in Old English. *NOWELE (North-Western European Language Evolution)* 60/61: 89-108.

Martín Arista, J. 2011a. Projections and Constructions in Functional Morphology. The Case of Old English *HRĒOW*. *Language and Linguistics* 12/2: 393-425.

Martín Arista, J. 2011b. Adjective formation and lexical layers in Old English. *English Studies* 92/3: 323-334.

Martín Arista, J. 2011c. Morphological relatedness and zero alternation in Old English. In P. Guerrero Medina (ed.), *Morphosyntactic Alternations in English*. Sheffield/Oakville: Equinox. 339-362.

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.

Martín Arista, J. 2012c. Nerthus. The reference list of Old English strong verbs. *Working Papers in Early English Lexicology and Lexicography* 2.

Martín Arista, J. 2013a. Recursivity, derivational depth and the search for Old English lexical primes. *Studia Neophilologica* 85/1: 1-21.

Martín Arista, J. 2013b. Nerthus. Lexical Database of Old English: From Word-Formation to Meaning Construction. Lecture delivered at the English Linguistics Research Seminar (Centre for Research in Humanities), University of Sheffield.

Martín Arista, J. 2014. *Noun layers in Old English. Mismatches and asymmetry in lexical derivation*. *Nordic Journal of English Studies* 13(3): 160-187.

Martín Arista, J., E. González Torres, G. Maíz Villalta, R. Mateo Mendaza, C. Novo Urraca, R. Vea Escarza and R. Torre Alonso. 2011. Nerthus. A lexical database of Old English. The initial headword list 2007-2009. *Working Papers in Early English Lexicology and Lexicography* 1.

Martín Arista, J. and R. Mateo Mendaza. 2013. Nerthus. Outline of a lexicon of Old English. *Working Papers in Early English Lexicology and Lexicography* 3.

Martín Arista, J. and F. Cortés Rodríguez. 2014. From directionals to telics: meaning construction, word-formation and grammaticalization in Role and Reference Grammar. In Gómez González, María Ángeles, Francisco Ruiz de Mendoza Ibáñez & Francisco Gonzálvez García (eds.), *Theory and Practice in Functional-Cognitive Space*. Amsterdam: John Benjamins. 229-250.

Martín Arista, J. and R. Vea Escarza. Assessing the semantic transparency of Old English affixation: adjective and noun formation. *English Studies*. Forthcoming.

Mitchell, B., C. Ball and A. Cameron. 1975. Short titles of Old English texts. *Anglo-Saxon England* 4: 207-221.

Mitchell, B. and F. Robinson. 1993. *A guide to Old English.* Cambridge: Blackwell.

Nerthus. A lexical database of Old English [Available at www.nerthusproject.com]

Prokosch, E. 1939. *A Comparative Germanic Grammar*. Philadelphia: University of Pennsylvania.

Pyles, T. and J. Algeo. 1982. *The origins and development of the English language.* New York: Harcourt Brace Jovanovich.

Ringe, D. 2006. *From Proto-Indo-European to Proto-Germanic. A Linguistic History of English. Volume I.* Oxford: Oxford University Press.

Robinson, O. W. 1993. *Old English and its closest relatives: a survey of the earliest Germanic languages.* London: Routledge.

Smith, J. J. 2009. *Old English: a linguistic introduction.* Cambridge: Cambridge University Press.

Stark, D. 1982. *The Old English weak verbs. A diachronic and synchronic analysis*. Tübingen: Niemeyer.

Sweet, M. 1893. The Third Class of Weak Verbs in Primitive Teutonic, with Special Reference to Its Development in Anglo-Saxon. *The American Journal of Philology* 14 (4): 409-455.

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