Are English Words Loners? An Inquiry into the Motivation and Word Family-Integration of the English Lexicon

Christina Sanchez
Friedrich-Alexander-Universität Erlangen-Nürnberg
Institut für Anglistik: Linguistik (PhD Student)
Bismarckstr.1
91054 Erlangen
Germany
Christina.Sanchez@web.de

Abstract
Ever since Ferdinand de Saussure's classification of English as a *langue lexicologique*, it has been widely believed that only a comparatively small part of the English lexicon is morpho-semantically motivated. Ernst Leisi's explanation for this phenomenon is the large proportion of Romance words that have become part of the originally Germanic language. I have subjected his hypothesis that many English words are completely isolated to an empirical analysis by examining the 3000 most frequent spoken and written words in the *Longman Dictionary of Contemporary English* with regard to their transparency and integration into word families.

1 Theories: English as a *Lexicological* or *Dissociated* Language

The following constitutes a preparatory study for my doctoral thesis. The aspect of English vocabulary I am interested in is a theory which goes back to Ferdinand de Saussure, who states in his *Course in General Linguistics* that all languages contain words which are relatively motivated and others which are not. For example, even though the form *singer* is phonetically arbitrary, it can be analysed into smaller elements which give a clue to the word's meaning: a singer is a person who sings. *Sing*, however, cannot be analysed further. De Saussure (1974: 133-134) claims that languages differ according to the degree to which such an analysis is possible: "we might say that languages in which there is least motivation are more *lexicological*, and those in which it is greatest are more *grammatical*. [...] motivation plays a much larger role in German than in English."

This claim was taken up by the reputed Swiss scholar Ernst Leisi in his book *Das heutige Englisch* (1955), which is widely used in academic teaching in Germany. Leisi believes that many English words are completely isolated and do not belong to any word family at all. One of the examples he gives is German *mündlich*, 'oral', which allows the language user to perceive the morpho-semantic relation with *Mund*, 'mouth'. English *oral* and *mouth*, however, share no formal features. Leisi explains this phenomenon, which he terms *dissociation*, as the direct result of the large number of Romance words that, in the course of its history, have become part of the originally Germanic English language. Indeed, Thomas Finkenstaedt's and Dieter Wolff's (1973) computerised study of the *Shorter Oxford English Dictionary* proves that overall, the proportion of Romance words exceeds 50 % –

but does this necessarily mean that there are hardly any morpho-semantic relations between the elements of the English lexicon?

2 The Research Project

2.1 Methodology

2.1.1 The Sample. As a first step, a sample of words had to be chosen in a sufficiently random way so that they were not influenced by my intentions. According to Finkenstaedt and Wolff (1973), there is a correlation between the first letter of a word and its etymological origin. Therefore, the sample was supposed to include words from the whole range of the alphabet and to be large enough to be representative – as well as relevant for learners of English. These criteria were all met by the most frequent words as marked in the *Longman Dictionary of Contemporary English* (LDOCE). The advantage of this sample is that it is based on a contemporary corpus not only of the written, but also of the spoken language.

2.1.2 The Item List. The combination of the 3000 most frequent items in both varieties yields a list of 3614 words. All figures refer to the third edition, the most recent one at the time of elaboration. This number was reduced to 3094 for two reasons: firstly, many items belong to more than one word-class, which is due to the fact that zero-derivation is a very frequent phenomenon in English. The noun *bottle* can thus be converted into the verb *bottle*. Conversely, the verb is motivated by the noun. As the aim of my analysis was to investigate the morpho-semantic integration of words in the analytical as well as in the synthetic direction, conversion would merely have distorted the results in one direction for one word-class of such a pair and in the other direction for the other word-class and was therefore excluded: where a formally identical second form was included in the original list, both items were merged into one. This treatment has the additional advantage of allowing us to avoid making statements about word-classes in the rest of the procedure. As Laurie Bauer (1983: 202) points out in the analysis of the word *rattlesnake*, *rattle* could be a verb (in the underlying sentence *the snake rattles*) or a noun (in the underlying sentence *the snake has a rattle*). A similar argument holds true for adverbs derived by means of the suffix *-ly*. According to the *Comprehensive Grammar of the English Language* (Quirk et al. 1985: 1556), "-LY [...] can be very generally added to an adjective in a grammatical environment requiring an adverb [...], so that it could almost be regarded as inflectional". As a consequence, adverbs ending in *-ly* were included under the corresponding adjective. This treatment also avoids the distortion mentioned above.

2.1.3 The Analyses. Subsequently, each item was subjected to the following procedures: firstly, it was determined whether it is analysable within the high-frequency list and then, whether there are motivating elements within the whole of the LDOCE, which also contains lemmatised affixes that were not included in the frequency count. For example, *specialize* is motivated by the adjective *special* and the suffix *-ize*. The second step deals with the opposite direction: word-formation. For *specialize*, none was found within the high-frequency vocabulary. However, the LDOCE contains the adjective *specialized*. Had this not been the case, another search would have been carried out in the *Shorter Oxford English Dictionary* (SOED). By restricting the analysis to words contained in the common reference

works, we can expect to deal with established items only. Wild card search of the corresponding electronic versions reduced the danger of overlooking any relevant word-formations to a minimum.

In the case of the expanded forms, we can safely assume that there is a very high level of agreement between different researchers. The study of motivation, however, deserves a few further comments. As one of my aims is to look at motivation and word-family integration from a cognitive point of view and to consider their effects on language acquisition and learning as well, the analysis necessarily had to be synchronic and could not rely on established etymological dictionaries. The underlying concept is best described by Gerhard August's *synchronic etymological competence*, i.e. the ability of "normal" speakers to analyse complex words until they arrive at core words which cannot be analysed any further (August 1998). According to August, the "normal speaker" is a layperson who cannot rely on particular knowledge of subject fields, foreign languages or even linguistics. The very enterprise of determining such synchronic motivation cannot avoid a certain degree of subjectivity: a computer could easily analyse words into graphic segments that correspond to morphemes, but a subsequent semantic analysis would need to be carried out by humans to rule out such analyses as *bishop* into *bi-* and *shop*, and *coalition* into *coal* and *-ition*. However, I have attempted to make my work as objective as possible in several different ways. Firstly, I only consider words to be motivated if semantically related shorter elements are completely integrated into the longer form. This may be on the phonetic and on the graphic level, such as in *singer*, which yields *sing* and *-er*, but also on the phonetic level only: the <i> in *daily* does not impede the analysis into *day* and *-ly*. From the point of view of foreign learners of English, who frequently encounter words for the first time in their written form – or at least soon afterwards – it also makes sense to admit words which are perfectly analysable on the graphic level only: *similarity* into *similar* and *-ity*, and *election* into *elect* and *-ion*. These self-imposed restrictions should ensure a maximum of objectivity. However, they are so strict that many words which we find intuitively transparent would have to be discarded. For example, it is unlikely that laypeople would find it difficult to decompose *demonstration* into *demonstrate* and *-ion*. *Capability* and *capable* share enough phonemes and graphemes to be reminiscent of each other, too. That is why I introduced the category of *partial transparency* and its corresponding synthetic counterpart, which includes cases like *remedy – remedial*. This allows me to maintain a clear dividing line between the evident cases and the formally or semantically slightly less evident analyses which I would still like to include. Within this intermediate category, there is necessarily a certain degree of gradience, because each case is unique: is the noun *account* motivated by the verb *count*? Is *business busy* in some way or another? Are *views* exchanged in an *interview*? No single rule can master all these variations, and different researchers can come to different conclusions here. That is why I have included a list with all analyses in my preparatory study, so that readers can compare their own decisions with mine. For the doctoral thesis, I intend to carry out a questionnaire study with informants to determine the range of interpersonal variation as well.

2.2 Results

If partial motivation can be considered as a valid form of motivation, which I consider to be the case, then the meaning of 42.8 % of the most frequent English words can be derived from their constituents – at least to a certain extent, because world knowledge plays an immensely important part, as Hans-Martin Gauger (1971) postulates. However, even if we decide to stick to the clear cases only, we are still left with a motivation rate of 30.3 %. It is unlikely that that of German would be significantly higher. The results for the opposite direction are even more striking: 93.7 % of the items on the list can serve as the basis for new word-formations. Here, the additional borderline cases amount to only 0.9 %. For the empirical consideration of Leisi's theory, the most interesting items are those which belong simultaneously to the 57.2 % of words which are not motivated and the 5.5 % that are not expandable in any way, that is, the completely isolated vocabulary. In my analysis, there are 26, which constitutes 0.8 % of the reduced word list. By relaxing standards, for example by allowing the inclusion of word-formations other than lemmas, and by consulting the *Oxford English Dictionary* (OED) as well as the *British National Corpus* (BNC), a word family could be found for all items but one. Thus, *both* could be expanded to *both-sided* in the OED, and *since* to *since-dissolved* in the BNC. The only exception was the interjection *ta*, 'thank you', where the number of hits containing the grapheme sequence is so immense that it was not possible to work through all of them.

In the end, we are left with a single word that is neither motivated nor expandable, which accounts for 0.03 % of the listed items. Is this really enough to claim that the English lexicon is dissociated, particularly in relation to the German language? Time and my doctoral thesis will tell.

3 References

Augst, G. (1998). Wortfamilienwörterbuch der deutschen Gegenwartssprache. Tübingen: Niemeyer.
Bauer, L. (1983). English Word-Formation. Cambridge: CUP.
Finkenstaedt, T. and Wolff, D. (1973). Ordered Profusion. Heidelberg: Winter.
Gauger, H.-M. (1971). Durchsichtige Wörter. Zur Theorie der Wortbildung. Heidelberg: Winter.
Leisi, E. and Mair, C. (1999). Das heutige Englisch. Wesenszüge und Probleme. 8., neubearbeitete Auflage. Heidelberg: Winter.
Longman Dictionary of Contemporary English with New Words Supplement (2001). 3. Ausgabe 1995, erweitert 2001. Mit CD-Rom. München: Langenscheidt-Longman.
The New Shorter Oxford English Dictionary (1996). Windows™ Version 1.0.3. Oxford: OUP.
The Oxford English Dictionary (1993). Second Edition on Compact Disc for the Apple Mackintosh®. Oxford: OUP.
Quirk, R. et al. (1985). A Comprehensive Grammar of the English Language. London: Longman.
de Saussure, F. (1974). Course in General Linguistics. Edited by Charles Bally and Albert Sechehaye in Collaboration with Albert Reidlinger. London: Fontana.