

## THE REGULARIZATION OF OLD ENGLISH WEAK VERBS

Marta Tío Sáenz  
 University of La Rioja

**Abstract:** *This article deals with the regularization of non-standard spellings of the verbal forms extracted from a corpus. It addresses the question of what the limits of regularization are when lemmatizing Old English weak verbs. The purpose of such regularization, also known as normalization, is to carry out lexicological analysis or lexicographical work. The analysis concentrates on weak verbs from the second class and draws on the lexical database of Old English Nerthus, which has incorporated the texts of the Dictionary of Old English Corpus. As regards the question of the limits of normalization, the solutions adopted are, in the first place, that when it is necessary to regularize, normalization is restricted to correspondences based on dialectal and diachronic variation and, secondly, that normalization has to be unidirectional.*

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

### 1. AIMS OF RESEARCH

The aim of this research is to propose criteria that limit the process of normalization necessary to regularize the lemmata of Old English weak verbs from the second class. In general, lemmatization based on the textual forms provided by a corpus is a necessary step in lexicological analysis or lexicographical work. In the specific area of Old English studies, there are several reasons why it is important to compile a list of verbal lemmata. To begin with, the standard dictionaries of Old English, including *An Anglo-Saxon Dictionary*, *A Concise Anglo-Saxon Dictionary* and *The student's Dictionary of Anglo-Saxon* are complete although they are not based on an extensive corpus of the language but on the partial list of sources given in the prefaces or introductions to these dictionaries. Secondly, *The Dictionary of Old English* is based on the *Dictionary of Old English Corpus*, which contains all the surviving texts with a total of six million words, but is still in progress (the letter G was published in 2008). And, thirdly, this work can be seen as a contribution to the research programme in the morphology and semantics of Old English as presented in 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 Veá Escarza (fc.).

The outline of this article is as follows. Section 2 focuses on the relevant aspects of the morphology of the weak verbs of Old English. Section 3 discusses the diatopic and diachronic features of Old English that can be applied to the normalization of weak verbs. Section 4 presents the results of the analysis by inflectional form and lemma, and puts the focus on the criteria that both motivate and constrain the process of normalization. To round off, Section 5 draws the main conclusions of this research.

### 2. THE INFLECTIONAL MORPHOLOGY OF THE OLD ENGLISH VERB

This section deals with the characteristics of the three subclasses of weak verbs and their specific features. The first part offers some diachronic perspectives on this verbal class, while the second part of the section provides a purely synchronic description of the morphology of the weak verbal class.

The Old English verbal endings derive from a number of Proto-Germanic endings, as Smith (2009: 113) remarks. The present indicative plural ending comes from the 3<sup>rd</sup> person plural of the present indicative in Proto-Germanic (Gothic *-and*) whereas the preterite indicative plural ending can be traced back to the 3<sup>rd</sup> person plural of the

preterite indicative (Gothic *-un*). Other correspondences can be established by comparing the Proto-Germanic and the Old English paradigms of *fremman* 'to do'. The Proto-Germanic forms are based on Prokosch (1939) as well as Hogg and Fulck (2011). The contrast between personal endings in Proto-Germanic and Old English personal endings can be seen in Figure 1.

	Proto-Germanic	Old English
<u>Infinitive:</u>	* <i>framjanan</i>	<i>fremman</i> 'to do'
<u>Inflected Infinitive:</u>	* <i>tō framjanjai</i>	<i>tō fremmenne</i>
<u>Present Participle:</u>	* <i>framjanðijaz</i>	<i>fremmende</i>
<u>Past Participle:</u>	* <i>framidaz</i>	( <i>ge-</i> ) <i>fremed</i>
<u>Imperative:</u>		
sg.	* <i>frami</i>	<i>freme</i>
pl.	* <i>framjanþ</i>	<i>fremmap</i>
<u>Present indicative:</u>		
sg. 1.	* <i>framjō</i>	<i>fremme</i>
2.	* <i>framjis</i>	<i>fremest</i>
3.	* <i>framjip</i>	<i>fremeþ</i>
pl.	* <i>framjanþ</i>	<i>fremmap</i>
<u>Present subjunctive:</u>		
sg.	* <i>framjai</i>	<i>fremme</i>
pl.	* <i>framjain</i>	<i>fremmen</i>
<u>Preterite indicative:</u>		
sg. 1.	* <i>framidōn</i>	<i>fremede</i>
2.	* <i>framidæs</i>	<i>fremedest</i>
3.	* <i>framidæ</i>	<i>fremede</i>
pl.	* <i>framidæðun</i>	<i>fremedon</i>
<u>Preterite subjunctive:</u>		
sg.	* <i>framidæðī</i>	<i>fremede</i>
pl.	* <i>framidæðīn</i>	<i>fremeden</i>

**Figure 1.** The weak verb *fremman* 'to do' in Proto-Germanic and Old English.

As presented in Figure 1, most consonants in the personal endings of the Old English verb can be directly related to the Proto-Germanic reconstruction. Apart from the endings mentioned above, the *-n* of the subjunctive plural, both in the present and preterite, is also noteworthy.

On the synchronic side, Pyles and Algeo (1982: 125) remark that 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*". In contrast to strong verbs, these forms do not modify the stem of the verb. Hogg and Fulck (2011: 258) also point out that those suffixes were dental consonants with the function of marking the preterite or past tense. Thus, weak verbs added dental consonants rather than using ablaut or reduplication. In this respect, the most accepted theory is that weak verbs developed their preterite forms from a periphrasis. Pyles and Algeo (1982: 125) hold that 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". Mitchell and Robinson (1993: 46) add 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.

Weak class 1 is one of the largest groups of verbs of all the verbal classes in Old English, among other reasons as a result of the above mentioned process of causative stem formation. Class 1 of weak verbs is subdivided into two classes, illustrated by the verbs *fremman* 'to do' and *hieran* 'to hear'. The paradigms of these weak verbs are presented in Figure 2, which is based on Mitchell and Robinson (1993: 46):

<u>Infinitive:</u>	subclass 1: <i>fremman</i> ‘to do’;	subclass 2: <i>hieran</i> ‘to hear’	
<u>Inflected Infinitive:</u>	subclass 1: <i>tō fremmenne</i> ;	subclass 2: <i>tō hierenne</i>	
<u>Present Participle:</u>	subclass 1: <i>fremmende</i> ;	subclass 2: <i>hierenne</i>	
<u>Past Participle:</u>	subclass 1: <i>(ge-)fremed</i> ;	subclass 2: <i>(ge-)nered</i>	
<u>Present indicative:</u>		<u>Present subjunctive:</u>	
Subclass 1	Subclass 2	Subclass 1	Subclass 2
sg. 1. <i>fremme</i>	<i>hīere</i>	sg. 1. <i>fremme</i>	<i>hīere</i>
2. <i>fremest</i>	<i>hīerst</i>	2. <i>fremme</i>	<i>hīere</i>
3. <i>fremeþ</i>	<i>hīerþ</i>	3. <i>fremme</i>	<i>hīere</i>
pl. <i>fremmaþ</i>	<i>hīeraþ</i>	pl. <i>fremmen</i>	<i>hīeren</i>
<u>Preterite indicative:</u>		<u>Preterite subjunctive:</u>	
Subclass 1	Subclass 2	Subclass 1	Subclass 2
sg. 1. <i>fremede</i>	<i>hīerde</i>	sg. 1. <i>fremede</i>	<i>hīerde</i>
2. <i>fremedest</i>	<i>hīerdest</i>	2. <i>fremede</i>	<i>hīerde</i>
3. <i>fremede</i>	<i>hīerde</i>	3. <i>fremede</i>	<i>hīerde</i>
pl. <i>fremedon</i>	<i>hīerdon</i>	pl. <i>fremeden</i>	<i>hīerden</i>
<u>Imperative:</u>			
Subclass 1	Subclass 2		
sg. <i>freme</i>	<i>hīer</i>		
pl. <i>fremmaþ</i>	<i>hīeraþ</i>		

Figure 2. The paradigm of class 1 weak verbs *fremman* ‘to do’ and *hieran* ‘to hear’.

A number of weak verbs had no vowel *i* before the dental preterite suffix in Proto-Germanic, with the consequence that they lacked umlaut in the Old English preterite and past participle. In addition, their stems all ended in *-l*, as presented in Figure 3, or velar consonant with the alternation of *tʃ* <cc> and *x* <h>, as shown in Figure 4 (Hogg and Fulk 2011: 274):

<i>cwellan</i> ‘to kill’	<i>cwealde</i>	<i>cweald</i>
<i>dwellan</i> ‘to mislead’	<i>dwealde</i>	<i>dweald</i>
<i>stellan</i> ‘to position’	<i>stealde</i>	<i>steald</i>

Figure 3. Stems in *-l*.

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

Figure 4. Stems in velar consonant.

Campbell (1987: 300) remarks that the 2<sup>nd</sup>. and 3<sup>rd</sup>. person of the singular (present indicative) of class 1 weak verbs are subject to assimilation. The assimilations of consonants are presented in Figure 5, with an instance of each pattern.

<i>-d-st</i> > <i>-tst</i>	<i>fētst</i> (infinitive <i>fēdan</i> ‘to feed’) then <i>-tst</i> > <i>-st</i> , <i>fēst</i>
<i>-p-st</i> > <i>tst</i>	<i>cȳpst, cȳtst</i> (infinitive <i>cȳpan</i> ‘to proclaim’)
<i>-g-st</i> > <i>-hst</i>	<i>bīhst</i> (infinitive <i>bīegan</i> ‘to bend’)
<i>-ng-st</i> > <i>-ncst</i>	<i>sprenst</i> (infinitive <i>sprengan</i> ‘to scatter’)
<i>-t-p, -d-p</i> > <i>-tt</i>	<i>mētt</i> , (infinitive <i>mētan</i> ‘to measure’)
<i>-s-p</i> > <i>-st</i>	<i>alȳst</i> (infinitive <i>alīesan</i> ‘to free’)
<i>-g-p</i> > <i>-hp</i>	<i>bīhp</i> (infinitive <i>bīegan</i> ‘to bend’)
<i>-ng-p</i> > <i>ncp</i>	<i>glencp</i> (infinitive <i>glengan</i> ‘to decorate’)

Figure 5. Assimilation in the 2<sup>nd</sup>. and 3<sup>rd</sup>. 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), the peculiarity of this class of verbs relies on the fact that 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 6 in order to compare their forms.

<u>Infinitive:</u>	subclass 1: <i>lufian</i> ‘love’;	subclass 2: <i>lofian</i> ‘praise’	
<u>Inflected infinitive:</u>	subclass 1: <i>tō lufienne</i> ;	subclass 2: <i>tō lofianne</i>	
<u>Present Participle:</u>	subclass 1: <i>lufiende</i> ;	subclass 2: <i>lofiende</i>	
<u>Past Participle:</u>	subclass 1: <i>(ge-)lufod</i> ;	subclass 2: <i>lofod</i>	
<u>Present indicative:</u>		<u>Present subjunctive:</u>	
<i>Subclass 1</i>	<i>Subclass 2</i>	<i>Subclass 1</i>	<i>Subclass 2</i>
sg. 1. <i>lufie</i>	<i>lofige</i>	sg. 1. <i>lufie</i>	<i>lofige</i>
2. <i>lufast</i>	<i>lofast</i>	2. <i>lufie</i>	<i>lofige</i>
3. <i>lufaþ</i>	<i>lofað</i>	3. <i>lufie</i>	<i>lofige</i>
pl. <i>lufiaþ</i>	<i>lofiað</i>	pl. <i>lufien</i>	<i>lofigen</i>
<u>Preterite indicative:</u>		<u>Preterite subjunctive:</u>	
<i>Subclass 1</i>	<i>Subclass 2</i>	<i>Subclass 1</i>	<i>Subclass 2</i>
sg. 1. <i>lufode</i>	<i>lofode</i>	sg. 1. <i>lufode</i>	<i>lofode</i>
2. <i>lufodest</i>	<i>lofodest</i>	2. <i>lufode</i>	<i>lofode</i>
3. <i>lufode</i>	<i>lofode</i>	3. <i>lufode</i>	<i>lofode</i>
pl. <i>lufodon</i>	<i>lofodon</i>	pl. <i>lufoden</i>	<i>lofoden</i>
<u>Imperative:</u>			
<i>Subclass 1</i>	<i>Subclass 2</i>		
sg. 1. <i>lufa</i>	<i>lofa</i>		
pl. 2. <i>lufiað</i>	<i>lofiað</i>		

Figure 6. The paradigm of class 2 weak verbs *lufian* ‘to love’ and *lofi(g)an* ‘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 7.

<u>Infinitive:</u>	<i>smēagan</i>
<u>Present participle:</u>	<i>smēagende</i>
<u>Past participle:</u>	<i>smēad</i>
<u>Present indicative:</u>	<u>Present subjunctive:</u>
sg. 1. <i>smēage</i>	sg. 1. <i>smēage</i>
2. <i>smēast</i>	2. <i>smēage</i>
3. <i>smēaþ</i>	3. <i>smēage</i>
pl. <i>smēagaþ</i>	pl. <i>smēagen</i>
<u>Preterite indicative:</u>	<u>Preterite subjunctive:</u>
sg. 1. <i>smēade</i>	sg. 1. <i>smēade</i>
2. <i>smēaest</i>	2. <i>smēade</i>
3. <i>smēade</i>	3. <i>smēade</i>
pl. <i>smēadon</i>	pl. <i>smēaden</i>
<u>Imperative:</u>	
sg. <i>smēa</i>	
pl. <i>smēagaþ</i>	

Figure 7. 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 8.

<u>Infinitive:</u>	<i>habban</i>	<i>libban</i>	<i>secgan</i>	<i>hycgan</i>
<u>Present participle:</u>	<i>hæbbende</i>	<i>libbende</i>	<i>secgende</i>	<i>hycgende</i>
<u>Past participle:</u>	<i>hæfd</i>	<i>lifd</i>	<i>sægd</i>	<i>hogd</i>
<u>Present indicative:</u>				
sg. 1.	<i>hæbbe</i>	<i>libbe</i>	<i>secge</i>	<i>hycge</i>
2.	<i>hæfst</i>	<i>leofast</i>	<i>sægst</i>	<i>hygst</i>
3.	<i>hæfþ</i>	<i>leofaþ</i>	<i>sægþ</i>	<i>hygþ</i>
pl.	<i>habbaþ</i>	<i>libbaþ</i>	<i>secgaþ</i>	<i>hycgaþ</i>
<u>Present subjunctive:</u>				
sg.	<i>hæbbe</i>	<i>libbe</i>	<i>secge</i>	<i>hycge</i>
pl.	<i>hæbben</i>	<i>libben</i>	<i>secgen</i>	<i>hycgen</i>
<u>Preterite indicative:</u>				
sg. 1.	<i>hæfde</i>	<i>lifde</i>	<i>sægde</i>	<i>hogde</i>
2.	<i>hæfdest</i>	<i>lifdest</i>	<i>sægdest</i>	<i>hogdest</i>
3.	<i>hæfde</i>	<i>lifde</i>	<i>sægde</i>	<i>hogde</i>
pl.	<i>hæfdon</i>	<i>lifdon</i>	<i>sægdon</i>	<i>hogdon</i>
<u>Preterite subjunctive:</u>				
sg.	<i>hæfde</i>	<i>lifde</i>	<i>sægde</i>	<i>hogde</i>
pl.	<i>hæfden</i>	<i>lifden</i>	<i>sægden</i>	<i>hogden</i>
<u>Imperative:</u>				
sg.	<i>hafa</i>	<i>leofa</i>	<i>sæge</i>	<i>hyge</i>
pl.	<i>habbaþ</i>	<i>libbaþ</i>	<i>secgaþ</i>	<i>hycgaþ</i>

**Figure 8.** The paradigms of class 3 weak verbs *habban* ‘to have’, *libban* ‘to live’, *secg(e)an* ‘to say’ and *hycg(e)an* ‘to think’.

### 3. DIATOPIC AND DIACHRONIC VARIATION IN OLD ENGLISH. APPLICATION TO THE NORMALIZATION OF VERBAL FORM.

This section presents some patterns of diatopic (interdialectal) and diachronic (intradialectal) variation that can both motivate and constrain the normalization of the noncanonical spellings of the forms of weak verbs. This section draws on the proposal made by de la Cruz (1986), who pays more attention to vocalic than to consonantal patterns of variation and opts for West-Saxon as the term of comparison when dealing with diatopic variation. As for diachronic variation, the discussion that follows is restricted to the West-Saxon dialect.

As regards diatopic variation in vowels, the contrast <æ>/<e> distinguishes the West-Saxon dialect from the others. West-Saxon favours <æ> as opposed to <e> as in other dialects. For instance, the preterite of *beran* ‘to bring’ is *bær* in West-Saxon whereas it has the form *ber* in Kentish and Southern Mercian. The infinitive *lætan* ‘to leave’ and *sætan* (preterite of *sittan* ‘to sit’) correspond, respectively, to *lētan* and *sēton* in Northumbrian, Kentish and Mercian. West-Saxon, Northumbrian and Mercian also present <æ> forms, as in *dælan* ‘to divide’, *hælan* ‘to heal’ and *lædan* ‘to lead’. Kentish, on the other hand, displays <e> forms in these forms, thus *dēlan*, *hēlan* and *lēdan*. As for the contrast between <ie> and <e, æ>, West-Saxon is the only dialect that presents <ie> forms in verbs such as *hliehhan* ‘to laugh’, *cierran* ‘to turn’, *hieran*, *geliefan* ‘to believe’ and *giefan* ‘to give’. Northumbrian, Kentish and Mercian rather show <e> or <æ> forms, as in *hlehhan/hlæhhan*, *cerran*, *hēran*, *gelēfan* and *gefan*. West-Saxon as a general rule opts for the diphthong <ea> where the other dialects display <e> or <æ>, thus *sceal* ‘shall’ in West-Saxon in contrast to *scel* or *scæl* in Northumbrian, Kentish and Mercian. The contrast between <eo> and <e> allows to distinguish West-Saxon and Kentish from Northumbrian and Mercian. While West-Saxon and Kentish prefer <eo>, as in *beorgan* ‘to protect’ and *flēogan* ‘to fly’, Northumbrian and Mercian show, respectively, *bergan* and *flēgan*. In Kentish, the forms related to <y>, which can be found in all the dialects, appear as <e>. For instance, verbs like *fyllan* ‘to fill’ and *ontynan* ‘to open’ become *fellan* and *ontenan* in Late Kentish. The contrast <e>/<eo> distinguishes West-Saxon, with <e> forms, from the other dialects, which display <eo> forms. For example, *beran* ‘to bring’ corresponds to *beoran* in Northumbrian, Kentish and Mercian. Along with West-Saxon, Northumbrian and Mercian also present <e> forms corresponding to Kentish <eo>, thus *sprecan* ‘to speak’ in all the dialects except in Kentish, which has *spreocan*. Another relevant contrast holds between West-Saxon, which has <i> forms where Northumbrian, Kentish and Mercian adopt <io> spellings. This is the case with *sidu* ‘habit’ and *wita* ‘adviser’, which correspond to *siodu* and *wiota* respectively in the other dialects. West-Saxon presents <ea> forms where the other dialects prefer <a> (but in Kentish <a> becomes <ea>, as in West-Saxon). Thus, *healdan* in West-Saxon and Late Kentish, as opposed to *haldan* in Northumbrian, Mercian and Early Kentish. West-Saxon is also characterized by the use of <ie> spellings where the other dialects have <io> or <eo>. For instance, *hierde* ‘shepherd’ and *gestriēnan* ‘to procreate’ in West-Saxon correspond to *hiorde* or *heorde* and *gestriōna* or *gestriōnan* in Northumbrian, Kentish and Mercian. Finally, West-Saxon presents <e> where the other dialects display <eo>

spellings (although in Kentish <oe> becomes <e>). Thus, *dēman* ‘to judge’ and *sēcan* ‘to seek’ in West-Saxon and late Kentish correspond, respectively, to *dōēman* and *sōēcan* in Northumbrian, Mercian and Early Kentish.

As far as the vocalic variation in West-Saxon is concerned, Early West-Saxon texts present <ie> forms, as in *hieran* ‘to hear’, *hīerde* (preterite of *hieran*), *begietst* (third person singular of the present indicative of *begietan* ‘to obtain’), *giefan* ‘to give’ and *wierp* (third person singular of the present indicative of *weorpan* ‘to become’). Nevertheless, this diphthong changes to <y> or <i>. Late West-Saxon has the following corresponding forms: *hȳran* or *hīran*, *hȳrde* or *hīrde*, *begystst* or *begitst*, *gyfan* or *gifan* and *wyrp* or *wirp*. Another diachronic contrast can be identified between <io> and <eo>. Early West-Saxon has <io> forms such as *clīopode* (preterite of *clīpian* ‘to call’) and *liofast* (second person singular present indicative of *libban* ‘to live’) where Late West-Saxon has *cleopode* and *leofast* respectively. Some vocalic contrasts are represented by consonants, such as <v[j]> / <v>. Thus, we have Early West-Saxon *frignan* ‘to ask’, *ligēþ* (third person singular present indicative of *licgan* ‘to lie’) and *sægde* (preterite of *secgan* ‘to say’) along with Late West-Saxon *frīnan*, *līp* and *sæde*. The contrast <y>/<i> holds between some nouns. For instance, Early West-Saxon presents *cyning* ‘king’, *cynn* ‘race’ and *dryhten* ‘lord’, which correspond, with the evolution <y> > <i>, to Late West-Saxon *cinig*, *cinn* and *drihten*. Early West-Saxon displays <ea> in verbal forms like *reahte* (preterite of *reccan* ‘to narrate’), *seah* (preterite of *sēon* ‘to see’), *geaf* (preterite of *giefan* ‘to give’) and *sceal* ‘shall’. The corresponding forms in Late West-Saxon are *rehte*, *seh*, *gef* and *scel*. To close this part, other verbal contrasts can be ultimately attributed to the process of simplification of inflections. For instance, canonical forms of strong verbs like *spreccen* (present subjunctive plural of *spreccan* ‘to speak’) and *sungon* (preterite indicative plural *singan* ‘to sing’) have weak forms like *spreccan* and *singan*.

#### 4. THE REGULARIZATION OF THE LEMMATA OF THE SECOND CLASS OF WEAK VERBS

As Burkhanov (1998) remarks, when organizing the corpus on which a dictionary is built it is necessary to lemmatize the textual (inflected) forms found in the corpus. In Burkhanov’s (1998: 122) words “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.” In this respect, Atkins and Rundell (2008: 325) point out that the headword “links all the information about one word together in one entry. In it goes the *canonical form* [italics as in the original] of the headword: the singular of nouns, the infinitive of verbs, the uninflected form of adjectives and adverbs, and so on”. Furthermore, as Jackson (2002: 179) puts it, “the criteria for determining what is a headword have important consequences for lexical description as well as for accessibility”.

In order to find the inflected forms of class 2 weak verbs, it is necessary to choose a set of inflectional endings of these verbs that are representative of their morphology in the sense of not being ambiguous with the inflectional endings of the other verbal classes. The inflections of class 2 weak verbs selected for lemmatization are 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 (*-odeþ*), the preterite indicative plural (*-odon*) and the preterite subjunctive plural (*-oden*).

The next step of the analysis is to search the *Dictionary of Old English Corpus* on the lexical database *Nerthus* for the inflectional endings presented above. In the process of lemmatization, the inflectional forms are grouped under the basic form or lemma. As illustration, Figure 9 presents the evidence and the lemmatization of *wilnian* ‘to wish’.

Inflectional form	Occurrences	Weak verb 2
<i>wilnast</i>	29	<i>wilnian(ge)</i>
<i>wilniað</i>	135	<i>wilnian(ge)</i>
<i>wilnian</i>	32	<i>wilnian(ge)</i>
<i>wilnianne</i>	3	<i>wilnian(ge)</i>
<i>wilniap</i>	4	<i>wilnian(ge)</i>
<i>wilnie</i>	10	<i>wilnian(ge)</i>
<i>wilniende</i>	13	<i>wilnian(ge)</i>
<i>wilnode</i>	96	<i>wilnian(ge)</i>
<i>wilnoden</i>	1	<i>wilnian(ge)</i>
<i>wilnodeþ</i>	3	<i>wilnian(ge)</i>
<i>wilnodon</i>	16	<i>wilnian(ge)</i>

Figure 9. The textual forms, occurrences and lemmatization of *wilnian*.

Throughout the process of lemmatization, some degree of regularization is necessary in order to adjust diachronic, dialectal or textual variants to the grammatical model. 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 *hiersumian(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 identified by de la Cruz (1986) as constituting instances of dialectal or diachronic variation. Vowel length is based on the information provided by the standard dictionaries of Old English, including the *Dictionary of Old English*.

A total of 187,000 inflectional forms have been searched for the inflectional endings of the infinitive (*-ian*), the inflected infinitive (*-ianne*), the present participle (*-iende*), the past participle (*ge-od*), the second person present indicative singular (*-ast*), the present indicative plural (*-iað/-iap*), the present subjunctive singular (*-ie/ge-ige*), the first and third person of preterite indicative singular (*-ode*), the second person of the preterite indicative singular (*-odest*), the preterite indicative plural (*-odon*) and the preterite subjunctive plural (*-oden*). A total of 1,064 lemmas of weak verbs from the second class have been found, which are listed in the Appendix.

The following inflected forms have been lemmatized by following the normalization patterns that result from the patterns of variation discussed in Section 3. Figure 10 presents the normalization based on diachronic contrasts while Figure 11 presents the diatopic criteria of normalization and their application.

## 1. &lt; y &gt; ≈ &lt; ie &gt;

Inflected Form	Lemma
<i>bescyrodest</i>	<i>bescierian</i>
<i>forgymeleasast</i>	<i>forgjemelēasian</i>
<i>forgymeleasodon</i>	<i>forgjemeleasian</i>
<i>gehyrsumast</i>	<i>hiersumian(ge)</i>
<i>gehyrsumige</i>	<i>hiersumian(ge)</i>
<i>gehyrsumod</i>	<i>hiersumian(ge)</i>
<i>gehyrsumodest</i>	<i>hiersumian(ge)</i>
<i>gehyrsumodon</i>	<i>hiersumian(ge)</i>
<i>gesmyrod</i>	<i>smierwan</i>
<i>geyrsod</i>	<i>iersian(ge)</i>
<i>gyrwast</i>	<i>gierwan(ge)</i>
<i>hyrsumast</i>	<i>hiersumian</i>

## 2. &lt; i &gt; ≈ &lt; ie &gt;

Inflected Form	Lemma
<i>gediglodon</i>	<i>dieglan(ge)</i>
<i>giddodest</i>	<i>gieddian</i>

## 3. &lt; i &gt; ≈ &lt; y &gt;

Inflected Form	Lemma
<i>asindrodest</i>	<i>äsyndran</i>
<i>drigast</i>	<i>drygan</i>
<i>gebricgod</i>	<i>brycgian</i>
<i>gemartirod</i>	<i>martyrian</i>
<i>gemartiroduon</i>	<i>martyrian</i>
<i>geminegod</i>	<i>mynegian(ge)</i>
<i>genihtsumige</i>	<i>nyhtsumian</i>
<i>gesingod</i>	<i>syngian(ge)</i>
<i>gestirod</i>	<i>styrian(ge)</i>
<i>underwirtwæloden</i>	<i>underwyrwalian</i>

## 4. &lt; e &gt; ≈ &lt; ea &gt;

Inflected Form	Lemma
<i>berefodon</i>	<i>berēafian</i>
<i>geernod</i>	<i>earnian(ge)</i>
<i>gemercod</i>	<i>mearcian(ge)</i>
<i>yrfewerdast</i>	<i>yrfewardian</i>

## 5. VCC ≈ VC

Inflected Form	Lemma
<i>forhttast</i>	<i>forhtian(ge)</i>
<i>geættrod</i>	<i>ætrian(ge)</i>
<i>geættrodon</i>	<i>ætrian(ge)</i>
<i>gehaddod</i>	<i>hādian(ge)</i>
<i>geh Wittod</i>	<i>hwitian(ge)</i>
<i>geliffæstast</i>	<i>lifæstan</i>
<i>geliffestast</i>	<i>lifæstan</i>
<i>gemannod</i>	<i>manian(ge)</i>
<i>gemicclige</i>	<i>miclian</i>
<i>gemicclod</i>	<i>miclian(ge)</i>
<i>gemicclodest</i>	<i>miclian</i>
<i>gerihtwissod</i>	<i>rihtwīsian</i>
<i>gesicclod</i>	<i>siclian</i>
<i>gewissod</i>	<i>wissian(ge)</i>
<i>innseglodon</i>	<i>inseglian(ge)</i>
<i>mannoden</i>	<i>manian(ge)</i>
<i>spellodon</i>	<i>spelian</i>
<i>weornnodon</i>	<i>weornian</i>
<i>widdast</i>	<i>wīdian</i>
<i>willnodon</i>	<i>wilnian</i>

Figure 10. Diachronic criteria of normalization.

1. < e > ≈ < æ >

Inflected Form	Lemma
<i>geclensod</i>	<i>clænsian(ge)</i>
<i>arefnodon</i>	<i>āræfnan</i>
<i>gefegnodon</i>	<i>gefægnian</i>
<i>geliffestast</i>	<i>lifæstan</i>

2. < e > ≈ < ie >

Inflected Form	Lemma
<i>forgemeleasoden</i>	<i>forgiemelesian</i>
<i>gedeglodon</i>	<i>dieglan(ge)</i>
<i>gehersumige</i>	<i>hiersumian(ge)</i>

3. < e > ≈ < ēa >

Inflected Form	Lemma
<i>berefodon</i>	<i>berēafian</i>

4. < e > ≈ < ea >

Inflected Form	Lemma
<i>geernod</i>	<i>earnian(ge)</i>
<i>gemercod</i>	<i>mearcian(ge)</i>
<i>yrfewerdast</i>	<i>yrfewardian</i>

5. < e > ≈ < eo >

Inflected Form	Lemma
<i>sweðerodon</i>	<i>sweoðerian</i>

6. < æ > ≈ < ēa >

Inflected Form	Lemma
<i>bescæwast</i>	<i>bescēawian</i>
<i>bescæwodon</i>	<i>bescēawian</i>
<i>forescæwode</i>	<i>forescēawian(ge)</i>
<i>gescæwige</i>	<i>scēawian(ge)</i>

7. < æ > ≈ < ea >

Inflected Form	Lemma
<i>gærcodest</i>	<i>gearcian(ge)</i>
<i>gærwode</i>	<i>gearwian(ge)</i>
<i>geærndod</i>	<i>earnian(ge)</i>
<i>gegærwige</i>	<i>gearwian(ge)</i>
<i>gegærwod</i>	<i>gearwian(ge)</i>
<i>gemærcod</i>	<i>mearcian(ge)</i>
<i>gemonigfældod</i>	<i>manigfealdian(ge)</i>
<i>monigfældodest</i>	<i>manigfealdian(ge)</i>
<i>yrfwærdast</i>	<i>yrfewardian</i>

8. < a > ≈ < ea >

Inflected Form	Lemma
<i>gemarcod</i>	<i>mearcian(ge)</i>
<i>gemonifaldod</i>	<i>manigfealdian(ge)</i>
<i>oferscadodest</i>	<i>ofersceadian</i>

9. < eo > ≈ < e >

Inflected Form	Lemma
<i>streowodon</i>	<i>strēwian(ge)</i>

10. < eo > ≈ < ie >

Inflected Form	Lemma
<i>cleopodon</i>	<i>cliepian</i>

11. < io > ≈ < i >

Inflected Form	Lemma
<i>cliopodon</i>	<i>clipian(ge)</i>

Figure 11: Diatopic criteria of normalization.

## 6. CONCLUSION

The main conclusions drawn from this research are the following. The list of lemmas of the second class weak verbs of Old English compiled in this article results from the analysis of the *The Dictionary of Old English Corpus*. A total of 187,000 inflectional forms have been searched for the least ambiguous inflectional endings in the paradigm of the second weak class. It seems feasible to conclude that after this analysis we have a more accurate knowledge of the relationship between Old English texts and the dictionaries of the language as regards the second class of weak verbs. The reason for this is that 1,064 lemmas of weak verbs from the second class have been found.

When it comes to regularizing the spelling of textual forms, so as to relate them to lemmas with more canonical orthography, this work has shown that normalization can be restricted to a number of correspondences based on dialectal and diachronic variation. Regularization, however, can also be limited by considering it unidirectionally. In spite of calling for further research in order to deal, for instance, with the alternative spellings of verbal prefixes, these criteria provide a principled motivation as well as a constraint on the limits of regularization.

## NOTES

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

## REFERENCES

- Atkins, B. T. S. and Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bosworth, J. and Toller, T. N. 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. diPaolo (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. diPaolo (ed.) with J. Price Wilkin and X.Xiang. (2004). *The Dictionary of Old English Web Corpus*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto. Available online at <http://www.doe.utoronto.ca/pages/pub/web-corpus.html>.
- Hogg, R. M. and Fulk, R. (2011). *A grammar of Old English. Volume 2. Morphology*. Oxford: Blackwell.
- Jackson, H. (2002). *Lexicography. An Introduction*. London: Routledge. <http://dx.doi.org/10.4324/9780203467282>
- Lass, R. and Anderson, J. (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. <http://dx.doi.org/10.1075/slcs.105.12mar>
- 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" *HREOW. 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. <http://dx.doi.org/10.1080/0013838X.2011.564776>
- 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. <http://dx.doi.org/10.1080/07268602.2012.744264>
- 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. <http://dx.doi.org/10.1080/00393274.2013.771829>
- 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., González Torres, E., Maíz Villalta, G., Mateo Mendaza, R., Novo Urraca, C., Veá Escarza, R. and Torre Alonso R. (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 Mateo Mendaza, R. (2013). "Nerthus. Outline of a lexicon of Old English". *Working Papers in Early English Lexicology and Lexicography*, 3.
- Martín Arista, J. and Cortés Rodríguez, F. (2014). "From directionals to telics: meaning construction, word-formation and grammaticalization in Role and Reference Grammar". In Gómez González, M. Á., F. Ruiz de Mendoza Ibáñez & F. González García (eds.), *Theory and Practice in Functional-Cognitive Space*. Amsterdam: John Benjamins. 229-250.
- Martín Arista, J. and Veá Escarza, R. "Assessing the semantic transparency of Old English affixation: adjective and noun formation". *English Studies*. Forthcoming.
- Mitchell, B., Ball, C. and Cameron, A. (1975). "Short titles of Old English texts". *Anglo-Saxon England*, 4: 207-221. <http://dx.doi.org/10.1017/S0263675100002799>
- Mitchell, B. and Robinson, F. (1993). *A guide to Old English*. Cambridge: Blackwell.
- Prokosch, E. (1939). *A Comparative Germanic Grammar*. Philadelphia: University of Pennsylvania.
- Pyles, T. and Algeo, J. (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. <http://dx.doi.org/10.1093/acprof:oso/9780199284139.001.0001>
- 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. <http://dx.doi.org/10.1017/CBO9780511812330>
- Stark, D. (1982). *The Old English weak verbs. A diachronic and synchronic analysis*. Tübingen: Niemeyer. <http://dx.doi.org/10.1515/9783111357140>
- 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. <http://dx.doi.org/10.2307/287733>
- Sweet, H. 1976 (1896). *The student's Dictionary of Anglo-Saxon*. Cambridge: Cambridge University Press.
- Nerthus. A lexical database of Old English*. Available at <http://www.nerthusproject.com>.



ofertrahtnian, offrian, oflīcian, ofmyrðrian, ofrian, ofsceamian, ofsetnian, ofsmorian, ofstician, ofwundrian, onāfæstnian, onclīfian, onclypian, onðenian, onðracian, ondruncnian, oneardian, onfægnian, ongefæstnian, onhātian, onhāwian, onhȳrsumian, onlōcian, onsceortian, onscunian, onstyrian, onwunian, openian, ortrūwian, pīlian, pinsian, plantian, platian, plegian, pluccian, prician, radian, rēadian, rēafian, rēafian, recenian, regnian, renian, rēnian, rēodian, rēonian, reordian, rēstan, rīcsian, rīhsian, rihtan, rihtwīsian, ripian, rīsan, rīxian, roscian, rūnian, rūnian, sacian, sadelian, sadian, sǣgan, sǣtian, samnian, sārgan, scamian, scēawian, sceorian, scēotan, scīmian, scirian, scotian, scrēadian, scrūtnian, scunian, scyldian, sēarian, seglan, segnian, selian, sengan, sēnian, seofan, seofian, seomian, sēowan, sibbian, sibsumian, siclian, sīdian, sīðian, sigefæstan, sigorian, simblian, singan, singian, slacian, slāwian, smēagan, smerian, smiðian, smocian, socian, sōðian, sorgian, sparian, spelian, spyrian, staðolian, stæððan, stalian, stician, stīfian, stigian, stihtan, stihtan, strælian, strangian, strēwian, stycian, styrian, sugian, sundfullian, sundrian, sūrian, suwian, swæðorian, swāmian, swearian, swefan, swefnian, sweogian, sweotolian, swerian, swician, swiðian, swīðrian, swīgian, swornian, sȳclian, sylian, symblian, syndrian, syngian, synnian, syrwan, sȳwian, taccian, tācnian, talian, tawian, telgian, temian, temprian, teochian, tēoðian, teohhian, teohian, tēorian, tīðian, tigðian, tihian, tilian, tīman, timbran, tīmian, tintegrian, tōclīfian, tōdihtnian, tōfēsian, togian, tōhaccian, tōliðian, tōlōcian, tōlȳsan, tōmearcian, torfian, tostihtan, tōsyndrian, tōtian, trahtian, trahtnian, trēowian, trēowsian, trucian, trumian, trūwian, trymian, tucian, twiccian, tȳdran, ðaccian ðeafian, ðafenian, ðafian, dagian, ðancian, ðancian, ðānian, ðeafian, ðeagian, ðeahtian, ðearfian, ðēawian, defran, ðegnian, delgian, ðēman, ðenian, ðēodan, ðēofian, ðēofian, ðēoran, ðēostrian, ðēowian, ðēowian, ðeowtian, derian, dīcian, dihtan, dihtnian, dīlegian, ðingian, ðingian, dolgian, ðolian, ðolian, ðracian, ðrafian, ðrēatian, ðrēatian, drēfan, ðreodian, ðreodian, ðrīstian, drohtian, drohtnian, dropian, droppetan, ðrōwian, ðrōwian, ðrōwian, drūgian, druncnian, drȳgan, ðūhsian, dunnian, ðurhborian, ðurhwunian, ðurhwunian, ðwærian, unārwurðian, unclānsian, undercrammian, underplantian, underwreðian, unðwærian, ungeðwærian, uninseglian, unmynegian, unrōtsian, unsyngian, untreowsian, untrumian, unweorðian, ūtlagian, wacian, wædlian, wǣgnian, wǣpnian, wealwian, wæterian, wagian, wandian, wanian, wānian, wansian, warian, warnian, weardian, wearnian, weddian, welgelīcian, wellīcian, wemman, wenian, wēodian, weorðian, weornian, wērgan, werian, wician, wīcian, wiðcostian, wiððingian, wiððingian, wiðerbrocian, wiðerian, wiðersacian, wiðerweardian, wiðerweardian, wiðheardian, wiðhogian, wīdlian, widmæran, wīfian, wilnian, wilnian, windwian, wīnhrēafetian, winian, wīsian, wisnian, wissian, wistfullian, wītegian, wītegian, witian, witnian, wlacian, wlancian, wlātian, woffian, wōgian, wracian, wracnian, wræcsīðian, wreðian, wrīdian, wrixlan, wuldorbēagian, wuldorfullian, wuldrian, wundian, wundrian, wunian, wunian, wynsumian, wyrðan, wyrnian, wyrnian, wyrntian, wyrntian, yðan, yðegan, yðgian, yfelian, yfelsacian, yflian, ymbðeahtian, ymbfrætewian, ymbhogian, yrfeweardian.