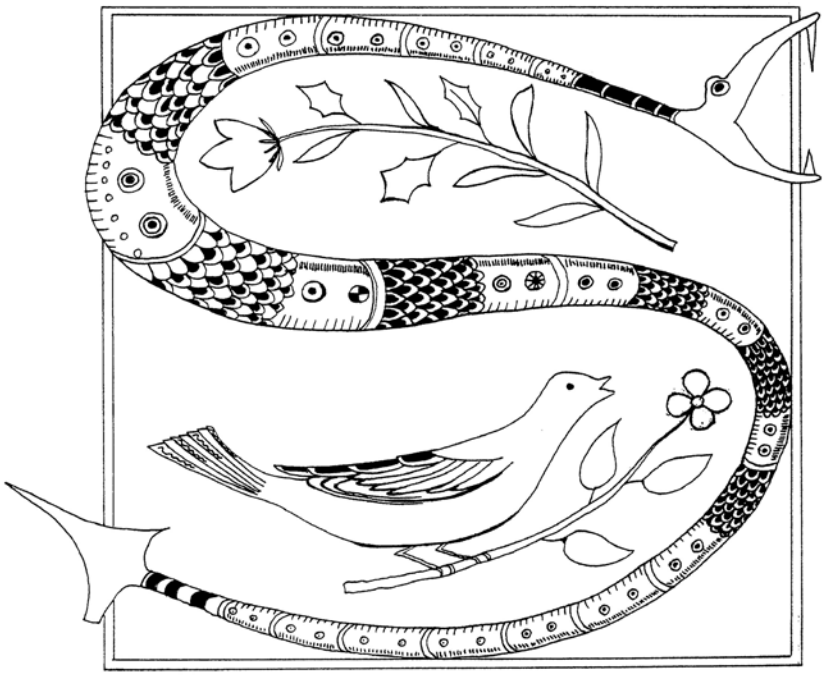
SELIM

Journal of the Spanish Society for Mediaeval English Language and Literature
Revista de la Sociedad Española de Lengua y Literatura Inglesa Medieval



Edited by - Dirigida por
Trinidad Guzmán & S. G. Fernández-Corugedo

Nº 14
Oviedo, 2007

Universidad de Oviedo &
Sociedad Española de Lengua y Literatura Inglesa Medieval
Oviedo – Murcia 2008

SEMANTIC PRIMES IN OLD ENGLISH: A PRELIMINARY STUDY OF DESCRIPTORS[1]

Abstract

The aim of this paper is to apply the methodology of semantic primes by Goddard and Wierzbicka (2002) to Old English in order to check whether it represents a suitable theoretical and methodological framework for the lexical and semantic study of this period. This constitutes a preliminary analysis of the semantic primes grouped as Descriptors: BIG/SMALL. The group is discussed taking into account a sample of texts provided by *The Helsinki Corpus of English Texts* and supplemented by the information contained in *The Dictionary of Old English Corpus*. The main sources of information on Old English definitions are *A Thesaurus of Old English* by Roberts and Kay (1995) and *A Concise Anglo-Saxon Dictionary* by Clark Hall (1931). The article attempts at being just a first approach to the topic, which could be further developed and extended to other semantic categories.

Keywords: Old English, semantic primes, linguistic corpus.

Resumen

El objetivo del presente artículo es aplicar la teoría de los primitivos semánticos de Goddard y Wierzbicka al inglés antiguo para comprobar si representa un marco teórico y metodológico adecuado para el estudio léxico-semántico del periodo. Éste constituye un análisis preliminar de los primitivos semánticos agrupados bajo la denominación de *Descriptors*: BIG/SMALL. El grupo se analiza teniendo en cuenta una muestra de textos extraídos del *Helsinki Corpus of English Texts*. El artículo intenta ser un primer acercamiento al tema, que podría seguir investigándose en el futuro, a medida que se amplíe el corpus de análisis a través de la información proporcionada por el *Dictionary of Old English Corpus*. Las principales fuentes de información para las definiciones en inglés antiguo proceden de *The Oxford English Dictionary*, A *Thesaurus of Old English* de Roberts y Kay, así como *A Concise Anglo-Saxon Dictionary* de Clark Hall.

Palabras clave: inglés antiguo, primitivos semánticos, corpus lingüísticos.

1 Introduction

Although two opposing views on linguistic facts have been alternating throughout the history of linguistics and even if some aspects are culture-dependent, as proposed by

[1] The research reported here is part of the project "Base de datos dinámica online de la morfología derivativa del inglés antiguo" ("Dynamic online database of Old English derivative morphology"), sponsored by the Spanish Ministry of Education, Reference HUM2005-07651-C02-02/FILO.

relativism, some other linguistic facts seem to obey to universally valid rules.[2] In fact, a uniform language was already reported in the Bible referring to the Tower of Babel. Therefore, the quest to find out what is invariable and shared by all languages has a long history and from different theoretical positions both philosophers and linguists have strived along centuries to discover these universal features common to all human languages. The debate was centred not only on the finding of universals but also on their own existence, inasmuch as some scholars deny the actual existence of universal properties common to all languages. However, proponents of linguistic primitives are found among several schools of thought: *realism* in the early Middle Ages, *nominalism* and *conceptualism* in the Renaissance period, *rationalism* in the 17th century as well as other movements and individuals in the next two centuries who tried to discover the relationship between the abstract universal entities and the particular languages that embody them, the best method to approach linguistic primitives and how to define them. Some of the ideas defended within the philosophical frameworks were put into practice by linguists from the 19th century onwards, although the objectives were not always alike. Thus, the Neogrammarians developed a whole network of connections between languages based on their genealogy or common origin; typological linguists focused on grouping languages according to the morphological constituents basically, while generativists centred their research on the syntactic structure of the different systems. Even if the approaches and the goals were diverse, the investigation on linguistic universals benefited from the advance in linguistics, as efforts were made to acquire a deeper knowledge of a huge amount of languages and to

[2] For a revision of the philosophical and linguistic origin and the development of language universals, as well as the different standpoints within this framework, see Mairal (2005), Mairal & Gil (2003) and Moure (2001), among others.

compare them to find out about their origin, their morphemes, their syntactic patterns or their divergences.

In the 20th century, especially in its second half, there was a renewed impetus in the search for universal linguistic properties. The debate on those general linguistic common traits was the focal point of investigation in two crucial conferences held at Dobbs Ferry (1961), "Conference on Universals of Language", organised by Greenberg and at Austin (1967), "Symposium on Universals in Linguistic Theory", by Bach and Harms. As a result of the arguments exposed in these two events, studies to test hypothetical sets of universals across a number of genetically and typologically diverse languages were undertaken. They were applied to the different planes of language analysis: phonology, morphology, syntax and lexis and semantics. Within the latter field, seminal work has been done on the part of Goddard and Wierzbicka. Thus, following the tenets of universal grammar, Wierzbicka (1996: 16) defends the idea that semantic primes or fundamental human concepts are universal and innate. Likewise, the Natural Semantic Metalanguage framework assumes that every natural language can be used as a metalanguage, which makes it unnecessary to resort to abstract semantic predicates to account for the meanings of a given language. Although there have been several previous attempts in which the set of hypothetical primitives was considerably expanded from one version to the next one, in its latest version the Natural Semantic Metalanguage includes (Goddard 2002: 14) the following:

- Substantives (I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY)
- Determiners (THIS, THE SAME, OTHER)
- Quantifiers (ONE, TWO, SOME, ALL, MUCH/MANY)
- Evaluators (GOOD, BAD)
- Descriptors (BIG, SMALL)
- Mental predicates (THINK, KNOW, WANT, FEEL, SEE, HEAR)

- Speech (SAY, WORDS, TRUE)
- Actions, events and movement (DO, HAPPEN, MOVE)
- Existence and possession (THERE IS, HAVE)
- Live and death (LIVE, DIE)
- Time (WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME)
- Space (WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE)
- Logical concepts (NOT, MAY BE, CAN, BECAUSE, IF)
- Intensifier, augmentor (VERY, MORE)
- Taxonomy, partonomy (KIND OF, PART OF)
- Similarity (LIKE)

Following the tenets rendered in Goddard and Wierzbicka (2002) and Wierzbicka (2002), this article aims at contributing to the model of universal grammar by carrying out the application of the Natural Semantic Metalanguage Research Program (henceforth NSMRP) to Old English. Martín Arista & Martín de la Rosa (2006) already applied the model to the analysis of the semantic primes of Old English that belong to the classes of substantives, determiners and quantifiers. The common characteristic to the three categories is that, being grammatical words, their combinations are far more restricted than those of lexical classes. As the latter article constitutes the first attempt to put the NSMRP into practice in Old English, the present research follows similar techniques and methods, although on this occasion I will concentrate on the analysis of descriptors BIG/SMALL, a lexical class. In choosing this specific approach my purpose is twofold: Firstly, to check whether it represents a suitable theoretical and methodological framework for the lexical and semantic study of the period. Secondly, to establish the 'hyperonym(s)' among all possible equivalents of BIG and SMALL extracted from the *Old English Thesaurus* by Roberts and Kay (1995).

2 Methodology

Before discussing the exponents of the semantic primitives it is necessary to ponder on two preliminary issues:

First of all, do all languages have units to express the concepts of BIG and SMALL? The evidence presented by different scholars, such as Goddard and Wierzbicka,[3] seems to suggest that they do. As one cannot resort to native speakers to confirm this claim in Old English, lexicographic works are one of the most valuable information sources. As Wierzbicka (1998: 152) suggests the right way to proceed is by identifying "in the language the local exponents of the universal concepts in question, with all their allomorphs and allolexes (i.e., lexical variants) and with the relevant grammatical frames". When one looks up for the equivalents of BIG in Roberts and Kay (1995: 161-162), it reads as follows:[4]

> **03.03.04.01.01 Greatness, bigness, size**: micelnes, micelu
> **. Greatness, bigness**: greatnes, gryto
> **. Fatness, bulkiness**: fætnes
> **. Something huge, a very great mass**: ormæte
> **. Large, big, broad**: brad, micel, wid
> **.. Large, capacious, ample**: sid
> **. (Of a ship) ample, with a large hold**: widfæþme

[3] Some of their more relevant contributions to the establishment of the theory and the progressive development of the different sets of universals can be consulted in the reference section. See, for instance, Goddard (1994; 2002a & 2002b), Goddard & Wierzbicka (1994; 2002) and Wierzbicka (1994 & 2002). In some of these publications, the reader will also find the application of the model to other languages.

[4] Although long vowels are marked with a macron on the paper version of the *Thesaurus*, the online version offers the possibility of searching with or without length-mark. As the *Helsinki Corpus* and the *The Dictionary of Old English Corpus* show no vowel length, this feature is omitted all through the article.

. Massive, bulky, huge: great
… (Of things) plump, full-bodied, subtantial: fætt
. Considerable: gehycglice, ungehwæde
. Not little, great: unlytel
. Very great, exceedingly great: swiþlic
. Huge, immense, enormous: eacencræftig, ormæte, ormætlic, ungefog, ungefoglic, unmæte, unmætlice, ungemet, uþmæte
. Giant, gigantic: entisc, eotenisc
. Of varying size: mismicel

The first thing that strikes the reader by having a look at the elements is the fact that the word *big* is not included, as it is not documented until the end of 13th century in writers of Northumbria and north Lincolnshire. The *Oxford English Dictionary* relates its location to a possible Norse origin. Likewise, *large* was not part of the word-stock yet, as it is first attested in the Middle English period. Apart from that, the second striking aspect is the great amount of lexical units that can be used. According to Wierzbicka, when trying to identify a proper evaluator (1989: 324) "difficulties involved in identifying these concepts cross-linguistically are due more to the superabundance of plausible candidates than to their absence". Although she is referring to evaluators GOOD/BAD, the great number of synonyms or quasi-synonyms documented in Old English for the descriptors BIG/SMALL seems to pose the same problem. It follows from here that the variation in use requires further investigation, as can be seen from the glance at the exponents of Old English semantic primitive of SMALL (Roberts and Kay, 1995: 166-167):

03.03.04.04 Littleness, smallness: gehwædnes, lytelnes

. **Little, a small amount**: hwæthwugu, hwega, hwon, hwugudæl, lyt, lytel, lythwon, medmicel, tohwega.
. **A particle, small piece, jot**: corn, egl, grot, lyttuc, mot, prica, pricele, spot, wloh
. **(Of amount) small, little**: hwæde, lytel, medeme, smæl, unginne, unmicel
.. **Little, small, slight**: hwonic, ieþelic, mæte, medemlic, medmicel
. **Very little**: forlytel, lytel
.. **Least**: læst
. **Somewhat, a little**: æthwara, æthwega, be sumum dæle, hwæthwara, hwæthwugu, hwæthwugununges, hwene, hwon, sumdæl, sume dæle, sumes, sum on dæle, tohwega
. **Slightly, little**: hwæthwugu, hwon, hwonlice, leohtlice, lyt, lytl, lythwon, lytle, lytlum, medemlice

The abundance of terms for both concepts (BIG and SMALL) makes it difficult to work out which one is the hyperonym among all the possible candidates. Two selecting criteria can be applied here: on the one hand, the frequency of occurrences; on the other, the range of different collocations a word shows. In order to apply the first criterion the group of descriptors is discussed taking into account a sample of texts included in The *Helsinki Corpus of English Texts*. The use of written corpora is not free from criticism, as has been pointed out by different scholars (Görlach, 1990: 164). If written corpora present a series of deficiencies, historical ones show more specific constraints. Some of limitations that are usually adduced for the study of historical texts and that are particularly related to the topic of representativeness of the corpus are:

(1) They just include written texts with their limitations, that is, written texts are not representative of all registers, genres, age, sex, or social condition of speakers.
(2) Modern readers do not have access to every text produced at a specific period. They are restricted to some types of texts that contemporary readers considered it was worth copying.
(3) The survival of the original texts is often arbitrary and by chance, which implies the data are not complete as a consequence of the random preservation. This means that there is a random selection of texts.

All these constraints may minimize the representativeness of the sample. The validity of the data has also been questioned on the part of some scholars. The written sources compiled in corpora are limited in size. It follows from here that the sample may be considered invalid because of its provenance and reduced size (Schneider, 2002: 81-90).

As the *Helsinki Corpus* has been widely contrasted and is worldwide accepted, it is taken for granted all the requirements for a suitable selection of the different sources have been met. Thus, the frequency parameter is based on the data retrieved by using this corpus. Nonetheless, the results will show that there are some elements that are not present in the *Helsinki Corpus*. That is why *The Dictionary of Old English Corpus* (henceforth *DOEC*) has also been consulted to try to solve some of the problems encountered when using the former corpus. The results of both will be contrasted to see if there are any significant differences.

The second criterion is the variety of collocation a particular item shows. Faber & Mairal (1999: 187), calling this feature the *Principle of Lexical Iconicity*, refer to the fact that "the greater the semantic coverage of a lexeme is, the greater its syntactic variation". Or put it the other way round, Cortes & Mairal (2002: 20) define it

as "the greater the syntactic coverage of a lexical unit, the higher its position in the semantic hierarchy within a given subdomain". One will be able to find out about the collocations of the items by consulting Roberts and Kay (1995), available online now, which enables users to search for phrases in Old English. In addition, those collocations that appear thrice or more in the *Helsinki Corpus* are also selected. The information contained in this source will be completed by that provided by *A Concise Anglo-Saxon Dictionary* (Clark Hall, 1931).

3 Discussion of results

The combination of both criteria, frequency and range of collocations, should give us reliable clues to be able to determine which elements might have been used as superordinates, although it may turn out difficult to decide which of several was the real hyperonym. Nonetheless, before having a look at some of the examples, it is necessary to comment on the researcher's need to go through the data personally. Obviously the automatic search enables the scholar to handle a wide variety of data, but also poses some problems:

First of all, as we are dealing with an untagged corpus, no distinction is made between homonymy among the different word classes. Thus, *sid* and *wid* can be both an adjective or an adverb. In fact, very often the adverbial combination *side and wide* or vice versa *wide and side* is read through the texts. *Side* can even be a noun, as well. In addition, by browsing an adjective like *mæte*, we find plenty of occurrences which correspond to a verb in third person singular; most of them show the construction *Gyf mon/man mæte*... ('If one considers...'). Even more, if the search is done with the adjective in nominative we can have no occurrences, as happens with *medemlic*. That is why the option with final asterisk is preferred to find all the possible inflections and that is how *medemlice* is found. However, when the option *mæt** is used, non desired examples like *mæterne* are

retrieved as well. It follows from here that the data must be revised in order to filter the ones scholars really need for their analysis.

The researcher must also be aware of language variation. Although the variety of written forms is not so wide as in the Middle English period, we find some alternants, like in the case of *micel* that is found 229 times either as quantifier or descriptor in the nominative case. Furthermore, if the search is carried out by *mycel* with ⟨y⟩ instead of ⟨i⟩ 149 tokens are displayed.

Finally the presence of some items is not so pervasive in the *Helsinki Corpus* as shown in Roberts and Kay (1995). In fact, even if the latter work provides several phrases for *great*, just 13 hints will be retrieved from the *Helsinki Corpus*, out of which 3 correspond to the name of the author, Pope Gregory the Great. This fact must be taken into account, as the search will include the title of the work as well as the name of the author.

From the whole set of items provided by Roberts and Kay (1995) those that could be interpreted as quantifiers have been disregarded, because Martín Arista & Martín de la Rosa handled those data in their article. Nevertheless, elements such as *hwon* or *hwæthwugu* or *micel* are found not only as quantifiers, but also as descriptors. Therefore, all the items that could function as descriptors have been analysed concentrating on those cases where they function as adjectives and not as other word categories. Likewise, nouns implying size have also been disregarded, as BIG and SMALL will basically appear as adjectives in attributive function.

Thus, for the Old English exponents of BIG 21 entries provided by Roberts and Kay (1995) have been revised. Out of these, no occurrences were found for *mismicel*, *swiþlic*, *ormætlic*, *unmætlice*, *uþmæte*, *ungefoglic*, *gehycglice* and *ungehwæde*. The rest of the elements show the following frequency and collocations either in

Roberts and Kay (1995) or in the *Helsinki Corpus* (henceforth *HC*), as displayed in Table 1:

Table 1. *OE exponents of BIG, their frequency and collocations*

	Item	Number of occurrences			Phrases
		0	1–10	> 10	
BIG	*entisc*	×			
	micel			×	*micel lic* ('elephantiasis') *micel wundor* ('great wonder', *HC*)
	mismicel	×			*micel folc* ('big crowd', *HC*)
	great		×		*great sealt* ('coarse salt') *greate wyrt* ('meadow saffron') *seo greate banwyrt* ('unidentified plant') *þa greatan netlan* ('nettle')
	brad			×	*brad hand* ('big hand') *brad sweord* ('broad sword') *brad þistel* ('big thistle')
	eotenisc		×		
	swiþlic	×			
	ormæte			×	
	ormætlic	×			
	unmæte			×	
	unmætlice	×			
	uþmæte	×			
	wid			×	*(on/to) widan feore* ('eternity, for ever')
	sid			×	*side rice* ('large kingdom', *HC*)
	unlytel			×	
	ungemet			×	
	eacencræftig		×		
	ungefog		×		
	ungefoglic	×			
	gehycglice	×			
	ungehwæde	×			

Similarly, for the Old English exponents of SMALL 17 items have been revised, out of which no information was retrieved on any of the following: *forlytel*, *hwonic*, *ieþelic*, *lytl*, as can be seen in Table 2:

Table 2. *OE exponents of SMALL, their frequency and collocations*

	Item	Number of occurrences			Phrases
		0	1–10	> 10	
SMALL	*ieþelic*	×			
	mæte		×		
	medmicel		×		
	medmicle		×		
	medemlic(e)		×		
	forlytel	×			
	lyt			×	
	lytel			×	*lytel forca* ('little fork')
	lytle			×	*se lytla finger* ('the little finger')
	lytlum			×	*lytlum fæce* ('a small interval of time', *HC*)
	lythwon		×		
	læst			×	*se læsta finger* ('the little finger')
	hwæthwugu			×	
	hwon			×	
	hwonic	×			
	hwonlice		×		
	leohtlice		×		

The lack of results about the mentioned units can only be taken as an indication of the low frequency of the terms, although obviously not being included in the compilation does not mean that the words were non-existent. A quick glance at the rest will make readers aware of the fact that there are several items that are etymologically

related in each set. Thus, regarding the primitive embodied by BIG the bigger group is made up of the stem *mæt* with different affixes: *ormæte*, *ormætlic*, *unmæte*, *unmætlice*, *uþmæte*. In the SMALL group we find *lytel* and its variants (*lyt*, *lytle*, *lytlum*, *lythwon*), *medmicel* (*medmicle*, *medemlice*), *hwon* (*hwonlic*, *hwonlice*) and then *mæte*, *læst*, *hwæthwugu* and *leohtlice*. From these the *lytel* family clearly outnumbers the occurrences found for the other items.

If a comparison between both descriptors is established, it can be observed that 9 words out of 21 show no occurrences for BIG; 4 are just represented by very few instances ranging from 1 to 10 and finally just 8 exponents of Old English BIG are found more than 10 times in the corpus. In the case of SMALL, no hints are retrieved for 3 of the 17 items; 7 lexical units show a frequency ranging between 1 and 10 times in the *Helsinki Corpus* and the pending elements, 7 in total, are found on more than 10 occasions. There is not always a coincidence between the most widely recorded words in the corpus used and those whose phases are mentioned as common in Old English by Roberts & Kay (1995), but both criteria can be combined to strive to figure out which ones could have been used as hyperonyms and are, because of that, more frequent.

These results highlight the insufficiency of the *Helsinki Corpus* to provide occurrences for all the searched items on this occasion. Therefore, it seems necessary to look for complementary data in a bigger corpus. As the *Dictionary of Old English Corpus* comprises at least one copy of each text surviving in Old English, it will surely meet all the requirements to be a suitable complement to the study. In fact, the retrieval of data from the *DOEC* proves to be an important supplement. The number of tokens extracted outnumbers those found in the *HC* in most cases. As a matter of fact, most of the items show a higher frequency in the *DOEC* than

in the *HC*, but the real difference is made in relation to the absence of some items.

Regarding the primitive BIG nearly all the items absent from the *HC* are now found in the *DOEC* (*mismicel*, *swiþlic*, *ormætlic*, *unmætlice*, *uþmæte*, *ungefoglic*, and *ungehwæde*) with the exception of *gehycglic*. For instance, Roberts and Kay (1995) mention that this word appeared just once in Anglo-Saxon documents. Furthermore, the term is not registered by Clark Hall (1931) as such, although *hycglic*—without the prefix—appears in *Die Dialoge Gregors den Grossen* 328:16. The only explanation for its absence from the *DOEC* would be that the variant text of Gregory's *Dialogues* included in the *DOEC* introduced a synonym of *hycglic*, rather than this very word.

On the contrary, all the three elements absent for SMALL in the *HC* are now present in the *DOEC*: namely *forlytel*, *hwonlic*, *ieþelic*. However, some of the problems have not been solved by the use of the *DOEC*: for instance, being also an untagged corpus, no distinction is established between parts of speech. Thus, cases of homonymy between word categories must still be worked out by the researcher. This task is even harder as the number of occurrences is much higher. Not only is this tendency observed in cases like the ones mentioned above for *lytel* or *micel* and other related forms, but quite a simple form like *brad* in the *HC*, where 71 tokens are found, appears with 546 hints in the *DOEC*. Thus, the *DOEC* helps in providing more reliable data regarding some items, but presents similar limitations to those found in the *HC* as well.

As the outcomes obtained through the use of the *DOEC* do not alter significantly those retrieved from the *HC*, the presence of the semantic primes will now be illustrated by using examples from those that have 10 or more instances as descriptors in the *Helsinki Corpus*. The exponent that displays the greatest number

of occurrences is *micel*. Even if *micel* and *lytle* can be found as quantifiers, as can be seen in the following passage from the *Anglo-Saxon Chronicle*, most of the occurrences happen to be adjectives in attributive position. Some of them would even be controversial, for instance *micle aþas sworon*, where *micle* could be interpreted as a quantifier and translated as 'swore many oaths' or as a descriptor and be rendered as 'swore great oaths'. The latter is the option chosen by both Savage (1995: 96) and Swanton (2000: 74-75) for the following sentence:

(1) *& micle aþas sworon*

Some other instances of *micel* are clearly quantifiers as in:

(2) *& geridon Wesseaxna lond & gesæton micel þæs folces*
'and they rode over the Wessex land and sat up many of the people'

From the same extract examples with *lytle* as quantifier (3) and descriptor (4) are taken:

(3) *& he lytle werede uniepelice æfter wudum for*
'and he defended a little with greater difficulty through the wood'

(4) *þæs on Eastron worhte ælfred cyning lytle werede geweorc*
'at Easter king Alfred built a little fort'

The other exponents for BIG showing more than 10 occurrences are *brad*, *ormæte*, *unmæte*, *sid*, *wid*, *unlytel* and *ungemet*. Let's see an example of each one:

(5) *ofer brad brimu Brytene soht*
'Brytene sought over the big waves'

(6) *gif ormæte hungor cymð*
'If big hunger comes'

(7) *com werod unmæte*
'a big multitude came'

(8) *we widefeorh weorcum hlodun geond sidne grund*
'we built (with) works through the ample ground for a long time'.

Example (8) could illustrate both *sid* and *wid*; the latter often appears with *feorh* 'life, time span' inasmuch as it is sometimes considered an adverb translated as 'always'.

(9) *to miclan bryne wæter unlytel*
'to a big fire, big/much water'

(10) *þæt is asolcennyss, ðæt is modes swærniss and ungemetegod slapulniss*
'that is laziness, that is sluggishness of spirit and excessive somnolence'.

By going through the various sentences retrieved for BIG, it has been observed that some of the descriptors function as intensifiers for other adjectives implying size, in the same way *extremely*, *enormously* or *immensely* can be used in present-day English. In fact, Clark Hall (1931) records *ormæte* and *ungemet* both as adjectives and adverbs. With this function are found in the *Helsinki Corpus*:

(11) *se mona is ormæte brad*
'the moon is immensely large'

(12) *winter bringeð weder ungemet cald*
'the winter brings extremely cold weather'.

Regarding the specific set of items for SMALL, apart from *lytel/lytle* and other etymologically related units like *lytlum* and *lyt*, the only

ones that are found 10 times or more are *læst*—that is also related to *lytle* as being its superlative—, *hwæthwugu* and *hwon*.

(13) *Babylonia, seo ðe mæst wæs & ærest ealra burga, seo is nu læst*
'Babylon, which was the largest and the first of all towns, is now the smallest'.

Hwæthwugu is documented as noun, adjective, pronoun and adverb in Clark Hall (1931). Thus, many of the occurrences will go under any of the other word classes. Besides, it is one of the few that can present graphic variants, since the last part of the compound can be recorded as *-hwigu*, *-hugu*, *-hwega*, *-hwegu* or *-hwygu*. An example of *hwæthugu* as descriptor is found in (14):

(14) *& gedyde hwæthugu getæse*
'and did little profit'.

Finally, one of the items showing the greater number of occurrences is *hwon*, as it can be a form of the interrogative *hwa*, found in adverbial phrases like *to hwon*, *for hwon* meaning 'why'; this is probably the most widely found case. Besides, it can be an adjective, a noun and an adverb. Most of the hints supplied are either as an inflected form of *hwa* or as an adverb. As a descriptor is found in:

(15) *geond þas eorðan æghwær sindon hiora gelican hwon ymbspræce*
'through the earth everywhere is their same little speech'.

4 Conclusions

In the introduction section, the goals of the present research were set up: Firstly, to check whether the Natural Semantic Metalanguage Research Program represented a suitable theoretical

and methodological framework for the lexical and semantic study of Old English and secondly, to establish the 'hyperonym(s)' among all possible equivalents of BIG and SMALL extracted from the Old English Thesaurus by Roberts and Kay (1995) by means of the interpretation of the data retrieved from the *Helsinki Corpus of English Texts* and the *Dictionary of Old English Corpus*.

Regarding the first objective of the investigation, it can be concluded that, although the Natural Semantic Metalanguage Research Program has dealt with various natural languages, English is probably one to which little attention has been paid. Thus, one of the initial aims of the article was to test the validity of the Natural Semantic Metalanguage Research Program as a useful tool in a former stage of English, namely Old English. Even if this piece of work attempted at being just a first approach to the topic, there is no doubt that the findings and proposals of the Natural Semantic Metalanguage Research Program may be tested in Old English. With this brief analysis I hope to have demonstrated that semantic primes represent a useful tool for the lexical and semantic analysis of Old English inasmuch as it can certainly be applied to other levels of linguistic analysis, such as morphology and syntax, for instance. Furthermore, it seems convenient to extend the research to other categories established by Goddard and Wierzbicka (2002) in order to find the exponents in Old English and to present the specific peculiarities this stage of the language shows.

Although the first objective was clearly achieved, there are some issues related to the second objective that need further research, such as which of the terms can clearly be considered hyperonyms. As a matter of fact, this second objective was tested and developed in the discussion of the data retrieved from Roberts and Kay (1995) and the two corpus used: *Helsinki Corpus* and *The Dictionary of Old English Corpus*. Even if some methodological problems were solved

by handling the data provided by the latter, such as the retrieval of some items absent in the former, there are still some deficiencies that prevail. Probably the most salient one is the fact that, being both untagged corpora, there is a need on the part of the researcher to filter the data in order to select the right information for the investigation.

Isabel de la Cruz Cabanillas
Universidad de Alcalá de Henares

References

Primary sources

Clark Hall, J. R. 1931: *A Concise Anglo-Saxon Dictionary*. Cambridge University Press, Cambridge.

The Dictionary of Old English Corpus. Online version available at: http://ets.umdl.umich.edu/o/oec/

Oxford English Diccionary. Online version available at: http://dictionary.oed.com

Rissanen, M. & O. Ihalainen. 1991: *The Helsinki Corpus of English Texts. Diachrnic and Dialectal*. Universidad de Helsinki, Helsinki.

Roberts, J. & C. Kay. 1995: *A Thesaurus of Old English*. King's College London Medieval Studies, London and online version available at: http://libra.englang.arts.gla.ac.uk/oethesaurus

Secondary sources

The Anglo-Saxon Chronicles. 1995: Translated and collated by A. Savage. Crescent Books, New York.

The Anglo-Saxon Chronicles. 2000: Translated and edited by M. Swanton. Phoenix Press, London.

Cortés, F. & R. Mairal 2002: A preliminary design for a syntactic dictionary of Old English on semantic principles. Díaz Vera, J. E. ed. *A Changing World of Words. Studies in English Historical Lexicography, Lexicology and Semantics*. Rodopi, Amsterdam: 3-46.

Faber, P. & R. Mairal 1999: *Constructing a lexicon of English verbs*. Mouton de Gruyter Berlin.

Goddard, C. 1994: Semantic theory and semantic universals. Goddard, C. & A. Wierzbicka eds. *Semantic and Lexical Universals. Theory and Empirical Findings*. John Benjamins, Amsterdam: 7-29.

Goddard, C. 2002a: The Search for the Shared Semantic Core of All Languages. Goddard, C. & A. Wierzbicka eds. *Meaning and Universal Grammar. Theory and Empirical Findings* (Volume I). John Benjamins, Amsterdam: 7-40.

Goddard, C. 2002b: The On-going Development of the NSM Research Program. Goddard, C. & A. Wierzbicka eds. *Meaning and Universal Grammar. Theory and Empirical Findings* (Volume II). John Benjamins, Amsterdam: 300-321.

Goddard, C & A. Wierzbicka 1994: Introducing lexical primitives. Goddard, C. & A. Wierzbicka eds. *Semantic and Lexical Universals. Theory and Empirical Findings*. John Benjamins, Amsterdam: 31-54.

Goddard, C. & A. Wierzbicka. 2002: Semantic Primes and Universal Grammar. Goddard, C. & A. Wierzbicka eds. *Meaning and Universal Grammar. Theory and Empirical Findings* (Volume I). John Benjamins, Amsterdam: 41-85.

Görlach, M. 1990: Corpus problems of text collections: linguistic aspects of the canon. *Studies in the History of the English Language*. Carl Winter, Heidelberg: 163-78.

Martín Arista, F. J. & Mª V. Martín de la Rosa 2006: Old English Semantic Primes: Substantives, Determiners and Quantifiers. *ATLANTIS* 28/2: 9-28.

Mairal Usón, R. 2005: Reflexiones en torno a la noción de 'universal lingüístico'. Fuertes Oliver, P. coord. *Lengua y sociedad: Investigaciones recientes en Lingüística Aplicada*. Universidad de Valladolid, Valladolid: 83-99.

Mairal, R. & J. Gil eds. 2003: *En torno a los universales lingüísticos*. Cambridge University Press, Cambridge.

Moure, T. 2001: *Universales del lenguaje y linguo-diversidad*. Ariel, Barcelona.

Schneider, E. 2002: Investigationg Variation and Change in Written Documents. Chambers, J. K., P. J. Trudgil and N. Schilling-Estes eds. *The Handbook of Language Variation and Change*. Blackwell, Oxford: 67-96.

Wierzbicka, A. 1989: Semantic primitives: The expanding set. *Quaderni di Semantica* 10/2: 309-332.

Wierzbicka, A. 1994: Semantic Primes Across Languages: A Critical Review. Goddard, C. & A. Wierzbicka eds. *Semantic and Lexical Universals. Theory and Empirical Findings*. John Benjamins, Amsterdam: 445-500.

Wierzbicka, A. 1996: *Semantics: Primes and Universals*. Oxford University Press, Oxford.

Wierzbicka, A. 1998: Anchoring linguistic typology in universal semantic primes. *Linguistic typology* 2/2: 141-194.

Wierzbicka, A. 2002: Semantic Primes and Linguistic Typology. Goddard, C. & A. Wierzbicka eds. *Meaning and Universal Grammar. Theory and Empirical Findings* (Volume II). John Benjamins, Amsterdam: 257-300.

Received 28 Feb 2007; *revision received* 28 May 2007; *accepted* 12 Jun 2007