Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

Cronfa - Swansea University Open Access Repository

_____

This is an author produced version of a paper published in :
Word-formation: an international handbook of the language of Europe

_____

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa22249

_____

_____

Tuggy, David
 2005   Cognitive approach to word-formation. In: Pavol Štekauer and Rochelle Lieber (eds.),
        Handbook of Word-Formation, 233−265. Dordrecht: Springer.
Windsor, Jennifer
 1993   The functions of novel word compounds. *Journal of Child Language* 20: 119−138.

Hilke Elsen and Karin Schlipphak, Munich (Germany)

121. Word-formation in second language acquisition

1. Introduction
2. The study of second language acquisition
3. Learning word-formation
4. Types of word-formation processes
5. Conclusion
6. References

Abstract

This article explores the role of word-formation processes in the acquisition of a second-language lexicon and the relevant concepts from second language acquisition theory, among them the "word family". The acquisition of word-formation knowledge influences the structure of a speaker's mental lexicon, but teaching word-formation explicitly is sometimes seen as an optional extension for advanced learners. Dictionaries can help learners develop their knowledge of word-formation, but only electronic versions can make full use of the relevant information.

1. Introduction

The interface of word-formation and second language acquisition is not a regular topic in either field. In the study of word-formation, the absence of attention to second language acquisition can be explained as a reflection of the focus on word-formation rules as such and their implementation in the (native) speaker's competence. Whereas data from first language acquisition can be seen as giving information about the origin and form of word-formation rules in the mental lexicon, it is not immediately clear how data from second language acquisition can add to this in any substantive way. Conversely, the field of second language acquisition does not pay much attention to word-formation studies. A recent handbook, Gass and Mackey (2012), does not have index entries for *word-formation*, *derivation* or *compounding*. Also in standard textbooks of second language

acquisition, e.g., Gass and Selinker (2008) or White (2003), word-formation is not mentioned.

2. The study of second language acquisition

Second (or foreign) language acquisition has been a topic of interest for a long time, but the scientific study of it is quite a young field, with the first important publications dating from the 1960s and 1970s (Gass and Selinker 2008). Second language acquisition is distinguished from first language acquisition by a focus on language acquisition by adolescents and adults who have learned another (first) language as a child. The reason for establishing second language acquisition as a separate field of study is that there are certain properties of the learning process that are relevant to language acquisition after L1 acquisition has been completed and that are shared by most, if not all learners.

2.1. Interlanguage

In the early days of interlanguage study, the obvious differences between texts produced by native speakers and those produced by learners provided the data for the field, with errors being particularly interesting, both for descriptive approaches and for those attempting to model this language form. The area of error analysis, often closely coupled with contrastive analysis, seemed to be a promising field in the 1950s, but the emergence of the mentalist approach in the 1960s and the rising interest in corpus linguistics in the 1980s changed the orientation of the field.

The Chomskyan revolution led to an even stronger focus on the systematic aspects in second language acquisition than had been the case in error analysis. The identification of developmental patterns common to all or most learners of a specific second or foreign language, not just those sharing a particular first language, became the research aim. In parallel to research in first language acquisition, linguists attempted to establish the order of acquisition of grammatical structures by second language learners.

As observed by Flynn (1996: 127), second language learners have two apparent advantages compared to those acquiring their first language. Whereas first language acquisition is concurrent with the child's cognitive development, adults involved in second language acquisition have already completed their cognitive development, which gives them a much wider range of problem-solving skills. Moreover, by definition, the second language learner already has a first language. It is therefore something of a paradox that second language acquisition seems more difficult and generally results in a considerably lower level of proficiency than first language acquisition.

In generative linguistics, the paradox of second language acquisition is explained with the help of a language acquisition device, which is available to children, but subject to a critical period. After this period, language acquisition can no longer take place in the same way as first language acquisition. Chomsky (1959: 42) used such observations to motivate the central place of language acquisition in linguistic theory. Following on from these ideas, Pit Corder and Larry Selinker developed the concept of *interlanguage* (cf. Selinker 1972, 1992). Interlanguage is the competence in a second language one is

learning. An essential observation concerning interlanguage is that it is a system that arises in the learner's mind/brain, not simply a deficient version of a standard language. Research into the common, systematic aspects of the second language acquisition process also contributed to the birth of this concept.

Most learners' interlanguage progresses to the point where their proficiency is sufficient for their needs, and then fossilizes to some extent. Fossilization (or stabilization) means that interlanguage competence ceases to develop towards the native-speaker standard, in spite of exposure to further data that would logically be sufficient to correct the divergences, cf. Han (2012) for an overview. In pronunciation, fossilization results in a stable foreign accent. Fossilization is relatively well researched where knowledge of syntax is concerned, but much less in other components of language.

Given the background of the concept of interlanguage, it is not surprising that most of the research into second language acquisition processes from this perspective concentrates on syntax and inflection. Textbooks such as White (2003), but also handbooks such as Ritchie and Bhatia (1996) deal almost exclusively with these two areas.

2.2. Vocabulary acquisition

Compared to the study of interlanguage syntax and phonology, the study of how learners manage to master a sufficiently large vocabulary has traditionally been less of a focus, despite the fact that language learners themselves typically recognize the need and potential benefits of investing their efforts in word learning activities. Recent decades have seen a rise of interest in the acquisition of the lexicon, however, both in terms of research on the lexicon and its acquisition, and in terms of practitioners' awareness of the relevance of systematic vocabulary learning.

Research questions include the typical speed of acquisition of second language vocabulary and the desired minimal vocabulary size for a given level of communicative competence. In order to answer the latter question, we have to measure a speaker's vocabulary size. A reliable measure of native speakers's vocabulary size would give researchers a *tertium comparationis*, while for learners, vocabulary size can be seen as one way of measuring L2 competence (Jeon 2011; Schmitt and Meara 1997), especially in a morphologically simple language such as English. However, the methodology for measuring the size of a speaker's vocabulary is not a trivial issue.

Estimates of the size of a native speaker's mental lexicon have occasionally been used as a starting point for deciding suitable target figures for second language learners' vocabulary. Published estimates of the number of words native speakers know originate from experiments where native speakers were tested on a sample of dictionary entries or from studies on reading. Nagy and Anderson (1984), for example, examined texts from US English school books to estimate the vocabulary size needed to understand these texts. They calculated that school children are expected to be able to cope with printed text that contains over 100,000 lexemes. This figure already takes into account inflectional variations and other morphological relationships within a word family that are deemed to be semantically transparent. Such a figure is clearly beyond the reach of practically all second language learners, even if the words are further grouped into larger word families.

A more realistic aim can be gauged from the vocabulary content of successful language course materials, and from the results of language tests that include measures for vocabulary size. In terms of order of magnitude, many of these agree with Nation (2001) who gives the figure of 2,000 words as a minimum necessary for basic communication for every learner. In order to be able to read and thus enlarge their lexicon autonomously, learners are thought to need about 5,000 words (in English).

Given that learners typically have limited time in which to reach the necessary level in their second language and the large number of words needed for communication, the questions for applied linguists are therefore which words of these 2,000−5,000 target words should be taught first, in which order extensions of this basis should be treated, and how learners can reach the vocabulary threshold in their interlanguage development as efficiently as possible. Word-formation processes can play a role in this context, especially for the comprehension of written texts, texts that often contain a wider range of vocabulary and more members of a word family than informal spoken language.

Concerning the order in which words should be taught in vocabulary acquisition, there is again general agreement (Nation 2001) that learners need to start their lexical acquisition process with the words that give them the highest coverage of texts, i.e. very frequent words. This basic vocabulary needs to be taught and learned actively. Only when learners reach a stage where their vocabulary covers between at least 95 and 98 % of running words in a text, will they be able to infer the meaning of unknown words from context.

It has commonly been assumed that learners acquire most of the words beyond their very basic vocabulary in an incidental manner, typically through reading and listening activities that focus on meaning. Especially among native speaker teachers of English as a foreign language, this led to the view that the teacher could do little during class to help learners with their vocabulary learning, and conveniently matched the widely adopted framework of communicative language teaching (Brumfit and Johnson 1979; Littlewood 1981). In fact, for teachers influenced by the American structuralist tradition going back to Fries (1945) and in essence continued in Halliday, McIntosh and Strevens (1964) in Britain, the main focus of language teaching was not vocabulary. Phonology and syntax were seen as the learners' main problems; a small number of content words would therefore be enough to practice the structural aspects of the foreign language.

In this perspective, the best approach to selecting vocabulary for teaching would be to keep the vocabulary manageable and ensure that the words chosen for phonological and grammatical exercises were the most useful ones in the circumstances, a task that has been attempted several times in the form of vocabulary control movements. The most prominent of these attempts is probably Ogden's (1930) *Basic English* (for a linguistic assessment, see Allerton 2002). A later and much more successful attempt is West's (1953) highly influential *A General Service List of English Words*, a list of 2,000 English words which not only includes frequency information for each word, but also lists major senses with their frequencies. West's list provided a basis for the development of graded reading programmes by a number of publishers. To date, there is no truly equivalent update for contemporary English, despite huge advances in corpus linguistics (but see Brezina and Gablasova 2013 and Browne 2014 on attempts for updates). Another practical application of a restricted defining vocabulary is found in LDOCE (1978). This learners' dictionary tried to help learners understand definitions by paraphrasing them using a basic vocabulary of only 2,000 words.

For French, the classic example of such a vocabulary list is Gougenheim's (1958) *Dictionnaire Fondamental de la Lange Française*. This list presents 3,000 French words divided into two levels (cf. also Gougenheim et al. 1964). Modern corpus linguistics has given a new lease of life to this area, hugely simplifying the collection of frequency information on individual words, which can then be used as one of the main criteria for inclusion in the vocabulary lists for learners. A modern example of this genre is Davies's (2006) list of 5,000 frequency-ranked words in Spanish. It is presented as "core vocabulary for learners" and the method developed for Spanish was subsequently applied to a range of other languages.

What is lacking in most of the modern applications, however, is any attempt at dealing with frequency of word senses. Frequency information is typically only available for orthographic words and derives from large corpora. The resulting word lists suffer from a fairly fundamental problem closely related to the idea of the word family, as discussed by Gardner (2007). Word frequency counts that have been compiled using computers (rather than human readers) are based on the orthographic word. Running text is segmented into word tokens, and these are then counted. Instances of *was* are counted separately from those of *were* at this stage. In a next step, the counts of inflected forms are grouped to give the total counts for the lemma. That *replace* and *replaces* are assigned to the lexeme *replace* is obvious, but there are numerous cases where the assignment to a lexeme is far from trivial. Should *left* be assigned to the adjective (opposite of *right*) or to the verb *leave*? Apart from these obvious cases of word forms belonging to different lexemes, there is also the question of polysemy. While it makes sense from a pedagogical point of view to group the singular *file* and the plural *files* for the purpose of frequency counts, it is much less obvious whether *file* in the senses of 'folder', 'data set on a computer', 'line' and 'tool' should be grouped into a single word that needs to be learned when the learner reaches the relevant stage in the frequency list.

2.3. Word families

In an attempt to find other ways of helping learners and their teachers cope with the huge task of learning large numbers of words, Bauer and Nation (1993) raised awareness of the potential of inflection and word-formation for the teaching of English as a second language. The word family is defined as consisting "of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (1993: 253). Bauer and Nation recognize that some etymologically related forms are too far apart semantically (*busy − business*; *fatal − fatality*) to count as belonging to the same word family; such items have to be learnt individually.

On the basis of criteria such as frequency of an affix, productivity, and regularity at various levels, Bauer and Nation (1993) sort English affixes into five levels, understood to go from easiest to hardest, with the lowest level containing inflectional affixes only. The next level contains the most frequent and regular derivational affixes, e.g., English *-able*, *-ness,* and *un-*. The authors' main aim in proposing this classification was to provide a basis for guidelines for the active teaching of derivation to support learners' receptive vocabulary knowledge. Other, related aims were to provide a tool for improving learners' dictionaries and for allowing for better comparison of vocabulary size estimates, going beyond the simple token-type-lemma series.

Although the authors suggest that their hierarchy needed to be tested, no programme of systematic testing has ever been carried out. As one of the few studies of a more limited scope, Mochizuki and Aizawa (2000) tested affixes from Bauer and Nation's hierarchy with Japanese learners of English. Their data show that learners clearly increase their affix knowledge in line with their overall acquisition of vocabulary. The study also supports the assumption that transfer plays a role in the acquisition of derivational knowledge in the second language. Despite the absence of supporting evidence from systematic testing, the principle of teaching learners at least the most common affixes in order to improve their vocabulary comprehension has been widely adopted in vocabulary teaching materials, especially for academic English, and is frequently used when target numbers for vocabulary knowledge for certain groups of learners are cited (e.g., Coxhead 2000).

An implicit or explicit understanding of the processes that link derived words to their bases is generally thought to be beneficial for widening the lexicon. Schmitt and Meara (1997) even use suffix knowledge as an indirect measure of vocabulary size, arguing that understanding derivational patterns gives access to more members of a word family. Similarly Olshtain (1987) uses knowledge of word-formation processes as an indicator of high second language competence and argues that advanced learners of Hebrew as a second language manage to approach native speakers in their ability to interpret new forms.

In Bauer and Nation's (1993) sense, a word family is a dynamic concept that evolves with the learner's proficiency, or the complexity of the learner's interlanguage. The size of the word family increases with the learner's knowledge of affixation, as more advanced learners are likely to be able to understand more of the semantic links that can be found within a word family, but does not stretch to cases where the meaning of the affixed word is too far away from the base word to understand it (as in *hard* and *hardly*). This incremental concept of the word family is easy to accommodate in theoretical models of vocabulary development, such as Jiang's (2000), but much harder to translate into static materials for language learning.

In general, awareness of word families is seen as facilitating the learner's task (Pons-Ridler and McKim 1985; Laufer 1997; Schmitt and Zimmerman 2002, among others). However, it is quite a complex issue to determine the decrease in learning effort brought about by their use in teaching materials. It is recognized that suppletive forms require a learning effort comparable to a new lexeme and thus differ substantially from regular inflectional processes in this respect. Word-formation processes have received much less attention and only tend to come to the attention of language teachers and researchers when learners are thought to have some problem with them. One area that has thrown some light onto the productive aspect of word-formation knowledge in second language learners is error analysis (cf. James 1998). Problems in comprehension are probably more frequently highlighted when learners' difficulties in reading can be attributed to lack of vocabulary knowledge. In such a context, teaching learners about the more common and productive word-formation processes is seen as a strategy that can help them deal with unknown words while reading (Huckin, Haynes and Coady 1993). However, it is probably safe to say that, along with many teachers, Bauer and Nation (1993) tend to overestimate the facilitative effect of word families in comprehension.

For production, the situation is even more complex. In a rare study of transfer of word-formation strategies, Whitley (2004) shows that learners make creative use of sec-

ond language morphology when they find themselves in a situation where they are trying to express an idea and need to fill a lexical gap in their second language lexicon. In Whitley's study, English-speaking learners of Spanish were asked to write expository essays in Spanish, without access to a dictionary or other linguistic help. Some word-formation strategies were used frequently, including zero-derivation. Suffixes that have a direct equivalent in English were used particularly frequently, e.g., Spanish -*mente*, which forms adverbs in a way similar to English -*ly*. The learners in this study have mental lexicons with incomplete word families and, when put into a situation with no access to help from other speakers or from dictionaries, they resort to using simplified and sometimes erroneous expressions to get (some form of) the intended meaning across. This can be interpreted as a transfer of first language intuitions for derivational patterns to the second language in a situation where the second language is not developed sufficiently well to allow the learner to produce a text without such support.

3. Learning word-formation

There are various ways in which language learners can use word-formation rules to enhance their knowledge of the language they are learning. They can be viewed from three perspectives: comprehension, production, and structuring the lexicon.

3.1. Comprehension

The gap in vocabulary between the learner's interlanguage lexicon and words they encounter in authentic texts may lead to comprehension problems. In a face-to-face conversation, meaning can often be negotiated with the interlocutor, but this option is not available when reading. Written texts also tend to contain a larger variety of words than spoken language. Learners can be taught strategies to deal with unknown words, including guessing from context and breaking up morphologically complex words. Being able to understand such complex words is often seen as a sign of high competence in the second language (cf. Schmitt and Meara 1997), as these are quite frequent in written and more formal registers. Especially the many derived words belonging to an academic register are a popular target area in courses for advanced learners. Frequently occurring word-formation types are seen as worthy of drawing learners' attention to them, with a view of helping learners cope with unknown words in written texts and of reducing the learning burden of acquiring a large lexicon (Schmitt and Zimmerman 2002).

After illustrating the long tradition of the study of affixation, Nation (2001) stresses the importance of learning word parts (morphemes) in order to improve learners' vocabularies. A number of studies cited in Nation (2001) show that a relatively small number of very frequent English prefixes (e.g., *non-*, *un-*, *in-*) and suffixes (e.g., -*able*, -*er*, -*less*, -*ness*) provide a sufficient coverage of English derivational affixes for most learners. The recommended teaching method is simply to have learners look at 10 or 20 words containing the target affix and bases they have learned before and asking them to work out the meaning of the derived words (and by implication the meaning of the word-formation process).

The acquisition of such derivational rules is one of the last stages in the acquisition of words (Jiang 2000), a sequence which mirrors that in first language acquisition, and many of a learner's word families are likely to fossilize or reach a lexical acquisition plateau well before the full set of derivations are acquired, unless the learner is taught to use morphological knowledge to decode the meaning of derived words. The incremental nature of vocabulary acquisition can be shown to be at work here as well, both in terms of the various derivations within a word family and of the continuum from passive to active word knowledge. Schmitt and Zimmerman (2002) show that adjective and adverb derivations in particular need support from teaching to help acquisition.

3.2. Production

In considering the use of word-formation rules in production, we encounter a crucial difference between word-formation rules and syntactic rules. Whereas syntactic rules are used to combine names of concepts into sentences, word-formation rules are used to name new concepts.

We can observe this by contrasting two approaches to productivity. Chomsky (1964: 7) states that "[t]he central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately". This clearly relates to syntactic productivity. In discussing word-formation, Corbin (1987: 176−178) distinguishes three types of productivity: *régularité* 'regularity', *disponibilité* 'availability' and *rentabilité* 'profitability'. Regularity refers to the predictability of the form and meaning of the outcome of a rule. Availability refers to whether the rule can be used to create new expressions. Profitability refers to the extent that the rule is actually used in cases where it can be used. This type of distinction is not necessary for syntactic productivity, but very usefully analyses the concept of morphological productivity.

The three concepts of productivity distinguished by Corbin enable us to express the learner's problems in using word-formation rules correctly in production. In the case of a syntactic rule, once the learner has learned it, they can use it in comprehension and production. However, a word-formation rule that is available is not necessarily used in all situations where it could be used (profitability). The prefix *un-* is certainly available for new formations, but alongside *unacceptable* we have *inconceivable*, not **unconceivable*. A learner who does not know *inconceivable* will have no reason to suspect that **unconceivable* is not correct. Lack of regularity constitutes a separate problem. Even when the learner comes across a particular form, they cannot assume that it has the regular meaning. Thus, *invaluable* is not the opposite of *valuable* and using it in such a sense would be very confusing.

Corbin (1987: 177) argues that availability is the central, underlying sense of productivity and that the other two are derived. In the study of word-formation in the lexicon, this position is warranted. After all, profitability is a property that is derived from availability in the way performance is derived from competence, because a rule must exist before it is possible to measure how often it is used. Regularity is a property of individual rule applications, which again presupposes that the rule is available. Corbin's conclusion is therefore that we should only concentrate on availability.

In the context of second language acquisition, the perspective is somewhat different. On one hand, higher degrees of profitability and regularity of a rule make it more worthwhile for learners to invest the effort of learning the rule, on the other they reduce the risk of accidentally creating non-existing words or using words that are lexicalized in an idiosyncratic way. This means that the study of regularity and profitability of word-formation rules is much more important in the context of selecting word-formation rules for course material in second language acquisition than it is for the study of word-formation as a component of linguistic competence.

It is interesting to compare, for instance, the negative prefixes *non-*, *un-*, and *in-*. The first two are available for new formations, but *non-* is less common than *in-*. This suggests that *in-* is more profitable than *non-*. However, *in-* is also less regular, as examples such as *invaluable* illustrate. Therefore, learners should be careful to use *in-* in production if they do not know what the resulting word means. For *un-*, it can more safely be assumed that any occurrence of the resulting word has the expected meaning, but learners cannot always assume that it exists (cf. **unconceivable*). *Non-* is a safer prefix to use, but it has a specialized meaning, transforming any property into a binary one (cf. *non-medieval*).

3.3. Internal structure of mental lexicon

The third category of use of word-formation in second language acquisition is in structuring the lexicon. These effects of the knowledge of word-formation rules are more difficult to measure than the effects on comprehension and production, because they are not directly reflected in performance. It is the structure of the mental lexicon as part of competence which is targeted.

The central idea is that a learner's mental lexicon with more structure is of a higher quality in at least three respects. First, it is easier to recall a particular lexical entry. Secondly, it is easier to retain entries. Thirdly, it is easier to learn additional entries. Word association studies point to the tentative conclusion that beginning learners' mental lexicons rely heavily on their first language lexicons and their word associations include more form-based associations, based on the phonology of the stimulus word, than those of native speakers. Even advanced learners do not seem to approach native speakers in terms of word association responses. Based on this kind of data, the structure of second language mental lexicons looks considerably more complex than those of a first language (Fitzpatrick 2009).

Structure in the relevant sense is more general than only the morphologically based word family. Purely semantic relations (e.g., antonym pairs), general and conceptual word associations, or even arbitrary mnemonics that create a link between a learner's first language and a word in their second language (e.g., Polish *drogi* 'expensive' starts with the consonants of *dear*) can create structure as well. However, word families based on word-formation rules are a very powerful means to enhance the internal structure of the lexicon, especially for more advanced learners.

3.4. Strategies for learning word-formation rules

The most natural way for word-formation rules to become part of a speaker's competence is through the reanalysis of individual formations. Examples of such a development can be found in the many word-formation rules in English that are based on borrowings from French. Mühleisen (2010) gives a detailed historical overview of the emergence and development of the suffix *-ee*. This suffix first appeared in English as a result of the reanalysis of a large number of borrowings from the time when French was replaced by English as the language of legal proceedings in England. Subsequently it became more generally productive.

In second language acquisition, this reanalysis can be supported in at least three ways. As an example, we will consider the nominalization of adjectives as in *happiness*. One type of support consists of providing analysed examples of output of a word-formation rule, i.e. making it explicit that *happiness* has the structure [[happy] ness]. A second type is to consider the rule as a class of items formed by it. In this case this would be a large set of adjective-noun pairs where the noun ends in *-ness*. Finally, the rule can be formulated explicitly. The first two of these are more targeted versions of what is available to native speakers in their language acquisition process, whereas the last one is based on explicit teaching (cf. also section 3.1 above).

Very little is known about transfer of word-formation concepts. It seems plausible or even logical that this should occur, at least where related languages are concerned, cf. Whitley (2004). There is evidence that procedural transfer relating to bound morphemes is possible (Jarvis and Odlin 2001). At the same time, a certain reluctance to assume that the L2 is similar to the L1 can be observed, a fear of false friends. Contrastive studies in word-formation among Slavic languages are more common and often oriented towards supporting second language acquisition, for instance, by identifying false friends. Examples of such studies are Vlček (1966), Lotko (1987), and Vaňko (2003, 2004).

3.5. Word-formation in learner's dictionaries

Learner's dictionaries are dictionaries for learners who are sufficiently advanced to use a monolingual dictionary of the language they are learning. This is an interesting target group for the study of the role of word-formation in second language acquisition, because these learners are at the level at which expanding and structuring the lexicon becomes one of the central learning aims. For a more general perspective on the issues involved in the representation of word-formation in dictionaries, cf. article 131 on dictionaries.

In paper dictionaries, the most common approach is the one exemplified by the *Longman Dictionary of Contemporary English* (LDOCE 2003). Here *happiness* is an entry of its own and its definition includes the word *happy*. Although not fully explicit, this means that the learner is invited to analyse this word as formed by the suffixation of *-ness*. In addition, *-ness* has an entry of its own, so that the learner can verify a hypothesis along these lines.

A different approach is taken in the *Dictionnaire du Français Contemporain* (DFC 1971). Although not using the term *learners' dictionary*, the preface states (1971: v) that it is aimed at "élèves de l'enseignement secondaire et aux étudiants étrangers" [second-

ary school pupils and foreign students]. In this dictionary, the basic unit of description is the word family. The entry for *approuver* 'to approve' includes as run-on entries, for instance, *approbateur* 'approving' or 'flatterer', as well as *désapprouver* 'to disapprove'. An obvious disadvantage of this approach is that it is more difficult to find words that are not described as the headword of a word family. However, DFC counters this by a generous amount of cross-references. Moreover, a prefix such as *dés-* is treated in a box in its alphabetical position (in this case the user is referred to "dé-, dés-") and the preface contains a systematic overview of suffixes and prefixes (1971: xix−xxii).

The main advantage of the DFC's approach is that it encourages the user to discover word families. Without making the structure of *approbateur* and *désapprouver* explicit, their position in the entry of *approuver* states clearly that there is a relationship. A mere mention of *approuver* in the definition of *approbateur*, as we saw for *happiness*, would not have the same effect here, because the form is quite different. The treatment of word-formation rules corresponds to the one we saw for *-ness*. The boxed entry for the prefix *dé(s)-* is arguably more eye-catching than the entry for *-ness*, but suffixes do not have the same treatment. The only place where *-ateur* is listed is in the preface. Given the general problems of space restrictions in paper dictionaries, this is an understandable compromise. In an alphabetical listing, connections with prefixed words are more difficult to find than with suffixed words.

Another problem with the use of word families as a basis for the macrostructure of a dictionary is the representation of compounds. Compounding combines two words, so this raises the question which of them is chosen as the place to treat the complex word. In French, this problem is arguably less pressing than in many other languages, because compounding is not such a prominent word-formation process. However, it is remarkable that DFC (1971) gives *chasse-neige* 'snowplough; lit. drive_away-snow' under *chasser* 'hunt', *tire-bouchon* 'corkscrew; lit. draw-cork' under *boucher* 'to block' and *pique-assiette* 'sponger; lit. pick-plate' is an entry of its own.

Whereas the idea of using word families as a basis is certainly conducive to learning word-formation, its implementation in a paper dictionary reduces the possibilities of taking optimal advantage of it. The only way to improve on the concept of the DFC is by making the dictionary electronic.

3.6. Electronic tools for learning word-formation rules

Electronic dictionaries have important advantages over paper dictionaries. One advantage is the possibility of text search, which, as van Sterkenburg (2003: 137) points out, makes the user less dependent on the alphabetical ordering. Another advantage is the lack of space restrictions. Thus, De Caluwe and Taeldeman (2003) argue that electronic dictionaries offer so much additional space that it is worth including many regular word-formations that would not have been given an entry in a paper dictionary. Nesi (1999) mentions that the most prominent advantage for users of learner's dictionaries is the shorter search time. As learners use the dictionary much more frequently than native speakers do, this is an important consideration.

In order to use the possibilities offered by electronic dictionaries fully, it is essential to adapt the macrostructure and the microstructure of the dictionary to the mode of presentation. In this respect, the DFC's (1971) concept of using the word family as the basis for the entry is a good idea, because the main problems it encounters are caused by the linear format of the paper dictionary.

An attempt to represent word-formation as a structuring device of the lexicon has been undertaken in the context of Word Manager (WM), a system for electronic lexical databases on a morphological basis. Ten Hacken (2009) gives a general overview of WM and references to further literature. The central notion in WM is the *lexeme*, taken in the sense of Matthews (1974) as an inflectional paradigm with a citation form as a name. In a WM database, IRules encode inflection and WFRules word-formation. IRules represent inflectional classes, i.e. classes of lexemes that are inflected in the same way. IRules can describe the inflectional paradigm or, for uninflected words such as *under*, specify that there is only one form. Each lexeme belongs to an inflection class, but only simplex lexemes, e.g., *settle* or *car*, are assigned to an IRule directly.

The treatment of complex words in word manager uses the word-formation system of a language to uncover the structure of the lexicon. In the development of the WM database, each complex lexeme is assigned to a WFRule. The WFRule combines the components of the complex lexeme and determines the IRule it belongs to. Thus *settlement* is assigned to the WFRule that produces nouns by attaching a suffix to a verb. The issue of determining the right IRule is particularly relevant in languages such as Italian and German, which have different inflection classes for nouns and verbs.

WM databases can be used to retrieve different types of information about word-formation. First, it can give the analysis of a particular word, e.g., *settlement* as a combination of the verb *settle* and the suffix *-ment*. Then, it can return the class of items that is formed from the same base, including, e.g., *settler*, *unsettle*. This is the word family as mentioned in section 2.3. The class of items can also be delimited in different ways, e.g., all deverbal nouns with *-ment*, which includes also *engagement*, *development*, etc. The class can even be broadened to include all deverbal nouns, also, for instance, *organization* and *refusal*. Finally, the actual word-formation rule can be retrieved.

Ten Hacken and Tschichold (2001) analyse the general potential of WM in enhancing second language vocabulary learning. As they note, the different perspectives on word-formation described in the previous paragraph correspond quite closely to the possibilities offered in the DFC, but without the drawbacks caused by the DFC's presentation as a book. A very useful device they describe is the so-called *tree browser*. The WM tree browser shows the size of classes in a WM database. The user can select the criteria that determine these classes. In this way, the number of lexemes formed by competing word-formation processes, e.g., English *-ment*, *-ation*, and *-al*, can be retrieved, giving a sense of their relative profitability. It is also possible to list all lexemes in a class, for instance in order to determine the regularity of the formations. As semantics is not covered in WM, checking semantic regularity has to be done manually.

WM databases are not meant to be used immediately by end users. They are rather conceived as reusable resources that can be used in the development of dedicated software. Ten Hacken, Abel and Knapp (2006) describe how this has been done in ELDIT. ELDIT is an electronic lexical resource for learners of Italian and German. It can serve at the same time as an electronic dictionary and as a vocabulary trainer. The word-formation component of ELDIT was developed by semi-automatic extraction of the rele-

vant morphological knowledge from the Italian and German WM databases. In this way, word families are represented from the starting point selected by the user, and word-formation rules can be represented as classes of entries at different levels of generality. The rules can also be used to analyse words that are not in the dictionary.

In conclusion, the potential of word-formation rules in supporting second language vocabulary acquisition can only be exploited in an electronic tool. The design of such a tool requires a considerable degree of rethinking of the organization of information compared to a paper dictionary.

4. Types of word-formation processes

Word-formation is generally divided into derivation and compounding. However, a more fine-grained classification is necessary to draw any generalizations as to the type of problems they pose for the second language learner.

In the case of prefixation, the prefix is prominent because it is the first element. This makes it generally easier to recognize. In a pair such as *happy* − *unhappy*, the prefix is also invariable and has a clearly identifiable, basic meaning. Other types of derivation complicate this picture in different ways.

One type of complication is the prominence of the affix. A suffix is perhaps less prominent, but it is still easy to recognize the Dutch pattern in (1):

(1) a.   eten − eetbaar   ‘to eat’ − ‘edible’
     b.   drinken − drinkbaar   ‘to drink’ − ‘drinkable’
     c.   lopen − loopbaar   ‘to walk’ − ‘walkable’

It is more difficult if the form of the affix is not stable (cf. *in-* with its variants *im-*, *il-*, *ir-*) or if it triggers changes in the stem (e.g., *strong* − *strength*). A further complication arises when affixation in the second language is of a type not attested in the first language. An example is the Dutch pattern in (2):

(2) a.   berg − gebergte   ‘mountain’ − ‘mountain range’
     b.   deel − gedeelte   ‘portion’ − ‘section’
     c.   vogel − gevogelte   ‘bird’ − ‘fowl’

Dutch has a circumfix in (2), as can be shown by the ungrammaticality of forms such as **geberg* and **bergte*. Cross-linguistically, this type of affixation is quite rare, so that a learner is less likely to expect it. A similar effect can arise in infixation. The most difficult cases are no doubt the ones where there is no affix at all, either conversion (e.g., $book_N$ − $book_V$) or changes in the stem (e.g., *sing* − *song*, *extend* − *extent*). It is well known, for instance that Chinese learners of English tend to overgeneralize conversion.

Perpendicular to the difficulty of recognizing the word-formation process on a formal basis, there is a scale of semantic transparency. Dutch suffixation with *-baar*, as illustrated in (1), has a fairly basic meaning. For native speakers of English (and many other languages), it also helps that their native language has a similar word-formation process. Even without the latter, it is not difficult to recognize the Polish pattern in (3):

(3)   a.   autobus − autobusowy   'bus$_N$' − 'bus$_A$'
      b.   owoc − owocowy   'fruit$_N$' − 'fruit$_A$'
      c.   naród − narodowy   'nation$_N$' − 'national$_A$'

The adjectives formed with *-owy* are used where English often has a compound, e.g.,
dworzec autobusowy 'bus station'. Although relational adjectives are much less common
in English and there is no parallel process for the cases in (3), it is not difficult to
recognize the word-formation process. Much more difficult in this respect are cases such
as (2). For speakers of Dutch there is an intuitive connection between the meaning
differences that correlate with the circumfix *ge-...-te*, but it is difficult to describe and
de Haas and Trommelen (1993: 256) do not give a unified semantic description. A well-
known stumbling block for learners of Slavic languages is the verbal prefixes that, al-
though not randomly, may have a very broad range of often overlapping meanings. As
an example, (4) gives some Polish cases:

(4)   a.   pisać − przepisać   'to write' − 'to transcribe, to copy'
      b.   kazać − przekazać   'to order' − 'to transfer'
      c.   siadać − przesiadać   'to sit' − 'to change (train/bus)'
      d.   prowadzić − przeprowadzić   'to carry out$_{IMPF}$' − 'to carry out$_{PERF}$'

The basic meaning in (4a) is often extended in various ways as illustrated in (4b−c). As
a side effect, prefixes often (but not always) make the verb perfective. (4d) is an example
where this is the only effect of the prefix.

In the case of compounding, there are three main problems in the context of language
acquisition. The first is a general, monolingual problem with the interpretation of com-
pounds. Compounds only specify two components, without making it explicit how they
are related to each other. It is well known that for the correct interpretation of contrasts
such as *watermill* and *papermill*, it is necessary to apply world knowledge. This is not
a big problem in second language acquisition, because second language learners have
reached an adult cognitive development stage. The only compounds for which such
problems arise are those whose cultural references are not known by the learner.

The other problems are related to the contrast between the learner's first and second
languages. One problem is that compounds in the second language may take a different
image. An example is German *Eisenbahn* 'railway; lit. iron road'. The other problem is
that languages use compounds to different extent. German and Dutch are known for
their extensive use of compounds, but any translator will know that apart from contexts
where a compound in these languages that corresponds to a more descriptive phrase in
English, there are also contexts in which an English compound has to be taken apart in
Dutch or German. Some examples are given in (5) and (6):

(5)   G.   Diebesbande   'thieves' pack'
      E.   pack of thieves

(6)   E.   hand car wash
      G.   Autowäsche von Hand   'car wash by hand'

Problems of interference in compounding only occur in production and they are not
visible until a high level of proficiency has been reached.

5. Conclusion

The central concept at the interface between word-formation and second language acquisition is the word family. Although it is clear that knowledge of word-formation is beneficial in second language acquisition, it is much less clear how this potential can be activated in practice. Moreover, word-formation is not a priority issue in the study of second language acquisition. Compared to inflection, it can be seen as an optional extension of linguistic knowledge, so that it usually only comes into play for advanced learners.

In the study of word-formation, data from second language acquisition are difficult to use because of their idiosyncratic nature. Each learner is different in so many ways that it is hardly possible to collect a homogeneous group of sufficient size. The problem is that for word-formation it is not only the first (native) language and the second language that is being learned that are relevant, but also a plethora of other factors, including the degree of literacy, other languages known, etc. Therefore, a lot of observations tend to remain anecdotal.

6. References

Allerton, David J.
  2002   C. K. Ogden's "Basic English": A critical assessment. In: David J. Allerton, Paul Skandera and Cornelia Tschichold (eds.), *Perspectives on English as a World Language*, 149−158. Basel: Schwabe.
Bauer, Laurie and Paul Nation
  1993   Word families. *International Journal of Lexicography* 6(4): 253−279.
Brezina, Vaclav and Dana Gablasova
  2013   Is there a core general vocabulary? Introducing the *New General Service List. Applied Linguistics* 34(1): 1−123.
Browne, Charles
  2014   A new general service list: The better mousetrap we've been looking for? *Vocabulary learning and instruction* 3(2): 1−10.
Brumfit, Christopher J. and Keith Johnson (eds.)
  1979   *The Communicative Approach to Language Teaching.* Oxford: Oxford University Press.
Chomsky, Noam
  1959   Review of Skinner, B. F. (1957), *Verbal Behavior*, Englewood Cliffs, NJ: Prentice-Hall. *Language* 35: 26−58.
Chomsky, Noam
  1964   *Current Issues in Linguistic Theory.* Den Haag: Mouton.
Corbin, Danielle
  1987   *Morphologie dérivationnelle et structuration du lexique.* 2 Vol. Tübingen: Niemeyer.
Coxhead, Averil
  2000   A new academic word list. *TESOL Quarterly* 34: 213−238.
Davies, Mark
  2006   *A Frequency Dictionary of Spanish. Core Vocabulary for Learners*. New York: Routledge.
De Caluwe, Johan and Johan Taeldeman
  2003   Morphology in dictionaries. In: Piet van Sterkenburg (ed.), *A Practical Guide to Lexicography*, 114−126. Amsterdam/Philadelphia: Benjamins.

DFC
   1971   *Dictionnaire du Français Contemporain*. Ed. by Jean Dubois, René Lagane, Georges
           Niobey, Didier Casalis, Jacqueline Casalis and Henri Meschonnic. Paris: Larousse.
Fitzpatrick, Tess
   2009   Word association profiles in a first and second language: Puzzles and problems. In: Tess
           Fitzpatrick and Andy Barfield (eds.), *Lexical Processing in Second Language Learners*,
           38−52. Bristol: Multilingual Matters.
Flynn, Suzanne
   1996   A parameter-setting approach to second language acquisition. In: William C. Ritchie
           and Tej K. Bhatia (eds.), *Handbook of Second Language Acquisition*, 121−158. San
           Diego: Academic Press.
Fries, Charles C.
   1945   *Teaching and Learning English as a Foreign Language*. Ann Arbor: University of Mi-
           chigan Press.
Gardner, Dee
   2007   Validating the construct of word in applied corpus-based vocabulary research: A critical
           survey. *Applied Linguistics* 28(2): 241−265.
Gass, Susan M. and Alison Mackey
   2012   *The Routledge Handbook of Second Language Acquisition*. London: Routledge.
Gass, Susan M. and Larry Selinker
   2008   *Second Language Acquisition. An Introductory Course.* Mahwah, NJ/London: Erlbaum.
Gougenheim, Georges
   1958   *Dictionnaire Fondamental de la Langue Française*. Paris: Didier.
Gougenheim, Georges, René Michéa, Paul Rivenc and Aurélien Sauvageot
   1964   *L'Élaboration du Français Fondamental (1er degré). Étude sur l'établissement d'un
           vocabulaire et d'une grammaire de base*. Paris: Didier.
de Haas, Wim and Mieke Trommelen
   1993   *Morfologisch Handboek van het Nederlands. Een overzicht van de woordvorming*.
           's-Gravenhage: SDU.
ten Hacken, Pius
   2009   Word manager. In: Cerstin Mahlow and Michael Piotrowski (eds.), *State of the Art in
           Computational Morphology*, 88−107. Berlin: Springer.
ten Hacken, Pius, Andrea Abel and Judith Knapp
   2006   Word formation in an electronic learners' dictionary. *International Journal of Lexicogra-
           phy* 19: 243−256.
ten Hacken, Pius and Cornelia Tschichold
   2001   Word manager and CALL: Structured access to the lexicon as a tool for enriching
           learners' vocabulary. *ReCALL* 13: 121−131.
Halliday, Michael A. K., Angus McIntosh and Peter Strevens
   1964   *The Linguistic Sciences and Language Teaching*. London: Longman.
Han, ZhaoHong
   2012   Fossilization − A classic concern of SLA research. In: Susan Gass and Alison Mackey
           (eds.), *The Routledge Handbook of Second Language Acquisition*, 476−490. London:
           Routledge.
Huckin, Thomas, Margot Haynes and James Coady
   1993   *Second Language Reading and Vocabulary Learning*. Norwood, NJ: Ablex.
James, Carl
   1998   *Errors in Language Learning and Use. Exploring Error Analysis.* London: Longman.
Jarvis, Scott and Terence Odlin
   2001   Morphological type, spatial reference, and language transfer. *Studies in Second Lan-
           guage Acquisition* 22(4): 535−556.

Jeon, Eun Hee

2011   Contribution of morphological awareness to second-language reading comprehension. *The Modern Language Journal* 95(2): 217−235.

Jiang, Nan

2000   Lexical representation and development in a second language. *Applied Linguistics* 21(1): 47−77.

Laufer, Batia

1997   What's in a word that makes it hard or easy: Some internal factors that affect the learning of words. In: Norbert Schmitt and Michael McCarthy (eds.), *Vocabulary. Description Acquisition and Pedagogy*, 140−155. Cambridge: Cambridge University Press.

LDOCE

1978   *Longman Dictionary of Contemporary English*. Ed. by Paul Procter. Harlow: Longman.

LDOCE

2003   *Longman Dictionary of Contemporary English*. Ed. by Stephen Bullon. Harlow: Pearson.

Littlewood, William

1981   *Communicative Language Teaching. An Introduction*. Cambridge: Cambridge University Press.

Lotko, Edvard

1987   *Zrádná slova v polštině a češtině*. Olomouc: Univerzita Palackého.

Matthews, Peter H.

1974   *Morphology. An Introduction to the Theory of Word Structure*. Cambridge: Cambridge University Press.

Mochizuki, Masamichi and Kazumi Aizawa

2000   An affix acquisition order for EFL learners: An exploratory study. *System* 28: 291−304.

Mühleisen, Susanne

2010   *Heterogeneity in Word-Formation Patterns. A Corpus-Based Analysis of Suffixation with -ee and its Productivity in English*. Amsterdam/Philadelphia: Benjamins.

Nagy, William E. and Richard C. Anderson

1984   How many words are there in printed school English? *Reading Research Quarterly* 19(3): 304−330.

Nation, I. S. Paul

2001   *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nesi, Hilary

1999   A users' guide to electronic dictionaries for language users. *International Journal of Lexicography* 12: 55−66.

Ogden, Charles K.

1930   *Basic English. A General Introduction*. London: Paul.

Olshtain, Elite

1987   The acquisition of new word formation processes in second language acquisition. *Studies in Second Language Acquisition* 9: 221−232.

Pons-Ridler, Suzanne and Fred McKim

1985   Etude du vocabulaire fondamental par la dérivation. *The French Review* 58(5): 633−649.

Ritchie, William C. and Tej K. Bhatia (eds.)

1996   *Handbook of Second Language Acquisition*. San Diego: Academic Press.

Schmitt, Norbert and Paul Meara

1997   Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition* 20: 17−36.

Schmitt, Norbert and Cheryl Boyd Zimmerman

2002   Derivative word forms: What do learners know? *TESOL Quarterly* 36(2): 145−171.

Selinker, Larry
   1972   Interlanguage. *International Review of Applied Linguistics* 10: 209−231.
Selinker, Larry
   1992   *Rediscovering Interlanguage*. Harlow: Routledge.
van Sterkenburg, Piet
   2003   Onomasiological specifications and a concise history of onomasiological dictionaries.
            In: Piet van Sterkenburg (ed.), *A Practical Guide to Lexicography*, 127−143. Amster-
            dam/Philadelphia: Benjamins.
Vaňko, Juraj
   2003   *Slovaško-slovenska homonimija. Slovar slovaško-slovenskih medjezikovnih homonimov.*
            Ljubljana: Filozofska fakulteta, Oddelek za slavistika.
Vaňko, Juraj
   2004   Slovinsko-slovenská interlingválná homonymia. *Slavia* 73(1): 59−70.
Vlček, Josef
   1966   *Úskalí ruské slovní zásoby. Slovník česko-ruské homonymie a paronymie.* Praha: Svět
            Sovětů.
West, Michael
   1953   *A General Service List of English Words*. London: Longman.
White, Lydia
   2003   *Second Language Acquisition and Universal Grammar.* Cambridge: Cambridge Univer-
            sity Press.
Whitley, Stanley
   2004   Lexical errors and the acquisition of derivational morphology in Spanish. *Hispania*
            87(1): 163−172.

Cornelia Tschichold, Swansea (Great Britain)
Pius ten Hacken, Innsbruck (Austria)

122. Word-formation in aphasia