

STRONG VERB LEMMAS FROM A CORPUS OF OLD ENGLISH. ADVANCES AND ISSUES

Darío Metola Rodríguez

Universidad de La Rioja

Abstract: *The aim of this article is to devise the method of lemmatisation of strong verbs from a corpus of Old English with a view to maximising the automatic search for the inflectional forms, with the corresponding minimisation of manual revision of the verbs under analysis. The search algorithm, which consists of query strings and filters, is launched on the lemmatiser Norna, a component of the lexical database of Old English Nerthus. The conclusions of the article insist on the limits of automatic lemmatisation as well as the paths of refinement of the lemmatisation method in order to accommodate less predictable forms.*

Keywords: *lemmatisation, Old English, lexical database, morphology, orthography.*

LEMAS DE VERBOS FUERTES DESDE UN CORPUS DE INGLÉS ANTIGUO: AVANCES Y PROBLEMAS

Resumen: el objetivo del presente artículo es idear un método de lematización de verbos fuertes de inglés antiguo, con el propósito de maximizar la búsqueda automática de formas flexivas, con la correspondiente reducción en la revisión manual de los verbos en estudio. El algoritmo de búsqueda consiste en cadenas de búsqueda y filtros, ejecutados en el lematizador *Norna*, un componente de la base de datos léxica de inglés antiguo *Nerthus*. Las conclusiones del artículo insisten en los límites de la lematización automática, así como en las posibilidades de refinamiento del método de lematización para acomodar las formas menos predecibles.

Palabras clave: lematización, inglés antiguo, base de datos léxica, morfología, ortografía.

1. AIMS AND SCOPE

This article deals with the morphology of Old English and, more specifically, with the lemmatisation of strong verbs based on the textual forms in the *Dictionary of Old English Corpus* (henceforth DOEC).¹ It focuses on the analytical steps required by lemmatisation as well as on the implementation of such steps in the lemmatiser *Norna*, an integral part of the lexical database of Old English *Nerthus* (www.nerthusproject.com). Along with the compilation of the initial inventory of lemmas of strong verbs and the design of a lemmatisation method, this article aims at maximising the automatic search for the inflectional forms of the verbs under analysis, with the corresponding minimisation of manual revision. With these aims, this article contributes to the research line in the linguistic analysis of Old English pursued, among others, by García García (2012, 2013), Martín Arista (2012a, 2012b, 2013, 2014, fc-a, fc-b), Mateo Mendaza (2013, 2014, 2015a, 2015b, 2016), Novo Urraca (2015, 2016a, 2016b) and Veá Escarza (2012, 2013, 2014, 2016, fc). The relevance of the undertaking lies in the lack of a lemmatised corpus of Old English. The corpus of reference in the field of Old English studies, the DOEC, is annotated at text level (edition, author, prose/poetry/gloss) but does not offer word tagging, neither by attested form nor by lemma. Other remarkable corpora, like the *The York-Helsinki Parsed Corpus of Old English Poetry*

¹ This research has been funded through the project FFI2014-59110 (MINECO), which is gratefully acknowledged.

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

and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*, which have been parsed for syntax and can be searched for syntactic categories and functions as well as for inflection, do not tag the attested forms with information on the corresponding lemma. For instance, an inflectional form such as *berað* is marked as finite verb, present of indicative, but no link is given to the lemma *beran* ‘to bear’, with which it is not possible either to gather the inflectional paradigm of this verb or to quantify the occurrences of *beran* in the corpus. This has been done by *The Dictionary of Old English* (hereafter DOE) for the letters published so far, A-G, but is hardly accessible or unavailable for the letters H-Y from the standard dictionaries of Old English. Of the lexical and morphological classes of Old English, this article concentrates on strong verbs for two reasons. The first is descriptive, namely that strong verbs represent the starting point of lexical derivation, as has been pointed out by authors like Mailhammer (2006, 2008) and Martín Arista (2012a). The second reason for restricting the scope of the article to the strong verb is methodological. Strong verbs, as is well known, form the preterite and the past participle by means of, as in *beran* (infinitive) - *bær* (1st. preterite) - *bæron* (2nd. preterite) - *(ge)boren* (past participle) ‘to bear’. The different *Ablaut* patterns give rise to seven classes of strong verbs, shown in figure 1 (based on Hogg and Fulk, 2011; see also Prokosch, 1939, Van Coetsem, 1990 and Hogg, 1992).

	INFINITIVE	FIRST PRETERITE	SECOND PRETERITE	PAST PARTICIPLE
I	<i>scīnan</i>	<i>scān</i>	<i>scinon</i>	<i>(ge)scinen</i> ‘to shine’
II	<i>crēopan</i>	<i>crēap</i>	<i>crupon</i>	<i>(ge)cropen</i> ‘to creep’
III	<i>feohtan</i>	<i>feaht</i>	<i>fuhton</i>	<i>(ge)fohten</i> ‘to fight’
IV	<i>beran</i>	<i>bær</i>	<i>bæron</i>	<i>(ge)boren</i> ‘to bear’
V	<i>giefan</i>	<i>geaf</i>	<i>geafon</i>	<i>(ge)giefen</i> ‘to give’
VI	<i>standan</i>	<i>stōd</i>	<i>stōdon</i>	<i>(ge)standen</i> ‘to stand’
VII	<i>slæpan</i>	<i>slēp</i>	<i>slēpon</i>	<i>(ge)slæpen</i> ‘to sleep’

Figure 1. The *Ablaut* patterns of the seven classes of strong verbs².

When searching a corpus for strong verbs, the queries can be aimed not only to inflectional endings, as in the lemmatisation of weak verbs, but also to the changes to the root vowel characteristic of each of the classes presented in figure 1, as, for instance, in the pattern eo-*ea*-u-o, which defines certain Class III strong verbs such as *feohtan-feaht-fuhton-(ge)fohten* ‘to fight’. Therefore, the lemmatisation of this class is likely to pave the way for the analysis of the other classes because, as has been pointed out above, a remarkable number of derivatives are based on strong verbs, which, moreover, are very frequent in the texts, as is the case with verbs like *cūman* ‘to come’ and *giefan* ‘to give’.

With the aims and scope thus defined, the remainder of the article is organised as follows. Section 2 presents the lemmatiser *Norna*, a building block of the relational lexical database of Old English *Nerthus*. Section 3 describes the search algorithm used for assigning lemmas to the inflections of the strong verbs found in the DOEC. Section 4 offers the results of the application of the query strings and filters of which the search algorithm consists to the data in the DOEC. The results of the analysis are discussed in section 5. Finally, section 6 draws the main conclusions, in the line of maximising the automatic search for the inflectional forms of the verbs under analysis with the corresponding minimisation of manual revision.

2. THE LEMMATISER *NORNA*

The lemmatiser *Norna* has been implemented in Filemaker database software (version 14) and is based on a concordance and an index to the DOEC. The corpus has been concorded by word and by fragment. The concordance by word, displayed in figure 2, consists of three million lines, one per word in the corpus. The concordance by fragment contains around two hundred thousand fragments of texts identified with the short title with which they appear in the DOEC, as in *Eala ðu cleric ne wana ðu æfre wexbreda fram sidan* [Abbo 000100 (103.1)].

² This classification of strong verbs follows the Clark Hall dictionary, except as regards the subclasses a, b, etc of Class III and Class VII, which are based on Krygier (1994). On strong verb classes, see also Levin (1964) and Mailhammer (2006, 2007, 2008).

Prefield	Conc Term	PostField
urh ofermodignysse & uppahafennysse þurh forligr & eafiac, gifernys, druncennis, oferfyll, forliger,	unrihtþæmed	þurh gitsunge & ydel wuldor, þurh yrre & unrotnys
e þone feorþan dæl þæs yrfes. Gyf hwylc wif sy þe	unrihtþæmed	, deofolgild, modignys, <forswarung>, þæt is mæne
æt is morþor & stala & mæncaþas & unrihtgitsung & ytsunge & æfest & ydelne gylp & stala & reafiac &	unrihtþæmed	fremme, hyre wite sy in hyre weres handum. <Wif- & gyfernysse & tælnysse & lease gewitnysse. & beo
yndon morþor & stala & mæncaþas & unrihtgitsung & rnon his sawle reste. & æretas & oferdruncolns & dinesse & þurh upahafennesse, þurh forliger & þurh yrre micel & teona & fylþ hyre na byþ oferhelud.	unrihtþæmed	& oferdruncennys & morþor & mæne aþas & leasunga & & gifernysssa & tælnysssa & lease gewitnysssa. Ac lu & idel gylp & unsibbe & stala & leasunga & mæncað
wif wrað þan geongan cnapan forþan þe he ascunode	unrihtþæmed	, þurh gitsunge & þurh idel wuldor, þurh yrre & þu
wif wearþ wraþ þam geongum cnapan; & he ascunode	unrihtþæmed	wifes on uppahafennysse eagena & on bræwum hyre b
yorðe, and <berenda> manna forhealdnessa, þæt is	unrihtþæmed	. Hit gelampt sic þa sume dæge þæt iosep was ana o
þe hyt geecearnað. þi we cow lærað þæt ge cow wið	unrihtþæmed	. Hit gelamp sume dæg þæt iosep was ana innan his
he his luste fulgæð, he hine genyðerað, oððe þurh	unrihtþæmed	, and morðæda, and stala, and mane aþas, and lybl
htra, þæt is oformodes & manslehtas & mæne aþas,	unrihtþæmed	georne beorgen, & ge scyldað cow wið þa bealewan
nesse & oferfyll & unrihtwisesse & unnytnesse & lice þa synleahtras, þe us forbodene synd, þæt is	unrihtþæmed	oððe þurh oðer yfel. Witodlice ne mæg sio hungrie
rman wyllað her on life. Uton eac beorgan us wið	unrihtþæmed	, morþor & gitsunga, lease gewitnysssa, nið & yrre,
lgan sang and forlætað erre and druncennesse and orþor and stala and mancaþas and unrihtgitsung and cennys, mansliht and leasung, reafiac and stalu,	unrihtþæmed	& æretas & calogalnesse & oferfyll & unsibbe & d
And we lærað, þæt man wið fulne galscype and wið	unrihtþæmed	and æretas and oferdruncennesse and idel gylp and
yllende and calle yfele dædan forletende, and eac	unrihtþæmed	and gescydan us wið þa bealewan synne and wendon
nd æghwylc Cristen mann eac for his Drihtenes ege	unrihtþæmed	and healdað þone halgan sunnandæg, and beþencað þ
e. And we lærað, þæt man wið fulne galscipe & wið	unrihtþæmed	and gifernesse and <tælnessa> and leasa witnysssa.
or Gode ge for worolde. & æghwylc Cristen man eac	unrihtþæmed	and geflit, æfest and eac þara mandæda, þe mannum
		and wið æghwylcne æwbrice warnic symble. And we l
		forfeon and eacne mansliht forbugan. Healdað min
		georne forbuge & godcunde lage rihtlice healde. A
		& wið æghwylcne æwbrice warnige symle. & we lærað
		georne forbuge & godcunde laga rihtlice healde. &

Figure 2. The word concordance to the DOEC.

The word concordance to the DOEC turns out an index of approximately one hundred and ninety thousand inflectional forms, which constitutes the target of the analysis. In Figure 3, the leftmost column lists inflectional forms, the occurrences are quantified next to the right; the column called *headword* shows the lemma that corresponds to each inflectional form, and the three columns to the right present the concordance prefield, the concorded word and the concordance postfield. As can be seen in figure 3, *abelgan* is the lemma attributed to the inflectional forms *abolgen*, *abulgon*, *abelge*, *abealh*, *abulge*, *abelgan*, *abelgeð*, *abealg*, *abelgað*, *abulgan*, *abolgenne*, *abelh*, *abulgen*, *abelged*, *abelige*, *abeligan* and *abelhð*.

InflectionalForm	Occurrences	Headword	DOEC_Conc_by_Word::Prefield	Conc Term	DOEC_Conc_by_Word::PostField	Stron
abacen	1	abacan (VI)	af þe bið of corne gegearcod and ðurh fyres	abacen	. mage beon awend to cristes lichaman. oððe	
abannan	9	abannan (VIIId)	ærmdon & slogon swa swa hi ferdon. þa het se	abannan	ut ealne þeodscepe of Wessexum & of	
abanne	2	abannan (VIIId)	on wærlagan þe þa wic bugað. þeah ge þa calle	abanne	>, and cow eac gewyrce widor sæce, ge her	
abarian	3	abarian (2)	lige þa cunnan gelacnian heora clyfmeda wunda	abarian	& geswutelian. godes dæges & nihtes sig caru	
abarast	1	abarian (2)	e ðe on ðam huse beo, hæbbe frîð mid ðe. & gyf	abarast	ure spræce, we ne beoð forsworene. Ð æt	
abolgen	20	abelgan (IIIb)	i þonne nabað nane unrihtwisesse, ne heora	abolgen	, þonne beo we calle to hospe gedone þurh	
abulgon	10	abelgan (IIIb)	eardan sægð, þæt we magon gegladian þone þe	abulgon	. Se þe þe his breþer hosp gecwyð, se bið þeahtes	
abelge	9	abelgan (IIIb)	d welwillendum dihte, þeah ðe ure yfelclys him	abelge	, and we þonne swingla for urum synnum	
abealh	8	abelgan (IIIb)	rædes behofað oððe gif he miltsað þam menn þe	abealh	oððe gif he gegerodne of æfnyde gedeð	
abulge	7	abelgan (IIIb)	on þone god, and his biggengon sædon, gif him	abulge	, þæt seo heofon sona sceolde <æfallan>, and	
abelgan	6	abelgan (IIIb)	wæron acwealde mid swordes ecge, þa þa hi	abelgan	heora scyppende in þam forbodenan &	
abelgeð	2	abelgan (IIIb)	we mildheortnysse ne habben ofer þa mæn, þe	abelgeð	, þæt on domesdæge drihtenes mildheortnysse	
abealg	2	abelgan (IIIb)	wilnige ðæt he ðone mon eft lufian mæge þe him	abealg	, ðonne he hit ðeah forgifan sceal, forðæm, gif	
abelgað	2	abelgan (IIIb)	tælan ure þa nyxtan ne ne <hyrwan> <hig>. Gif	abelgað	ure efenhæfden, þonne wregað we <þæt>. &	
abulgan	2	abelgan (IIIb)	an mid godum dædum. þeah ðe we hine ær mid	abulgan	, he wile sona onfon þa soðan hreowe and	
abolgenne	2	abelgan (IIIb)	larward is from fieder minum. & geherende þa	abolgenne	werun be þam twæm broþrum. hælend þa	
abelh	1	abelgan (IIIb)	i sin yfele. & wilt, þæt þin lif si yfel? On hwon	abelh	þe þin lif? Forhwon wilt þu beon ana yfel	
abulgen	1	abelgan (IIIb)	n, þy les þa halgan treow þurh heora wop &	abulgen	. Ond ne geherde ða ondsware þara treowa ma	
abealch	1	abelgan (IIIb)	hrusan hordærna sum, eacencræftig, oððæt hyne	abealch	mon on mode; mandryhtne bæc fited wæge,	
abelged	1	abelgan (IIIb)	cop, seo genemned is Utricolensis. Se hæfde	abelged	þone wellhreowan cuning Totila, þa þa he mid	
abelige	1	abelgan (IIIb)	e tæleð, oððe his gesceafte wyrgeð, þeah hine	abelige	; & þurh þyllicu þing gefirenað seo tunge ofi.	
abeligan	1	abelgan (IIIb)	anna bearnum. And eft ymbe lytel ongan	abeligan	god for sunnandæges weorcum, and þa ongan	
abelhð	1	abelgan (IIIb)	t wordes oððon weorces, he dryhð deofles willan	abelhð	his Drihtne swiðor þonne he beforfte. Ne	
abeað	29	abeodan (II)	sona to ðam mædene þæt ylce ærende þe his	abeað	, ac Agnes wifsoc, sæde þæt heo nolde þæs	
aboden	12	abeodan (II)	is hrvre lifes brucan. þa was ærende æðelum	aboden	in burum. ne was him bleað hwe. ah he wæs	

Figure 3. Layout of the lemmatiser Norma.

On the lemmatiser *Norna*, inflectional forms are assigned a lemma on the basis of a reference list of verbs from each strong class that has been retrieved from the lexical database *Nerthus* and supplemented with information from Krygier (1994) and Hogg and Fulk (2011). For example, the reference list of strong verbs from Class VI is as follows in figure 4 (verbs with weak forms excluded).

ābacan	bestandan	hlædan(ge)	oðstandan
ācalan	bewacan	inafaran	oðwadan
acan	bewadan	inefaran	orhlædan
ādragan	calan	infaran	sceacan
āfaran	dafan(ge)	inhebban	singalan
āgalan	dragan(ge)	inhlædan	spanan(ge)
āgnagan	eftalan	instæppan	standan(ge)
āgrafan	eftdragan	instandan	tacan
āhladan	emfaran	inwadan	tōfaran
āhliehhan	faran(ge)	midfaran	tōhlædan
alan	forbacan	midstandan	tōsceacan
analan	forebacan	midwadan	tōstandan
andswerian	forefaran	misfaran	ðurhfaran
ānstandan	foresacan	ōfaran	ðurhforfaran
ānswerian	forestandan	ofāsceacan	ðurhstandan
ānwacan	forewadan	ofcalan	ðurhwadan
ānwadan	forfaran	oferāhebban	ðurhwrecan
āsceacan	forgnagan	oferfaran	undergestandan
āscieppan	forsacan	ofergestandan	underhebban
āspanan	forsceacan	oferstæppan	understandan
āstandan	forscieppan	oferstandan	unfaran
āwacan	forspanan	oferwadan	unhebban
āwascan	forstandan	offaran	untōsceacan
æalan	forwadan	ofsceacan	ūphladan
ærwadan	forōfaran	ofstandan	ūtāfaran
æspanan	forōgefaran	oftacan	ūtdragan
ætforan	forōsacan	onāhebban	ūtforan
ætstandan	forōsceacan	onalan	wacan
ætwardan	forōscefafan	onfaran	wadan(ge)
ætwascan	forōwadan	ongalan	wascan
bacan	framāscæcan	ongēanstandan	wiðerstandan
bedafan	framstandan	ongespanan	wiðfaran
bedragan	framwadan	onhlædan	wiðsacan
befaran	fullfaran	onsacan	wiðscefafan
begalan	fullwadan	onsceacan	wiðstandan
begnagan	galan(ge)	onstandan	ymbfaran
begravan	geondfaran	onwacan	ymbstandan
behlædan	geondwadan	onwadan	
behliehhan	gnagan	oðfaran	
bespanan	grafan	oðsceacan	

Figure 4. The reference list of Class VI strong verbs.

Given the inflectional forms from the DOEC and the reference list, the assignment of lemmas can be done automatically or manually. In order to automatise the process of lemmatisation, it is necessary to define a search algorithm of which results do not require a great deal of manual revision. Nevertheless, the diachronic and diatopic variants included in the DOEC as well as the various spellings with which many words appear in the corpus seem to exclude a fully automatic search procedure and to predict manual revision. This question is tackled in the next section.

3. THE SEARCH ALGORITHM

This section describes the steps taken in order to develop a search algorithm with which to launch queries on the lemmatiser *Norna* and find the inflectional forms of the strong verbs in the DOEC. The target of the first step of the design of the algorithm is the simplex word. The underived verbs in the reference list of the seven strong classes have been inflected for the infinitive, inflected infinitive, present participle and past participle; present indicative singular and plural, present subjunctive singular and plural, preterite indicative singular and plural, preterite subjunctive singular and plural, imperative singular and plural. The list of inflections for *crēopan* ‘to creep’ is offered in (1):

(1)

- a. Finite forms: *crēopan, crēap, crupon, crēope, crēopest, crēopst, crīepest, crīepst, crēopeð, crēopeþ, crēopð, crēopp, crēopt, crīepeð, crīepeþ, crīepð, crīepþ, crīept, crēopað, crēop, crēopaþ, crēopen, crupe, crupen*.

Non-finite forms: *cropen, crēopanne, crēopenne, crēopande, crēopende, crēopinde*.

- b. Finite forms: *crýpan, crēap, crupon, crýpe, crýpest, crýpst, crýpeð, crýpeþ, crýpð, crýpþ, crýpt, crýpað, crýpaþ, crýpe, crýpen, crýp*.

Non-finite forms: *cropen, crýpanne, crýpenne, crýpande, crýpende, crýpinde*.

Several comments are in point with respect to the inflectional paradigm given in (1). Firstly, the inventory of forms has been duplicated so as to account for the potential effects of i-mutation. The Bosworth-Toller and Sweet dictionaries agree in giving the strong verb *crēopan* ‘to creep’ but Clark-Hall adds the 3rd. person singular present indicative forms *crýpeð, crýpð* in a cross-reference to *crēopan*. Indeed, the DOEC has two occurrences of the syncopated form *crýpð*. For this reason, the i-mutated inflectional forms in (1b), such as *crýpeð* and *crýpð*, have been included for Class II strong verbs in the search algorithm. For the other classes of strong verbs, the effects of i-mutation might be seen in endings like the ones provided in (2), following Campbell (1987).

(2)

-icð, -icð, -icþ, -iðð, -iþþ, -ieçð, -ieçþ, -ieðð, -ieþþ, -ielpð, -ielþþ, -ielt, -iett, -ilpð, -ilþþ, -ilt, -itt, -iðð, -iþþ, -ycð, -ycð, -ycþ, -ycþ, -yðð, -yþþ, -ylpð, -ylþþ, -ylt, -ytt

Secondly, the alternative spelling <ð> / <þ> justifies the duplication of forms like *crēopeð / crēopeþ, crēopð / crēopp, crīepeð / crīepeþ, crīepð / crīepþ* and *crēopað / crēopaþ*. Thirdly, the syncope of the vowel of the second third person singular present indicative explains the inclusion of pairs like, respectively, *crēopest / crēopst* and *crēopeð / crēopð*. In the fourth place, the inflectional endings with consonant assimilation such as *crēopt, crīept* and *crýpt* appear in the list along with the canonical forms *crēopeð, crīepeð* and *crýpeð*. Finally, dialectal variation motivates the presence of forms like *crēopenne, crēopande* and *crēopinde* together with the West-Saxon inflections *crēopanne* and *crēopende*.

The second step of the design of the algorithm focuses on the complex word. It consists of the compilation of a list of elements that may be attached to simplex strong verbs to form derived or compound verbs. *A priori*, the inventory of preverbal elements, which has been retrieved from the lexical database of Old English *Nerthus*, includes affixes with a very specific meaning, such as the negative prefix *un-*, the pejorative prefix *mis-* as well as the aspectual prefixes *eft-* and *ed-*; the Germanic pure prefixes *ā-*, *be-*, *for-*, *ge-*, *of-*, *on-*, *tō-* (de la Cruz, 1975); the spatial and temporal adverbs and prepositions that are going through grammaticalisation resulting in a telic marker (Brinton and Traugott, 2005; Martín Arista and Cortés Rodríguez, 2014), including *adūn-*, *æfter-*, *æt-*, *āweg-*, *beforan-*, *betwux-*, *ðurh-*, *forð-*, *fore-*, *fram-*, *geond-*, *in-*, *niðer-*, *oð-*, *ofer-*, *onweg-*, *under-*, *ūp-*, *ūt-*, *wið-*, *wiðer-*, and *ymb-*; and fully free forms that appear in compound verbs such as *āgēn-*, *and-*, *ðri-*, *dyrn-*, *efen-*, *ful-*, *hearm-*, *mæg-*, *mān-*, *nyd-*, *riht-*, *twi-*, *wyrg-*. With the preverbal elements, the roots and the set of inflections as presented above, the third step of the design of the search algorithm is the definition of query strings that can be launched on Filemaker. Four query strings (QS) have been defined. QS1 is aimed to the stems and inflections given in (1), and the corresponding stems in the remaining six classes of strong verbs. With the operator for exact matches in Filemaker (==), the part of QS1 that search the corpus for the inflections of *bēodan* can be seen in (3).

(3)

==beodan, ==bead, ==budon, ==beode, ==bead, ==biedest, ==biedst, ==bietst, ==biest, ==bude,
==beodeð, ==beodeþ, ==biett, ==bietð, ==bietþ, ==bead, ==beodaþ, ==beodað, ==budon, ==beode,
==bude, ==beoden, ==buden, ==beod, ==beodað, ==beodaþ

The target of the second query string is the prefixation with *ge-*, the most frequent prefix in Old English (Martín Arista, 2012b) to such an extent that most strong verbs have a simplex and a complex form prefixed with *ge-*. QS2 for *gebēodan* is shown in (4).

(4)

==gebeodan, ==gebead, ==gebudon, ==gebeode, ==gebead, ==gebiedest, ==gebiedst, ==gebietst,
==gebiest, ==gebude, ==gebeodeð, ==gebeodeþ, ==gebiett, ==gebietð, ==gebietþ, ==gebead,
==gebeodaþ, ==gebeodað, ==gebudon, ==gebeode, ==gebude, ==gebeoden, ==gebuden, ==gebeod,
==gebeodað, ==gebeodaþ

QS3 has been defined for accounting for complex strong verbs with preverbs different from *ge-*. The wild card * in (5) stands for any preverbal element attached to the base *bēodan* and its inflections.

(5)

==*beodan, ==*bead, ==*budon, ==*beode, ==*bead, ==*biedest, ==*biedst, ==*bietst, ==*biest, ==*bude,
==*beodeð, ==*beodeþ, ==*biett, ==*bietð, ==*bietþ, ==*bead, ==*beodaþ, ==*beodað, ==*budon,
==*beode, ==*bude, ==*beoden, ==*buden, ==*beod, ==*beodað, ==*beodaþ

QS4 is the least specific query. It search the corpus for the stems of strong verbs with any preverbal and inflectional segment, thus the introduction of the wild card * to the left and to the right of the stem. This can be seen in (6).

(6)

==*beod*, ==*bead*, ==*bud*, ==*bod*, ==*bied*, ==*biet*, ==*biest*

These query strings have been launched sequentially: QS1, QS2, QS3, QS4. After the submission of each query, its hits have been tagged on the lemmatiser *Norna*, so that the tags from previous queries could guide the tagging of the hits of subsequent queries. This simplifies the overall task because, in spite of being likely to find some unexpected spellings, QS4 is redundant with respect to QS1 (endings) as well as QS2 and QS3 (preverbs). Moreover, given its wide scope, it is predictable that this query string turns out a remarkably high number of matches. For this reason, the final step in the design of the search algorithm is the definition of filters that can put aside at least part of the undesired results of SQ4, so that manual revision can be kept to a minimum. Two filters have been devised. Filter (F) 1 is intended to isolate verbal forms. It narrows down the hits of QS4 to inflectional forms that end with *on*, *-odon-*, *-ast*, *-est*, *-ost*, *-ð*, *-þ*, *-iað* and *-iap*, thus the operators *==* and ***. The application of F1 to the 17,138 hits of SQ4 reduces this figure to 1,939. F1 is presented in (7).

(7)

==*-on, ==*-odon-, ==*-ast, ==*-est, ==*-ost, ==*-ð, ==*-þ, ==*-iað, ==*-iap

F2 is aimed at finding spelling variations in the consonantal endings of verbal forms. It is applied in two steps. The first step selects the inflectional forms that end in a consonant, as can be seen in (8).

(8)

==*b, ==*c, ==*d, ==*f, ==*g, ==*h, ==*l, ==*m, ==*n, ==*p, ==*r, ==*s, ==*t, ==*w, ==*x, ==*y, ==*ð, ==*þ

The second step of F2 targets members of the non-verbal classes as well as weak verbs by deleting inflectional forms that end in -on, -en, -an, -es, -um, -end, -as, -est, -ost, -ed, -od, -ig, -ic, -ing, -ung, -un, -us, -nes, -er, -or, -ur, -iað, -iaþ. It must be noted that F2 also puts aside the endings -iað, -iaþ, which are selected by F1. When applied to the outcome of SQ4, the first step of F2 reduces its hits from 17,138 to 10,305, which, after the application of the second step of F2, result in 3,533 hits. The second step of F2 is displayed in (9).

(9)

==*on, ==*en, ==*an, ==*es, ==*um, ==*end, ==*as, ==*est, ==*ost, ==*ed, ==*od, ==*ig, ==*ic, ==*ing, ==*ung, ==*un, ==*us, ==*nes, ==*er, ==*or, ==*ur, ==*iað, ==*iaþ

4. THE OUTCOME OF THE SEARCH ALGORITHM

The application of the algorithm to the search for Class IV strong verbs turns out the results presented below. The inflectional forms attributed to each lemma can be seen in the Appendix. The overall quantitative results by class are tabulated in table 1.

Table 1: Hits by strong verb class.

Class	QS1	QS2	QS3	QS4+		Total
				F1	QS4+F2	
I	359	150	582	60	42	1,190
II	293	134	263	12	75	774
III	565	206	775	290	175	1,926
IV	116	60	287	69	44	688
V	343	145	747	123	68	1,428
VI	215	68	532	35	37	883
VII	509	252	1,136	77	69	1,827
Total						8,716

As can be seen in Table 1, the most frequent verb classes are III, V, and I, with 1926, 1428 and 1190 inflectional forms respectively. QS3 turns out the highest number of hits in all the classes, except Class II. These results are discussed in the following section.

Apart from the quantitative data, the analysis has turned out a normalised list of preverbal and inflectional endings. The inventory of attested preverbal forms includes the ones presented in Figure 5 (variants between brackets). Throughout the lemmatisation process, the forms between brackets have been normalised to the canonical forms preceding them in Figure 5.

ā-, æfter- (æft-, æftyr-, efter-, eftyr-), æt- (ad-, æd-, at-, ēt-, et-, ot-), and- (iand-, ond-, ont-), be- (b-, beah-, beo-, bet-, bew-, bi-, bī-, bic-, big-, bio-, bis-), ed-, efen- (æfen-, efn-, emn-), eft- (æft-, oft-), for- (f'-, f'r-, fær-, færn-, far-, feor-, fer-, fern-), fore- (foren-, fores-, forn-, fors-), forþ- (fort-, ferþ-), fram- (frum-, from-, frun-, frym-, frymft-, frymftþ-), full- (ful-), ge- (cg-, īg-, ēh-, eīg-, i-, g-, ga-, gæ-, gæan-, gær-, gad-, gan-, gar-, ged-, gen-, gem-, ger-, gi-, gif-, gim-, gy-), geond- (giond-, gind-, gio-, gion-, gin-, geon-, gon-, geo-), in- (ine-, ing-, inn-), mid- (me-, met-, mi-, med-, mið-, mod-), mis- (miss-, mus-), of- (æf-, af-, off-), ofer- (eofer-, eofor-, ofær-, ofern-, ofor-, of'-, ofyr-, ouer-, ouyr-, ofer-), on-, oþ- (oep-, ūþ-), tō-, twi- (twig-, twy-), þurh- (þorh-), un-, under- (und-, undern-, ynder-), ūp- (upp-), ūt- (utt-, vt-), wiþ-, wiþer (wiþere-, wiþyr-), ymb- (ym-, ymbe-, emb-, embe-, eme-, imb-).

Figure 5. Attested preverbal forms and their normalisation.

The inflectional endings in figure 6 have also been identified. For the sake of clarity, the variants are grouped under the canonical form and displayed between brackets.

-að; -ap (-ad; -æd; -æð; -æp; -at; -ath); -an (-æn); -ð; -þ (-d; -th); -e (-eg; -cg; -ch; -cht; -gc; -h; -o); -eð; -eþ (-cg; -ch -cht; -et; -eth; -eað; -eaþ; -ed; -eid; -gc; -h, -id; -ið; -iþ; -ieð; -ieþ; -it; -ith; -oð; -ud; -up; -yd; -yð; -yþ; -yt); -en (-in; -yn); -est (-æs; -æst; -ast; -æt; -esð; -esp; -is; -ist; -ust; -ys; -ysð; -yst); -on (-don; -onn); -st (-sð; -s); -tst (-dst); -t (-tt).

Figure 6. Inflectional endings and their variants.

5. DISCUSSION

The lemmatisation method, based on a set of query strings and filters implemented on a lexical database, has clear advantages but also raises some issues. A comparison with the inflectional forms provided by the DOE (A-G) has shown that the accuracy of the search algorithm is around 80%. In other words, 10% of the hits do not correspond to inflections listed in the DOE and, conversely, 10% of the forms in the DOE have not been found by the lemmatiser in its present state. This is due to a number of reasons that must guide future research. Beginning with the hits that are not found in the DOE, three main shortcomings have been identified: strong nouns that coincide with the third person singular of the present indicative of potential lemmas, such as *bisceop* (from *biscop*, not from ****besceopan**); adjectives derived from strong verbs that coincide in form with the third person singular of the present indicative of a verbal lemma, such as *eacene* (from *ēacen*, not from *ēacan*); forms overlapping with the paradigm of weak verbs (*beslepen*, from *beslēpan*, weak, not from *beslæpan*). Among the inflections provided by the DOE that have not been found by means of automatic searches we find early forms like *adriofan* (*ādrīfan*), late forms such as *arysað* (*ārisan*), consonantal contrasts due to Verner's Law, such as *seoðan*, *seað* vs. *sudon*, *sodden* (*seoðan*); unpredictable <k> or <ch> for <c>, as in *belukeð*, *belochene* (*belūcan*); unpredictable <v> for <w>, as in *bisvicen* (**beswicen**); unpredictable <c> for <g>, as in *forcnad* (**forgnīdan**); unpredictable <y> for <i>, as in *byðed* (**biddan**), this example also illustrating unpredictable <ð> for <d>; assimilation of the third person singular, present indicative ending to <t>, as in *aworpet* (*āweorpan*); and loss of <h> before <r>, as in *ætrinð* (*æthrinan*). Finally, the DOE has inflected participles that were not considered in this work, such as the present participle *cēosendum* (*cēosan*) and the past participle *besmitenes* (*besmītan*).

Apart from the accuracy of the lemmatiser, two issues have arisen throughout the research that need some comment. In the first place, the hits of the search algorithm not only show a remarkable degree of variation, as has been described above, but also require desambiguation. That is to say, some hits certainly constitute accurate results in the sense of belonging to the inflectional paradigms of strong verbs, but they may pertain to more than one class, as in the pairs in (10). Provisionally, these inflections have been assigned to both potential lemmas.

(10)

abudon: *ābēodan* (II);
abudon: *ābūan(ge)* (VIIg; st. with wk. pret./p.p.)
abærst: *āberstan* (IIIc);
abærst: *āberan* (IV)
bitt: *bītan(ge)* (I);
bitt: *biddan(ge)* (V; st. with wk. pres.)
feall/feallap/gefeoll/fealleð: *fēolan(ge)* (IIIb);
feall/feallap/gefeoll/fealleð: *feallan(ge)* (VIIc)
fersceat/forsceat: *forscēotan* (II);
fersceat/forsceat: *forscēadan* (VIIa; st. with wk. pret./p.p.)
onhlad: *onhlīdan* (I);
onhlad: *onhlædan* (VI)

The final problem has to do with the existence of a derived verb for which no underived base can be proposed. All the sources account for the existence of the strong verb *forlēosan* 'to lose, abandon, let go'. However, none of them considers that the verb **lēosan* is attested and concur in marking it with the asterisk used for reconstructed forms, thus, for instance, the DOE. Given the stepwise character of the search strings proposed above, it is necessary to add the verb *lēosan(ge)* to the reference list in order to account for its derivatives.

6. CONCLUSION

This article has described the process of lemmatisation of the strong verb forms in the DOEC by means of a lemmatiser that is implemented on a lexical database and consists of an index, a concordance, a reference list and a set of query strings and filters.

The implementation of the search algorithm on a lexical database has proven remarkably useful and convenient not only because information can be easily stored, searched and retrieved but also because the results can be sorted in alphabetical order, which implies that similar forms appear close to each other. This has two advantages. First, the identification task is not as time consuming as it would be if the corpus was searched manually. Second, and more important, variant spellings can be identified more easily and quickly compared with nearby words, which results in both the identification of a higher number of inflectional forms and, in some cases, the identification of new lemmas.

The methodology based on a search algorithm on a lexical database has had some effects on the resulting inventory of lemmas, at least in two respects. Given that a canonical form cannot be a variant form of another prefix, two lemmas have been proposed when inflectional forms have been found containing both preverbal items. At the same time, the separation of some affixes has led to the replacement of a lemma for a new one, instead of keeping both.

Overall, the automatic searches on the lemmatiser *Norna* have an accuracy of approximately 80%. Considering that the inflectional forms of the strong verbs beginning with the letters H-Y remain largely unpublished, finding eight out of every ten forms may represent a significant advance. On the other hand, future research must be conducted with the aim of refining the search algorithm in the directions indicated in this article.

APPENDIX: THE LEMMAS AND INFLECTIONS OF CLASS IV

āberan: abær, abære, abæron, aber, aberað, aberan, aberð, abere, abereð, aberon, abireð, aboren, aborene, *Æbær*; *ābrecan*: abræc, abræcan, abræce, abræcen, abræcenne, abræcon, abrecað, abrecan, abrecane, abrecanne, abrecap, abrecenne, abrecð, abrece, abrecende, abrecenne, abrecon; *ācuman*: acom, acoman, acome, acomon, acuman, acume, acumen, acymð, acymp; *ācwelan*: acwæl, acwælan, acwæle, acwælon, acwelað, acwelan, acwelð, acwele, acwellan, acwelon, acwið, acwolen; *āðweran*: aþwer; *ætberan*: *Ætbær*, *Ætbæron*, *Ætbærst*, *Ætberan*, *Ætbere*, *Ætberst*, *atbærst*; *āniman*: anim, Animað, animan, animð, anime, anumen, anumenne, anumenre, *āteran*: aterað; *beberan*: beborene; *becuman*: becom, becoman, become, becomen, becomon, becomun, becum, becumað, becuman, becume, becumen, becumende, becumene, becumenne, becumon, becwom, becwome, becwomon, becymð, becyme, becymeð, becymen, becymest, becymeþ, becymst, becymþ, beocuman, bicom, bicoman, bicomme, bicomen, bicomon, bicumad, bicumen, bicwom, bicymæð, bicymeð, bicymo, bycuman, bycumð; *becwelan*: becwæl; *behelan*: beholen, beholene, biholen; *beniman*: beneoman, benimað, beniman, benimð, benime, benimeð, benimen, benom, benumen, benumene, benumenne, binom, binoman, binumen; *beran(ge)*: bær, bærað, bære, bæren, bæron, ber, berað, beræð, berap, berð, bere, beren, berende, berene, berenne, beryð, birð, bireð, bireþ, boren, borene, borene, borene, gebær, gebæran, gebære, gebæron, geber, geberan, gebere, geberen, gebereþ, gibær; *igberan*: ibære, iboren; *igcuman*: icumen; *igdelan*: idæl, idælen, iddel, idel, idelan, idele; *igfelan*: ifel, ifele, ifeleð; *ighelan*: ihælen; *ignumman*: inumen; *bescieran*: bescear, bescearen, besceoren, besciran, bescire, bescoren, bescorene; *bestelan*: bestæl, bestæle, bestælon, bestelan, bestele, bestolen, bestolene, bistið; *brecan(ge)*: bræc, bræcan, bræcc, bræce, bræcen, bræcg, bræcon, brec, brecað, brecan, breccanne, brecð, brece, breceð, brecende, brecenne, brecest, breceþ, brech, brecon, brocen, brocenan, gebræc, gebræcan, gebræcc, gebræce, gebræcg, gebrec, gebrecan, gebrecanne, gebrecap, gebrecceð, gebrece, gebrecende, gebrecendes, gebrecest, gebreco, gebrecon, gibræc, gibrece, gibrecon; *cuman(ge)*: com, coma, coman, come, comen, comm, comon, comth, comun, cum, cumad, cumað, Cumæð, cumæþ, cuman, cumap, cumen, cummað, cumon, cwom, cwoman, cwome, cwomon, cwomon, Cym, cymas, cymd, cymð, cyme, cymed, cymeð, cymen, cymes, cymest, cymet, cymeþ, cymid, cymmeð, cymmest, cymo, cymst, cymyð, cymyst, cymþ, gecom, gecome, gecomon, gecum, gecuman, gecume, gecwome, gecymð, gecyme, gecymeð, gecymes, gecymmest, gicom, gicome, gicomon, gicyme, gicymes; *cwelan*: cwelað, cwelan, cwele, cwelende, cwellað, cwelleð; *ðurhbrecan*: Þurhbræc; *ðwæran(ge)*: Ðwære, Ðweran, geðwæra, geðwæran, geðwære, geðwærest, geðwære, geþwær, geþwæra, geþwæran, geþwære, geþwærede, geþwærende, geþwærian, geþwærode, geþwer, geþwere, geþworen, Þwere; *ðwerian*: geðwærian, geðwærie, geþwæriað, geþwæriende, geþwæriendan, Þwerian; *efencuman*: efencomon, efencuman, Efencwomon, efncomon, efncomon; *eftcuman*:

totoren, totorene, totorene; **underberan**: underberende, vnderbær; **undercuman**: vndercyme, vndercymende; **underniman**: undernim, undernimað, underniman, undernimð, undernumen; **unfelan**: Unfæle, unfelende, ungefelan, ungefele; **unhelan**: unhæl, unhæle, unhele; **unscieran**: unscoren; **ūpābreca**: uppabrece; **ūpcuman**: upcumen, upcumende, upcumene, upcymð, upcyme, upcymeð, Upcymeþ, Uppcumyn, uppcymyð; **ūpniman**: upnimende; **ūtābreca**: utbærst; **ūtniman**: utniman, utnimð; **wiðerbrecan**: wiðerbrecan, wiðerbrecap, wiðyrbrecan, wiþerbrecan; **ymbberan**: ymbboren; **ymbcuman**: ymcyme.

DATABASES AND CORPORA

- 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.
- Martín Arista, J. (ed.), García Fernández, L., Lacalle Palacios, M., Ojanguren López, A. E. and Ruiz Narbona, E. 2016. *NerthusV3. Online Lexical Database of Old English*. Nerthus Project. Universidad de La Rioja.
- Rissanen, M., Kytö, M., Kahlas-Tarkka, L., Kilpiö, M., Nevanlinna, S., Taavitsainen, I., Nevalainen, T. and Raumolin-Brunberg, H. 1991. *The Helsinki Corpus of English Texts*. Department of Modern Languages, University of Helsinki.
- Pintzuk, S. and Plug, L. 2001. *The York-Helsinki Parsed Corpus of Old English Poetry*. Department of Language and Linguistic Science. University of York.
- Taylor, A., Warner, A., Pintzuk, S. and Beths, F. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Department of Language and Linguistic Science. University of York.

REFERENCES

- Bosworth, J. and Toller, T. N. 1973 (1898). *Anglo-Saxon Dictionary*. Oxford: Oxford University Press.
- Brinton, L. and Closs Traugott, E. (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511615962>
- Campbell, A. 1987 (1959). *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. (1975). "Old English Pure Prefixes: Structure and Function". *Linguistics*, 13/145: 47-82. <https://doi.org/10.1515/ling.1975.13.145.47>
- García García, L. (2012). "Morphological causatives in Old English: the quest for a vanishing formation". *Transactions of the Philological Society*, 110/1: 122-148. <https://doi.org/10.1111/j.1467-968X.2012.01287.x>
- García García, L. (2013). "Lexicalization and morphological simplification in Old English jan-causatives: some open questions". *Sprachwissenschaft*, 38/2: 245-264.
- 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.
- Hogg, R. M. and Falk, R. D. 1992 (2011). *A Grammar of Old English*. Oxford: Wiley-Blackwell.
- Hogg, R., (ed.) (1992). *The Cambridge History of the English Language I. The Beginnings to 1066*. Cambridge: Cambridge University Press.
- Krygier, M. (1994). *The Disintegration of the English Strong Verb System*. Frankfurt am Main: Peter Lang.
- Levin, S. R. (1964). "A Reclassification of the English Strong Verbs". *Language*, 40/2: 156-161. <https://doi.org/10.2307/411574>
- Mailhammer, R. (2006). "On the origin of the Germanic strong verb system". *Sprachwissenschaft*, 31/1: 1-52.
- Mailhammer, R. (2007). "The Germanic Strong Verbs: Foundations and Development of a New System", *Trends in Linguistics. Studies and Monographs*, 183. Mouton De Gruyter. <https://doi.org/10.1515/9783110198782>
- Mailhammer, R. (2008). The Typological Significance of Ablaut in the (Pre-)history English. In W. Rudolf, T. Honegger and A. J. Johnston (eds.), *Clerks, Wives, and Historians: Essays on Medieval English Language and Literature*. Bern: Peter Lang, 185-212.
- 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. <https://doi.org/10.1080/07268602.2012.744264>
- Martín Arista, J. (2013). "Recursivity, derivational depth and the search for Old English lexical primes". *Studia Neophilologica*, 85/1: 1-21. <https://doi.org/10.1080/00393274.2013.771829>
- Martín Arista, J. (2014). "Noun layers in Old English. Asymmetry and mismatches in lexical derivation". *Nordic Journal of English Studies*, 13/3: 160-187.
- Martín Arista, J. "El paradigma derivativo del inglés antiguo". *Onomazein*. Forthcoming-a.
- Martín Arista, J. "The Semantic Poles of Old English. Towards the 3D Representation of Complex Polysemy". *Digital Scholarship in the Humanities*. Forthcoming-b.
- Martín Arista, J. and Cortés Rodríguez, F. (2014). From directionals to telics: meaning construction, word-formation and grammaticalization in Role and Reference Grammar. In M. A. Gómez González, F. Ruiz de Mendoza Ibáñez and F. González García (eds.), *Theory and Practice in Functional-Cognitive Space*. Amsterdam: John Benjamins. 229-250. <https://doi.org/10.1075/sfsl.68.10mar>
- Mateo Mendaza, R. (2013). "The Old English exponent for the semantic prime TOUCH. Descriptive and methodological questions". *Australian Journal of Linguistics*, 33/4: 449-466. <https://doi.org/10.1080/07268602.2013.857574>
- 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. <https://doi.org/10.1075/nowele.67.1.04men>
- Mateo Mendaza, R. (2015a). "Matching productivity indexes and diachronic evolution. The Old English affixes *ful-*, *-isc*, *-cund* and *-ful*". *Canadian Journal of Linguistics*. 60/1: 1-24. <https://doi.org/10.1017/S0008413100000517>
- Mateo Mendaza, R. (2015b). "The search for Old English semantic primes: the case of HAPPEN". *Nordic Journal of English Studies*, 15: 71-99.
- Mateo Mendaza, R. (2016). "The Old English exponent for the semantic prime MOVE". *Australian Journal of Linguistics*, 36/4: 542-559. <https://doi.org/10.1080/07268602.2016.1169976>
- Novo Urraca, C. (2015). "Old English Deadjectival Paradigms. Productivity and Recursivity". *NOWELE-North-Western European Language Evolution*, 68/1: 61-80. <https://doi.org/10.1075/nowele.68.1.02urr>
- Novo Urraca, C. (2016a). "Old English suffixation. Content and transposition". *English Studies*, 97/6: 638-655. <https://doi.org/10.1080/0013838X.2016.1183955>
- Novo Urraca, C. (2016b). "Morphological relatedness and the typology of adjectival formation in Old English". *Studia Neophilologica*, 88/1: 46-55. <https://doi.org/10.1080/00393274.2016.1150788>
- Prokosch, E. (1939). *A Comparative Germanic Grammar*. Philadelphia: University of Pennsylvania.
- Sweet, H. 1976 (1896). *The Student's Dictionary of Anglo-Saxon*. Cambridge: Cambridge University Press.
- Van Coetsem, F. (1990). *Ablaut and Reduplication in the Germanic Verb*. Heidelberg: Winter.
- Vea Escarza, R. (2012). "Structural and Functional Aspects of Morphological Recursivity". *NOWELE-North-Western European Language Evolution*, 64/65: 155-179. <https://doi.org/10.1075/nowele.64-65.09esc>
- Vea Escarza, R. (2013). "Old English Adjectival Affixation. Structure and Function". *Studia Anglica Posnaniensia*, Vol. 48/2-3: 1-21. <https://doi.org/10.2478/stap-2013-0005>
- Vea Escarza, R. (2014). "Split and unified functions in the formation of Old English nouns and adjectives". *Revista de Lingüística y Lenguas Aplicadas*, 9: 110-116. <https://doi.org/10.4995/rlyla.2014.2086>
- Vea Escarza, R. (2016). "Recursivity and inheritance in the formation of Old English nouns and adjectives". *Studia Neophilologica* 88/1: 1-23. <https://doi.org/10.1080/00393274.2015.1049830>
- Vea Escarza, R. (2016). "Old English affixation. A structural-functional analysis". *Nordic Journal of English Studies*. 15(1): 101-119.