Hapax Legomena and the Productivity of the Old English Weak Verb Suffixes[1]

Roberto Torre Alonso, University of La Rioja and Gema Maíz Villalta, The National Distance Education University (UNED)

Abstract

The aim of this article is to measure the productivity of the Old English weak verb suffixes *-ettan*, *-læcan*, *-sian*, *-nian*, *-lian*, *-erian* and *-cian* from a synchronic point of view by taking into account the role played by hapax legomena. Productivity in the narrow sense *P* and global productivity *P** are measured and frequency is calculated in terms of type and token. Three types of hapax legomena are distinguished, namely absolute hapaxes (unique formations that appear in one text), relative hapaxes (formations that appear in different texts, but only once in each text) and mixed hapaxes (a subsumption of both types). This typology of hapaxes puts the focus on *-sian*, *-erian*, *-lian* and *-cian*, which range between very low and zero productivity.

Keywords**:** Old English, weak verb, suffix, productivity, frequency, hapax legomenon.

1. Introduction

This article engages in Old English word-formation. More precisely, it focuses on the weak verb suffixes *-ettan*, *-læcan*, *-sian*, *-nian*, *-lian*, *-erian* and *-cian* and aims at assessing their morphological productivity from a synchronic point of view. To fulfill this goal, a combined study of dictionary and corpus data is conducted that produces descriptive and methodological conclusions related to the relative and absolute indexes of productivity of the affixes as well as the assessment of productivity in a historical language.

The topics of Old English word-formation and lexical semantics have drawn a remarkable amount of attention recently. Kastovsky (1986, 1989, 1990, 1992, 2005, 2006) describes the main units and processes involved in Old English word-formation and identifies a typological change in the lexicon from stem-based morphology to word-based morphology. In a more theoretical approach, Martín Arista (2008, 2009, 2011a) lays the foundations of a theory of derivational morphology compatible with functional models and applies it to Old English in

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

Torre Alonso, Roberto and Gema Maíz Villalta. 2014. "Hapax Legomena and the Productivity of the Old English Weak Verb Suffixes." *Nordic Journal of English Studies* 13(3):188-211.

general and, more specifically, to zero derivation (Martín Arista 2011c), lexical layers (Martín Arista 2011b), morphological recursivity (Martín Arista 2010a, 2010b, 2013a) and derivational paradigms (Martín Arista 2010c, 2012, 2013b). On the specific question of morphological productivity, Author 2 (2011, 2012) and Mateo Mendaza (2012, 2014) concur on the necessity of combining textual and lexicographical sources as well as on the importance of checking statistical analyses against qualitative judgements on the grammaticalisation or loss of the affixes at stake. Apart from these works, the question of the quantitative assessment of the morphological productivity of Old English word-formation has been addressed by Trips (2009) and Haselow (2010), who do not discuss the lexical category of the verb and, moreover, opt for a more absolute approach than the one adopted in this research.

Considering this background, the article is organised as follows. Section 2 raises the question of how to measure productivity in a historical language, with special emphasis on low indexes. Section 3 establishes the methodological and theoretical underpinnings of this work, including the relevant sources, the treatment of the data and the formulae of productivity. Then, section 4 presents the results of the frequency and productivity analysis and discusses the different accounts of productivity obtained on the basis of a typology of hapaxes that distinguishes absolute, relative and mixed hapax legomena. Finally, section 5 offers a summary and the main methodological and descriptive conclusions.

2. Measuring productivity in a historical language

The topic of productivity has recently been debated within the area of morphological theory. This is probably the case because productivity constitutes, in several respects, the meeting point of lexicology, inflectional morphology and derivational morphology. Indeed, the studies on productivity by Bauer (2004, 2005), Baayen (1992, 1993, 1994, 2009), Plag (1999, 2003) and Rainer (2005), among others, highlight the complex character of this phenomenon, which comprises, at least, the following aspects: (1) the relationship between the likelihood that a certain morphological process becomes operational and the establishment and spread of neologisms; (2) the limits between derivation and inflection as reflected by the use of two types of sources,

namely lexicographical (dictionaries containing lemmatised forms) and textual (corpora displaying unlemmatised forms); (3) the relative and absolute measure of the productivity of a certain process; and (4) the impact of the number of hapax legomena or unique formations on the overall index of productivity.

While the assessment of productivity in natural languages constitutes a challenging exercise, historical languages face further problems related to the fragmentary nature of the written records, which often results in a lack of sufficient representative data or reliable linguistic material. Moreover, as linguists like Kastovsky (1992) and Lass (1994) have put forward, an account of the productivity of a historical language poses three main problems. First, there is no direct way of testing productivity; we have to rely on indirect evidence such as the number of occurrences in a text or the continuity of a given process of word-formation. Second, productivity and transparency can vary diachronically. In Kastovsky's (1992: 357) words, "when one has to deal with a linguistic period such as Old English, stretching over some 600 years, there are bound to have been many changes. Only the output of the patterns recorded in the later documents is available for study." This goes in the line of Lass (1994: 193), who remarks that it is difficult to determine whether a given occurrence of a derived form represents an institutionalised lexical item or not, or whether it is a new formation. And third, when a given word-formation process loses its productivity, it may leave at least some of its output as part of the vocabulary.

In recent times, new approaches have tried to overcome the first problem recognized by Kastovsky. Säily and Suomela have developed a series of software programs that allow the direct observation of types and hapaxes. By using Suomela's (2007) *Types 1* tool, random permutations of the elements under study can be done, thus identifying the upper and lower bounds for each significance level. This software, and its second version (*Types 2*) have been successfully applied in studies by Säily and Suomela (2009), Säily (2011) and Gardner (fc.)

In the framework of a historical language, the morphological productivity of a word-formation process has to make reference to the number of attested types and tokens produced by the process in question. This calls for a reflection on the relative character of the concepts of type and token, which, moreover, refers us to the relationship between inflectional and derivational morphology. It has to be determined, in

other words, if lemmatised types can be distinguished from unlemmatised types, in such a way that a given lemmatised type in the dictionary corresponds to a number of unlemmatised types (forms from the inflectional paradigm of the type) each of which has a number of occurrences (tokens) in texts.

As part of the effort to enlarge and improve the methodology for the assessment of productivity in a historical language, it is also necessary to pay special attention to low productivity indexes, that some of the affixes under analysis happen to display, and to reconsider the role played by hapax legomena in the calculation of the indexes. Beginning with the former question, Fernández-Domínguez *et al*. (2007) hold that there is a tendency in word-formation research to focus on high productivity measures to the exclusion of low indexes of productivity. Fernández-Domínguez *et al.* (2007: 35) distinguish three major models of productivity assessment: (1) frequency models, which centre on type frequency, token frequency and relative frequency; (2) probabilistic models, which concentrate on productivity in the narrow sense, global productivity and the degree of productivity based on the count of hapax legomena; and (3) Štekauer's (2005) onomasiological model, in which word-formation patterns are regular, predictable and productive. Fernández-Domínguez *et al.* (2007) identify a number of shortcomings of these models of productivity assessment. In the first place, frequency models are restricted to affixation measurement, which excludes other morphogical processes of lexical creation. Secondly, probabilistic models rely almost exclusively on figures to interpret the data. Finally, the onomasiological model is hardly compatible with the others because it considers non-quantitative aspects of word-formation processes, such as the need of the speech community for a given neologism. Whatever model of productivity assessment is implemented, Fernández-Domínguez *et al*. (2007: 51) note that "high figures unequivocally correspond to high productivity, it is not entirely clear whether low figures correspondingly match low productivity or whether they imply a decrease in measurement accuracy." It is necessary, therefore, to take additional perspectives on processes with low productivity indexes. This has to be done with respect to the question of the role played by hapax legomena in the assessment of productivity. It is generally accepted (thus Bauer 2004, for instance) that the higher the figure of tokens the lower the index of productivity and, conversely, the lower the figure of tokens,

the higher the index of productivity. Consequently, processes with one instance represent the maximal degree of productivity. This is reflected by the formula that calculates productivity (Baayen 1992, 1993), in which the index of productivity results from dividing the number of hapaxes by the number of tokens. The figure of hapaxes, therefore, is in direct proportion to the index of productivity.

Productivity, as put forward by Bauer (2004: 87), is a matter of availability and profitability. Availability makes reference to whether a given process can be used for producing new words, whereas profitability refers to the frequency of use of a morphological process. For assessing the productivity of the processes of a historical language, the assessment of productivity cannot be restricted to availability. On the contrary, the question of profitability, or how much a process is used, has to be central to the analysis. In this respect, carrying out several accounts of productivity relative to three types of hapax legomena offers a more faithful description of the situation of a certain morphological process and is more compatible with qualitative judgements on loss of semantic analysability, lexicalisation and loss.

Given these considerations, the main point made by this article is that if hapaxes are considered in a relative way, based on a type-token continuum, and a distinction can be drawn between absolute hapaxes and relative hapaxes (as well as mixed hapaxes, which subsume both types), the index of productivity can be calculated in a more accurate way and, moreover, the qualitative aspects of the analysis are reinfored.

3. Sources, data and formulae

To recapitulate, this research aims at gauging the productivity of the seven suffixes that form weak verbs in Old English (*-ettan*, *-læcan*, *-sian*, *-nian*, *-lian*, *-erian* and *-cian*) from a synchronic point of view, as well discussing the role that the different types of unique formations play in the measure of productivity.

The sources of this study are both textual and lexicographical. For token analysis, the source is *The Dictionary of Old English Corpus*, which comprises around 3 million words corresponding to approximately 3,000 texts, divided into prose, poetry and glosses. For the calculation of types, the source is the lexical database of Old English *Nerthus* (www.nerthusproject.com, consulted on May 2011), which contains

approximately 30,000 headwords based on the information provided by standard dictionaries of Old English like Bosworth-Toller (1973), Sweet (1976) and Clark Hall (1996).

Turning to the data, a preliminary question arises regarding the inventory of the suffixes that form weak verbs in Old English, given that, as Figure 1 shows, there is no consensus in the literature.

Suffix	Jember *et al*. (1975)	Kastovsky (1992)	Lass (1994)	Quirk and Wrenn (1994)
-ettan	X	X	X	X
-læcan	X	X		X
-sian	X	X	X	X
-nian	X	X	X	
-lian				
-erian	X			
-cian				

Figure 1 The Old English weak verb suffixes in the literature.

Whereas there is agreement on the suffixes *-ettan* and *-sian*, whose bound and derivational character is unanimously acknowledged, the other affixes analysed in this article are not listed in all secondary sources. For this reason, in the remainder of this section we will argue and provide evidence in favour of the derivational status of *-læcan, -nian, -lian, -erian and -cian*.

The suffix *-ettan*, as in *lāþettan* ‘to loathe’ or *ōnorrettan* ‘to perform with effort’, is considered a derivational morpheme without exception (Marckwardt 1942: 275; Kastovsky 1992: 391; Quirk and Wrenn 1994: 116-118; Lass 1994: 203). The case with *-læcan* is different. For Kastovsky (1992: 391), the verbal suffix *-læc(an)* forms deadjectival verbs with the meaning ‘be, become, make’ (*dyrstlæcan* ‘to dare’, *geanlæcan* ‘to make one, join’, *rihtlæcan* ‘to put right’) and denominal verbs with the meaning ‘produce, grow, become’ (*æfenlæcan* ‘to become evening’, *loflæcan* ‘to promise’, *sumorlæcan* ‘to become summer’). Quirk and Wrenn (1994) describe the suffix *-læcan* as being used to form verbs, usually from adjectives and nouns, as in *gēanlæcan* ‘to unite’ or *nēalæcan* ‘to approach’. Along with these formations, the affix *-læcan* coexists with the weak class 1 verb *lǣcan* ‘to spring up, rise, flare up’, a zero derivative of the class VIIa strong verb *lācan* ‘to move up and down, leap, jump, swing, fly; play (instrument); play upon, delude; fight,

contend'. Such a coexistence of a lexeme and a morpheme in the lexicon indicates that grammaticalisation is underway.[2] The data available indicate that, from the semantic point of view, the derivation of *lǣcan* from *lācan* is motivated by meaning specialisation. Indeed, a change has taken place from a literal meaning of movement 'to spring' to a figurative meaning 'beginning a state or an action'. There is also a change from a more specific lexical meaning to a more general grammatical meaning, to code inchoative internal aspect (beginning of an action or state) and causative internal aspect (with a secondary predication as second argument). When the affix is attached to nouns, its function is usually to convey an inchoative internal aspect as in the example *fālǣcan* 'to be hostile to' <*fāh* 'enemy', except in the causative formations such as *gewundorlǣcan* 'to make wonderful' <*wundor* 'wonder' and *gehīwlǣcan* 'to form' <*hīw* 'form'. When the affix is attached to adjectives, it conveys a causative meaning, as in *(ge)cūðlǣcan* 'to make known' <*(ge)cūð* 'known'. It can be concluded then that *-læcan* results from grammaticalisation lexeme > derivational morpheme that can be identified on the grounds of a change from specific to general meaning and from literal to figurative meaning.[3]

As for the suffix *-ian*, a decision has to be made regarding the question whether it is simply an inflectional ending or it takes part of larger morphemes with derivational function, namely, *-sian*, *-nian*, *-lian*, *-erian* and *-cian*. As shown in figure 2, *-sian* is considered a derivational suffix in the literature. However, Kastovsky (1992: 392), who lists *-ettan*, *-læcan*, *-sian* and *-nian*, remarks that "the suffix-*n(ian)* results from misanalysis of zero-derived verbs such as *fægenian* 'rejoice'< *fægen*, *openian* 'open'< *open, tacenian* 'make a sign'< *tacen*, and leads to a few analogical formations, such as *berhtnian* 'glorify', *læcnian* 'heal', *þreatnian* 'threaten'". Lass (1994: 203) also distinguishes between the suffix *-s-ian* and a formative *-n-*. The suffix *-s-ian* appertains to class II

[2] Grammaticalisation is a change from lexical into grammatical status (Hopper and Traugott 2003: 18). According to Lehmann (2002: 15) *grammaticalisation reduces the autonomy of a unit, shifting it to a lower, more strictly regulated grammatical level*. Givón (2009: 301) lays the emphasis on the *desemanticisation of lexical forms, which gain more abstract meanings.*

[3] At the same time, some derivatives displaying this suffix undergo lexicalisation, throughout which the meaning of the derivative is not predictable from the meanings of the base of derivation and the affix.

weak verbs that have an /-s-/ formative, as *clǣn-s-ian* 'cleanse' <*clǣne* 'clean', *rīc-s-ian* 'rule' <*rīce* 'kingdom', *milt-s-ian* 'take pity on' <*mild* 'mild'. Lass (1994: 203) describes the /-n-/ formative as "reflecting an extended suffix*/-*in-ōn*/ [that-RTA & GMV] appears in a number of class II weak verbs, especially denominal and deadjectival*: fæst-n-ian* 'fasten'< *fæst, for-set-n-ian* 'beset' (*for-settan* 'hedge in, obstruct'), *lāc-n-ian* 'heal, cure' (*lǣce* 'physician')". Hallander (1966) points out that it is difficult to distinguish between those verbs that have an *-s-* as a derivational suffix and those that have an *-s-* merely ending the root. In these cases, the etymon can give us the clue. Hallander (1966: 9) explains that "the etymon of an Old English s-verb should exist in Old English in a form without -s-" but he goes on to say that "in certain cases, the probable etymon has double forms in OE: *egesian* (*ege>egesa*), *halsian* (*hǣl>hālor>hāls*) are examples of this". The solution put forward by Hallander (1966) and adopted in this research is that those verbs whose etymon includes the *-s-* are excluded from the data. Moreover, the arguments in favour of the derivational character of *-sian* are also applicable to the other suffixes in this group. It is the case that the base of derivation of the Old English weak verb suffixes does not always present a thematic consonant, that is *-s-*, *-n-*, *-l-*, *-r-* and *-c-*, as can be seen in (1), respectively:

(1)
(ge)bēnsian "to pray, supplicate" <*(ge)bēn* "prayer" (noun, m.)
(ge)dihtnian "to arrange, dispose" <*diht* "arrangement" (noun, n.)
nestlian "to make a nest" <*nest* "nest; young bird, brood" (noun, n.)
swīðrian "to avail, become strong" <*swīð* "strong, mighty" (adj.)
tamcian "to tame, soothe" <*tam* "tame; tracktable, gentle" (adj.)

For the reasons just given, the set of suffixes that form weak verbs in Old English include not only the more generally studied *-ettan*, *-læcan*, *-sian* and *-nian* but also *-lian*, *-erian* and *-cian*.

Regarding the formulae of productivity, we aim, first of all, at calculating the type-frequency and the token-frequency of the suffixes under analysis, for which we resort to Bauer's (2004, 2005) distinction. According to Bauer (2004: 102-104), type-frequency refers to the number of items of a particular word-formation process found in the dictionary, while token-frequency expresses the number of occurrences

of a given word-formation process in a certain corpus or a text. A high type-frequency corresponds to productive processes, while a high token-frequency indicates less productive processes.

As for the index of productivity, there is a distinction between productivity in the narrow sense and global productivity. Beginning with productivity in the narrow sense, Baayen (1989) develops a statistical measure of productivity based on Aronoff's (1976) formula for measuring the degree of productivity of a certain word-formation process. Baayen's formula can draw a distinction between productive and non-productive processes by establishing a degree of productivity among them. The formula is given in Figure 2:

$P = n_1 / N$

Figure 2: Index of productivity (Baayen and Lieber 1991: 809).

The productivity *P* of a given word-formation process is defined as the quotient of the number of hapax legomena n_1 or unique formations and the total number of tokens *N* of all the words resulting from a given word-formation process. In Baayen and Lieber's words (1991: 809) productivity "P expresses the rate at which new types are to be expected to appear when N tokens have been sampled, n_1 expresses the number of types of a given affix which only occurs once on the corpus (the so-called *hapax legomena*), and the total number of tokens of that given affix in a given corpus". Baayen and Lieber (1991: 124) also propose the measure of global productivity *P** of a word-formation process, which expresses the relation between the index of productivity *P* and the number of types *V*. This measure is innovative for two reasons. In the first place, global productivity is not a numerical calculation but a visual representation in which *P* appears on the horizontal axis and *V* on the vertical one. Secondly, this measure requires textual analysis, given that it relates narrow productivity to the number of types in a corpus. This means that whereas the index of narrow productivity *P* tends to calculate already productive processes, the index of global productivity *P** deals with both productive and unproductive processes.

Baayen and Lieber (1991: 124) propose global productivity in order to reduce the relative weight of hapaxes in productivity measures. Indeed, if the number of hapaxes is high, then the index of narrow productivity will increase, whereas if the number of types is high, it will

decrease. In this way, the number of hapax legomena is directly proportional to the index of productivity. For this and other reasons, there is a certain degree of controversy among scholars on the importance of unique formations. Lass (1994: 193), for instance, remarks that it is not clear if the existence of a hapax legomenon represents a piece of solid linguistic evidence or is simply a question of language survival. Plag (2006: 542), on the other hand, holds that "the number of hapaxes of a given morphological category should correlate with the number of neologisms of that category, and that the number of hapaxes can be seen as an indicator of productivity." Plag (2006: 544) also states that, when measuring productivity by counting the neologisms in a given period, "the greater the number of neologisms in that period, the higher the productivity of a given affix in that period."

Although the role played by hapaxes in the assessment of productivity is not uncontroversial, it is true that hapaxes are taken into account by most authors. In a recent study in the diachrony of the English affixes *-hood, -dom* and *-ship*, Trips (2009) "puts forward the following criterion of productivity: a productive set of formations is defined as the occurrence of formations with a morphological category with *at least* [emphasis in original] two hapaxes where a hapax is a new type built by a new rule and a new type exploiting that new rule." The criterion of productivity, as stipulated by Trips, has the advantage of marking a cut-off point, under which no morphological productivity can be found. On the other hand, Trips's criterion does not allow for gradation, since it divides processes into unproductive and productive thus leaving aside the question of the different degrees of productivity.

To summarise, Plag's (2006) position on the importance of hapaxes for determining productivity constitutes one of the main guidelines of this research, although global productivity and frequency are also assessed in order to check the results of measures directly based on hapaxes. Moreover, the criterion of productivity as defined by Trips (2009) adds an extra perspective and contributes to a more accurate measure on the productivity of morphological processes of word-formation. This is the reason why several measures have been taken into account, in such a way that the following formulae are required. To calculate type-frequency we make use of the formula shown in Figure 3:

$$\text{Type-frequency} = \frac{\text{Number of derivatives of suffix}}{\text{Number of headwords in the dictionary}}$$

Figure 3: Type-frequency (based on Bauer 2001, 2005).

For the calculation of token-frequency, the formula presented in Figure 4 is used:

$$\text{Token-frequency} = \frac{\text{Number of tokens of derivatives of suffix}}{\text{Number of words in corpus}}$$

Figure 4: Token-frequency (based on Bauer 2001, 2005).

Finally, to calculate the productivity of a given affix, we have used the formula displayed in Figure 5:

$$\text{Index of productivity } P = \frac{\text{Number of hapax legomena of suffix}}{\text{Number of tokens of derivatives of suffix}}$$

Figure 5: Index of productivity (based on Baayen and Lieber 1991).

The different measures of productivity as well as the discussion of the types of hapaxes follow in the next section.

4. Results of the analysis.

4.1. Frequency and productivity.

Overall, there are 186 weak verbs derived by means of the suffixes displayed in Figure 2. The inflected forms of these verbs are considered (relative) types when repetitions of the same form are discarded, and tokens when all repetitions of a given inflectional form are taken into account. With these definitions of type and token, there are 1,498 types and 6,737 tokens in the texts from *The Dictionary of Old English Corpus*. The figures of the synchronic analysis of the suffixes are presented in Table 1:

Affix	Verbs	Types	Tokens	Type-frequency	Token-frequency	Hapaxes	Index of productivity
-ettan	74	252	390	0.00245	0.0001294	27	0.069230
-læcan	32	295	1,114	0.00106	0.0003698	9	0.008078
-sian	30	557	3,386	0.00099	0.0011242	6	0.001772
-nian	22	289	1,694	0.00073	0.0005624	4	0.002361
-lian	17	59	81	0.00056	0.0000268	5	0.061728
-erian	8	29	39	0.00026	0.0000129	1	0.025641
-cian	3	17	33	0.00010	0.0000109	1	0.030303
Total	**186**	**1,498**	**6,737**	**0.00615**	**0.0022368**	**53**	**0.007867**

Table 1: Frequency and productivity indexes of the Old English weak verb suffixes.

The resulting indexes can be displayed in a number of hierarchies. By type-frequency, that is, the ratio of derivatives in the lexicographical source to the total number of types, the following hierarchy of suffixes can be established (> means *is more type-frequent than*):

-ettan > *-læcan* > *-sian* > *-nian* > *-lian* > *-erian* > *-cian*
Figure 6: The hierarchy of type-frequency.

By token-frequency, that is, the ratio of occurrences of the derivatives in the textual source against the total number of tokens in the textual source, the hierarchy in Figure 7 can be established (> means *is more token-frequent than*):

-sian > *-nian* > *-læcan* > *-ettan* > *-lian* > *-erian* > *-cian*
Figure 7: The hierarchy of token-frequency.

The index of productivity has been calculated by dividing the number of hapaxes by the number of tokens. The hierarchy of productivity is as follows in Figure 8 (> means *is more productive in the narrow sense than*):

-ettan > *-lian* > *-cian* > *-erian* > *-læcan* > *-nian* > *-sian*
Figure 8: The hierarchy of the index of productivity.

By applying Baayen and Lieber's (1991) concept of global productivity we get the graphic presented in Figure 9:

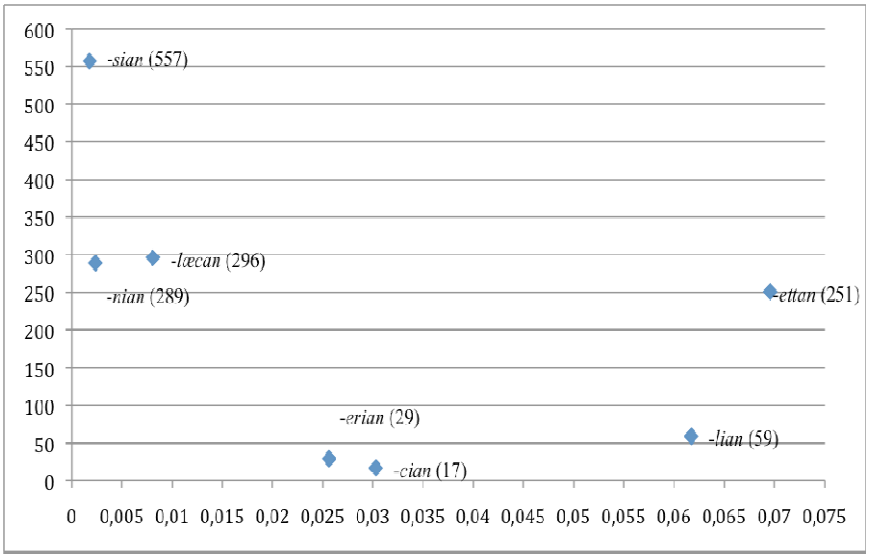

Figure 9: Global productivity of the Old English weak verb suffixes.

In Figure 9, The horizontal axis *V* refers to types while the vertical axis *P* is the index of productivity of each suffix. Notice that while *-sian* has the highest number of types *V*, its index of productivity is very low, whereas *-ettan* does not have a high number of types but enjoys the highest index of productivity. In general, the suffixes with a higher number of types have a lower index of productivity, whereas the suffixes with fewer types have higher indexes of productivity. Regarding this question, Plag (2006), in his analysis of the English suffixes *-ness*, *-ion*, *-ity*, *-ist*, *-less*, *-ish* and *-wise*, demonstrates that the suffix *-wise*, in spite of having very low rates of tokens and types, has a very high index of productivity. Plag (2006: 545) points out that "the proportion of unknown words among all the -wise derivatives is high, indicating the suffix's potential to be easily used for the coinage of new forms (...) The OED ranking reflects the fact that -wise words are, though easily derivable, not often used." In other words, when the suffix does not take place in many different words that are not very frequent in the texts, the

rates of the number of tokens and hapaxes and token-frequency are very low, but the fact that an affix would be used easily to coin new forms makes the index of *P* very high. That is, suffixes with high frequency, number of tokens and hapaxes, but with a low index of *P*, are used on stable combiantions root-suffix, but seldom to coin new words.

4.2. Absolute and relative hapax legomena.

Despite Baayen and Lieber's (1991) attempt to restrict the relative importance of hapaxes for calculating productivity, productivity indexes crucially depend on the number of hapaxes. For this reason, the concept of hapax legomena requires some further attention at this point. The analysis carried out so far is based on an absolute definition of this concept according to which one type is realized by one token in one text. Defined in relative terms, a hapax legomenon is one type that is realized by the same inflectional form in two or more texts. That is to say, an absolute hapax legomenon is a unique formation that appears in one text whereas a relative hapax legomenon remains a unique formation but it appears in more than one text. Ultimately, this definition is based on the distinction between type and token, which is not absolute but relative. Indeed, the inflected forms of a dictionary form constitute tokens of the dictionary form in question, but when they appear in more than one text, the abstract inflective form becomes a type with respect to the concrete inflective forms as they can be found in the texts. Put in other words, the absolute hapax is a purely derivational concept whereas the relative hapax has more to do with the textual realizations of the abstract morphological paradigm. Finally, mixed hapax legomena represent the sum of the absolute and relative hapaxes of a word and offer comprehensive view of the hapaxes identified.

The application of these types of hapaxes to the affixes that form weak verbs in Old English is the following. Begining with *-ettan*, this suffix presents 27 absolute hapax legomena and 6 relative hapax legomena. They appear in Figure 10:

Absolute hapax legomena (27): *beþete* (*beðettan*, PeriD), *cloccet* (*cloccettan*, LchII), *efnette* ((*ge)efnettan*, GDPref), *firmetton* (*firmettan*, Or), *fnærettende* (*fnǣrettan*, HlGl), *gealchatte* (*gealchattan*, PsGlI), *galpettað* (*gealpettan*, HomU), *grunnettan* (*grunnettan*, CorpGl), *hliapettan* (*hlēapettan*, Bede), *hospetęt* (*hospettan*, CorpGl), *huncetton* (*huncettan*, PsGlD), *leasliccettan* (*lēaslīcettan*, ClGl), *miscroccetan* (*miscrōcettan*, LS), *muþetton* (*mūðettan*, ÆHomM), *salletað* (*sallettan*, PPs), *scofett* (*scofettan*, CP), *slecgete* (*slecgettan*, Lch II), *spigette* (*spīgettan*, Num), *sporetteð* (*sporettan*, PsCaA), *swolgettan* (*swolgettan*, Lch II), *togetteþ* (*togettan*, Lch II), *towettan* (*towettan*, WPol), *geþæfetæþ* (*þafettan*, PsCaE), *þametaþ* (*þamettan*, PsGlJ), *þoddetton* (*þodettan*, HomU), *wincettað* (*wincettan*, PPs (prose)), *winhreafetiaþ* (*wīnreafetian*, PsGlI).
Relative hapax legomena (6): *cancettende* (*cancettan*, ThCap, ClGl), *onretteþ* (*onōrettan*, ClGl (2)), *plicet* (*plicettan*, Ch, PrudGl), *sarette* (*sārettan*, CP (2)), *spornette* (*spornettan*, ClGl (2))*tolcetende* (*tolcettan*, AldV (2)).
Mixed hapax legomena: 33.
Figure 10: The absolute, relative and mixed hapax legomena of *-ettan*.

The hapaxes of the suffix *-læcan* can be seen in Figure 11:

Absolute hapax legomena (9): *gesamodlæceð* (*gesamodlǣcan*, PsGlG), *gewærlæht* (*gewærlǣcan*, ÆIntSig), *geweredlæhþ* (*gewerodlǣcan*, LibSc), *gewistlæcan* (*gewistlǣcan*, Lk (WSCp)), *gewundorlæc* (*gewundorlǣcan*, PsGlF), *loflæcað* (*loflǣcan*, PsGlI), *sumorlæhð* (*sumorlǣcan*, ÆCHom I), *swæðlæhte* (*swæðlǣcan*, PsCaC), *wyþerlecað* (*wiðerlǣcan*, PsGlE).
Relative hapax legomena (1): *gelimplæcan* (*limplǣcan*, AldV (2)).
Mixed hapax legomena: 10.
Figure 11: The absolute, relative and mixed hapax legomena of *-læcan*.

The hapaxes of the suffix *-sian* appear in Figure 12:

Absolute hapax legomena (6): *dwelsode* (*dwelsian*, PsGlI), *frecelsod* (*frecēlsian*, ClGl), *gedyrsod* (*gedȳrsian*, Jud), *hrywsode* (*hrȳwsian*, PsGlC), *hwinsianne* (*hwinsian*, LS), *wrænsiaþ* (*wrǣnsian*, HomU).
Relative hapax legomena (0).
Mixed hapax legomena: 6.
Figure 12: The absolute, relative and mixed hapax legomena of *-sian*.

Figure 13 displays the hapaxes of *-nian*:

Absolute hapax legomena (4): *gefætnodest* (*gefǣtnian*, PsGlI), *gefultnede* (*gefulwihtnian*, LS), *geliffæstnast* (*gelīffæstnian*, PsGlD), *þreatniað* (*ðrēatnian*, ÆCHom I).
Relative hapax legomenon (1): *gecocanade* (*gecōcnian*, ClGl (2)).
Mixed hapax legomena: 5.
Figure 13: The absolute, relative and mixed hapax legomena of *-nian*.

The hapax legomena found for *-lian* can be seen in Figure 14:

Absolute hapax legomena (5): *bræclade* (*bræclian*, GDPref), *cneowlian* (*cnēowlian*, LS), *gefystlude* (*gefȳstlian*, LibSc), *nestliað* (*nestlian*, PsGlI), *spearnlode* (*spearnlian*, judg).
Relative hapax legomena (0).
Mixed hapax legomena: 5.
Figure 14: The absolute, relative and mixed hapax legomena of *-lian*.

The hapax legomena of *-erian* verbs as found in *The Dictionary of Old English Corpus* are given in Figure 15

Absolute hapax legomenon (1): *woperiende* (*wōperian*, LS).
Relative hapax legomenon (0).
Mixed hapax legomenon: 1.
Figure 15: The absolute, relative and mixed hapax legomena of *-erian*.

Finally, the hapax legomena of the suffix *-cian* can be seen in Figure 16:

Absolute hapax legomena (1): *tamcyan* (*tamcian*, ChrodR).
Relative hapax legomena (0).
Mixed hapax legomenon: 1.
Figure 16: The absolute, relative and mixed hapax legomena of *-cian*.

If productivity is measured on the basis of mixed hapaxes, *-ettan* is the most productive, followed, in this order, by *-lian, -cian, -erian, -læcan, -nian* and *-sian*, which qualifies as the least productive. The figures are shown in Table 2.

Affix	**tokens**	**mixed hapaxes**	**index of productivity**
-ettan	390	33	0.084615
-læcan	1,114	10	0.008976
-sian	3,386	6	0.001772
-nian	1,694	5	0.002951
-lian	81	5	0.061728
-erian	39	1	0.025641
-cian	33	1	0.030303

Table 2: Index of productivity with mixed hapaxes.

If the account of productivity is based on the figure of relative hapaxes, it turns out that the productivity of *-sian*, *-lian*, *-erian* and *-cian* is zero. As for the rest of the suffixes, *-ettan* is the most productive, followed by *-læcan* and *-nian*. This is shown in Table 3:

Affix	**Tokens**	**Relative hapaxes**	**Index of productivity**
-ettan	390	6	0.031578
-læcan	1,114	1	0.000897
-sian	3,386	0	0
-nian	1,694	1	0.000590
-lian	81	0	0
-erian	39	0	0
-cian	33	0	0

Table 3: Index of productivity with relative hapaxes.

Table 4 provides an account of the index of productivity considering the number of absolute hapaxes. Given that *-sian*, *-lian*, *-erian* and *–cian* do not have any relative hapaxes, their figures coincide with those based on mixed hapaxes.

Affix	**tokens**	**absolute hapaxes**	**index of productivity**
-ettan	390	27	0.069230
-læcan	1,114	9	0.008078
-sian	3,386	6	0.001772
-nian	1,694	4	0.002361
-lian	81	5	0.061728
-erian	39	1	0.025641
-cian	33	1	0.030303

Table 4: Index of productivity with absolute hapaxes

Although the indexes of productivity based on the different types of hapaxes vary, the relative productivity of the affixes calculated on the grounds of mixed and absolute hapaxes coincide: *-ettan* is the most productive, followed, in this order, by *-lian*, *-cian*, *-erian*, *-læcan*, *-nian* and *-sian*. However, the productivity indexes based on relative hapaxes bring about an interesting change in perspective: *-ettan* remains the most productive suffix, which is in keeping with its low number of types and tokens and high figure of hapaxes; since they have no relative hapaxes, the productivity of the suffixes *-sian*, *-lian*, *-erian* and *-cian* is zero, which is consistent with their high number of tokens (*-sian*) or low number of types (*-lian*, *-erian* and *-cian*); and the productivity of *-læcan* and *-nian* is clearly lower than that of *-ettan*, which is in agreement with its lower number of types and hapaxes. In sum, the greatest advantage of the measure on productivity based on relative hapaxes is that it considers *-læcan* and *-nian* more productive than *-lian*, *-erian* and *-cian*. In this respect, if Trip's (2009) criterion of productivity requiring a minimum of hapaxes is applied, *-lian*, *-erian* and *-cian* cannot be productive, while the productivity of *-læcan* and *-nian* is questionable. Leaving aside hapaxes, it is hardly compatible with the concept of productivity as has been discussed in section 2 that an affix with a higher type:token ratio that results from a low number of types (like *-lian*, *-erian* and *-cian*) is more productive than an affix with a lower type:token ratio due to a relatively high number of types (as is the case with *-læcan* and *-nian*). Therefore, the measure based on relative hapaxes is more accurate than

the one based on absolute hapaxes, the reason being that considering hapaxes in their relative version reduces the relative weight of hapax legomena in the statistical measure on productivity.

5. Summary and conclusions

The previous discussion has shown that the combination of a lexicographical and a textual source allows us to gauge the morphological productivity of the Old English weak verb suffixes in a principled way. Given the central role played by hapaxes in the quantitative assessment of productivity, two new types of hapax legomena have been distinguished, namely relative hapax legomena and mixed hapax legomena. An absolute hapax legomenon occurs when one type is realised by one token in one text. A relative hapax legomenon occurs when one type that is realised by the same inflectional form appears in two or more texts. In other words, the difference between an absolute and a relative hapax legomenon is that the former appears in one whereas the latter turns up in two or more texts (mixed hapaxes constitute a combination of the other two types).

If productivity is assessed on the grounds of absolute hapaxes, the suffix *-ettan* enjoys the highest index of productivity, followed by the suffix *-læcan*, while the suffixes *-lian*, *-erian* and *-cian* display the lowest indexes. In the formula used for calculating productivity, the index of productivity is in direct proportion to the number of hapax legomena (or unique formations) and in inverse proportion to the number of tokens (or textual occurrences in all texts). Thus, the comparatively high productivity of *-ettan* is a consequence of the high number of hapax legomena containing this suffix while the comparatively low productivity of *-nian* and *-sian* results from the low figures of hapaxes shown by these affixes. Additionally, affixes with a high number of tokens, such as *-sian*, *-nian* and *-læcan* are less productive than affixes with a low figure of tokens such as *-lian*, *-cian* and *-erian.*

If the account of productivity is based on the relative hapaxes, *-ettan* remains the most productive suffix, the productivity of the suffixes *-sian*, *-lian*, *-erian* and *-cian* is zero and the productivity of *-læcan* and *-nian* is lower than that of *-ettan.* The indexes of productivity calculated on the grounds of relative hapaxes have the advantage over absolute hapaxes in considering *-læcan* and *-nian* more productive than *-lian*, *-erian* and

-cian. Given that an affix with a higher type:token ratio that results from a low number of types cannot be more productive than an affix with a lower type:token ratio due to a relatively high number of types, the measure based on relative hapaxes is more accurate because it lessens the importance of hapax legomena.

Regarding the relationship between absolute and relative hapax legomena, it seems to be the case that lemmatised types in the dictionary correspond to more than one unlemmatised type in the corpus, which, in turn, are reflected by a given number of tokens in the corpus. The opposite is less frequent. That is, a correspondence between one lemmatised type in the dictionary and one unlemmatised type in the corpus reflected by two or more tokens in the corpus (as the concept of relative hapax legomenon requires) occurs very infrequently. Actually, this happens less often than absolute hapax legomena arise, which comprise one lemmatised type in the dictionary and one unlemmatised type in the corpus reflected by one token in the corpus. The conclusion can be drawn, therefore, that, absolute hapaxes are more frequent than relative hapaxes.

Finally, it can be concluded that a high token frequency has resulted in a low productivity index. This is coincidental with the view, already stated by Stanley (in Amos 1980: 141) that the generalised derivation of weak verbs with *-ian* is characteristic of the end of the Old English period. On the diachronic axis, the loss of the affixes for derivational or inflectional reasons due to the decline of inflections also points to a low level of productivity of the weak verb suffixes. Against this background, the typology of hapaxes allows more fine-grained distinctions to be drawn because it puts the focus on *-sian*, *-erian*, *-lian* and *-cian*, which range between very low and zero productivity.

References

Amos, Ashley C. 1980. *Linguistic Means of Determining the Dates of Old English Literary Texts*. Cambridge, Massachusetts: The Medieval Academy of America.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, Massachusets: MIT Press.

Baayen, Harald. 1989. *A Corpus-Based Approach to Morphological Productivity*. Doctoral dissertation. Free University of Amsterdam.

Baayen, Harald. 1992 "Quantitative aspects of morphological productivity". Yearbook of Morphology 1991 Eds Geert Booij and Jaap van Marle. Kluwer Academic Publishers, Dordrecht. 109-149.

Baayen, Harald. 1993. "On Frequency, Transparency and Productivity." Yearbook of Morphology 1992. Eds. Gert E. Booij and Jaap Van Marle Dordrecht: Kluwer Academic Publishers.181-208.

Baayen, Harald. 1994 "Productivity in production". *Language and Cognitive Processes* 9: 447-469.

Baayen, Harald. 2009 "Corpus linguistics in morphology: morphological productivity". *Corpus Linguistics*. *An international handbook*. Eds. Anke Lüdeling and Merja Kyto. Berlin: Mouton De Gruyter .900-919.

Baayen, Harald & Rochelle Lieber. 1991. "Productivity and English derivation: a corpus-based study." *Linguistics* 29: 801-843.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2004. *A Glossary of Morphology*. Edinburgh: Edinburgh University Press.

Bauer, Laurie. 2005. "Productivity: Theories." *Handbook of Word-Formation.* Eds. Pavol Štekauer and Rochelle Lieber. Dordrecht: Springer. 315-334.

Bosworth, Joseph and Thomas N. Toller. 1973. *An Anglo-Saxon Dictionary*. Oxford: Oxford University Press.

Clark Hall, John. R. 1996. *A Concise Anglo-Saxon Dictionary*. Toronto: University of Toronto Press.

Fernández-Domínguez, Jesús, Ana Díaz-Negrillo and Pavol Štekauer. 2007. "How is low morphological productivity measured?" *Atlantis* 29: 29–54

Gardner, Anne. Derivation in Middle English: Regional and Text Type Variation. Unpublished PhD.

Givón, Talmy. 2009. *The Genesis of Syntactic Complexity*. Amsterdam: John Benjamins.

Hallander, Lars-Gunnar. 1966. *Old English Verbs in -sian: A Semantic and Derivational study*. Stockholm Studies in English: University of Stockholm.

Haselow, Alexander. 2010. *Typological Changes in the Lexicon. Analytic Tendencies in English Noun Formation.* Berlin: Mouton de Gruyter.

Healey, Antonette diPaolo (ed.). 2004. *The Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Hopper, Paul and Elisabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambrige University Press.

Jember, Gregory K. *et al.* 1975. *English-Old English, Old English-English Dictionary*. Boulder, Colorado: Westview Press.

Kastovsky, Dieter. 1986. "Deverbal nouns in Old and Modern English: from stem-formation to word-formation". *Historical Semantics-Historical Word Formation.* Ed. Jacek Fisiak. Berlin: Mouton de Gruyter. 221-61.

Kastovsky, Dieter. 1989. "Typological Changes in the History of English Morphology". *Meaning and Beyond*. *Ernst Leisi zum 70. Geburstag*. Ed. Udo Fries and Martin Heusser. Tübingen: Narr. 281-93.

Kastovsky, Dieter. 1990. "The typological status of Old English Word Formation". *Papers from the 5th International Conference on English Historical Linguistics.* Eds. Sylvia M. Adamson, Vivien A. Law, Nigel Vincent and Susan Wright. Amsterdam: John Benjamins. 205-224.

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

Kastovsky, Dieter. 2005. "Conversion and/on zero: word-formation theory, historical linguistics, and typology". *Approaches to Conversion/Zero Derivation.* Eds. Laurie Bauer and Salvador Varela (eds.). Münster: Waxmann. 31-50.

Kastovsky, Dieter. 2006. "Typological Changes in Derivational Morphology". *The Handbook of The History of English*. Eds. Ans van Kemenade and Los Bettelou (eds.). Oxford: Blackwell. 151-177.

Lass, Roger. 1994. *Old English: A Historical Linguistic Companion*. Cambridge: Cambridge University Press.

Lehmann, Christian. 2002. "Myths and the prehistory of grammars". *Journal of Linguistics* 38. 113-136.

Maíz Villata, Gema. 2011. Assessing the productivity of the Old English *–læcan*". *Miscelanea. A Journal of English and American Studies* 43: 55-72

Maíz Villalta, Gema. 2011. "Low Productivity indexes: The Old English verbal suffixes *-ettan* and *–læcan*". *RESLA-Revista Española de Lingüística Aplicada* 25: 117-131.

Marckwardt, Albert H. 1942. "The Verbal Suffix *-ettan* in Old English". *Language* 18: 275-281.

Martín Arista, Javier. 2008. "Unification and separation in a functional theory of morphology". *Investigations of the Syntax-Semantics-Pragmatics Interface*. Ed. Robert D. Van Valin Jr. Amsterdam: John Benjamins. 119-145.

Martín Arista, Javier. 2009. "A Typology of Morphological Constructions". *Deconstructing Constructions.* Eds Christopher Butler and Javier Martín Arista. Amsterdam: John Benjamins .85-115

Martín Arista, Javier. 2010a. "OE strong verbs derived from strong verbs". *SKASE Journal of Theoretical Linguistics* 7/1: 36-56.

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

Martín Arista, Javier. 2010c. "Building a lexical database of Old English: issues and landmarks". *Current projects in historical lexicography.* Ed. John P. Considine. Newcastle: Cambridge Scholars Publishing. 1-33.

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

Martín Arista, Javier. 2011b. "Adjective formation and lexical layers in Old English". *English Studies* 92/3: 323-344.

Martín Arista, Javier. 2011c. "Morphological relatedness and zero alternation in Old English". *Morphosyntactic Alternations in English*. Ed. Pilar Guerrero Medina. London: Equinox. 311-336.

Martín Arista, Javier. 2012. "Lexical Database, Derivational Map and 3D Representation". RESLA-*Revista Española de Lingüística Aplicada*. (Extra 1): 119-144.

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

Martín Arista, Javier. 2013b. "Old English Lexical Primes: Corpus Analysis and Database Compilation". *Creation and Use of Historical*

Linguistic Corpora in Spain. Ed. Nila Vázquez. Newcastle: Cambridge Scholars. 335-357.

Martín Arista, Javier. *et al*.2009. *Nerthus: An Online Lexical Database of Old English.* [Available at: http://www.nerthusproject.com]

Mateo Mendaza, Raquel. 2012. "The Old English Adjectival suffixes -*cund* and -*isc*: textual occurrences and productivity". *ES-Revista de Filología Inglesa* 33: 197-213.

Mateo Mendaza, Raquel. 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.

Plag, Ingo. 1999. *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Plag, Ingo. 2006. Productivity. *Handbook of English Linguistics*. Eds. Bas Aarts, B. and April McMahon. Oxford: Blackwell. 537-556.

Quirk, Randolph and Charles L. Wrenn 1994. *An Old English Grammar*. DeKalb, IL: Northern Illinois University Press.

Rainer, Franz. 2005. "Semantic change in word formation". *Linguistics* 43/2: 415-441.

Säily, Tanja and Jukka Suomela. 2009. "Comparing type counts: The case of women, men and -ity in early English letters." In *Corpus Linguistics: Refinements and Reassessments*. Eds. Antoinette Renouf and Andrew Kehoe. Amsterdam: Rodopi. 87-109.

Säily, Tanja 2011. "Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations". *Corpus Linguistics and Linguistic Theory* 7/1: 119-41.

Štekauer, Pavol. 2005. Meaning predictability in word-formation: Novel, context-free naming units. Amsterdam: Benjamins.

Sweet, Henry. 1976. *The Student's Dictionary of Anglo-Saxon*. Cambridge: Cambridge University Press.

Trips, Carola. 2009. *Lexical Semantics and Diachronic Morphology. The Development of -hood, -dom and -ship in the History of English*. Tübingen: Niemeyer.