

---

# Neoclassical Formatives in Dictionaries

Pius ten Hacken, Renáta Panocová

Leopold-Franzens-Universität Innsbruck, P.J. Šafárik University Košice

Pius.ten-Hacken@uibk.ac.at, Renata.Panocova@upjs.sk

## Abstract

Neoclassical formatives are elements that occur in neoclassical word formations such as *ortho* and *paed(o)* in *orthopaedic*. We can be sure that a separate system of neoclassical word formation is in place when we find words such as *orthopaedic* that use classical elements but cannot have been borrowed from a classical language, because the concept they refer to did not exist yet. However, it must be taken into account that such words can also have been borrowed from a modern language that has such a system. In the lexicographic treatment of neoclassical word formation, a central question is whether neoclassical formatives should be treated in separate entries. We investigated how different English and Russian dictionaries treat them. In order to arrive at an unbiased sample of formatives, we used a Catalan word formation dictionary.

The results of our investigation support the hypothesis that neoclassical word formation constitutes a separate system in English, but not in Russian. This means that in English neoclassical formatives should have entries. In electronic dictionaries they can usefully be connected to the full class of words they appear in. In Russian, neoclassical formations are borrowings from languages such as English or French and their internal structure belongs to the domain of etymology.

**Keywords:** neoclassical word formation; neoclassical formatives; combining forms; English; Russian

In this paper we investigate the optimal treatment of neoclassical word formation in dictionaries. The main problem of neoclassical word formation in lexicography is to determine how to treat the internal structure and the components that are not words by themselves. First, we give a general introduction to the phenomenon of neoclassical word formation (section 1) and present some considerations as to the treatment of word formation in dictionaries (section 2). In order to investigate the coverage of neoclassical word formation in dictionaries, we present a sampling method that is not directly dependent on the languages and dictionaries under investigation (section 3). Then we compare the realization of neoclassical word formation and its treatment in some standard general dictionaries in two languages, English (section 4) and Russian (section 5). These languages were chosen because, as we will argue, the status of neoclassical word formation in them is different in interesting ways. On the basis of our observations in the preceding sections, we conclude in section 6 that there are good arguments for treating neoclassical formatives in separate entries in English, but not in Russian.

## 1 Neoclassical Word Formation

The discussion of neoclassical word formation has a long history. In many cases, it is connected to the opposition between learned and non-learned word formation. Bloomfield (1933: 153-54) discusses the opposition between unmarked and learned forms in relation to the way the lexicon of a language is organized and gives examples for various languages. He discusses the phenomenon of learned language mainly as a matter of register. Here, however, we will be interested only in the specific type of word formation involving Greek and Latin stems that is found in a wide range of European languages. An example from English is *orthopaedic*.

An issue that arises immediately in the context of neoclassical word formation is the boundary between word formation and borrowing. This issue arises in two shapes. The first can be illustrated on the basis of the contrast between *orthogonal* and *orthopaedic*. Both of these are based on Ancient Greek, but not in the same way. In Ancient Greek, the word ὀρθόγωνος ([*orthógonos*] ‘rectangular’) is attested and can be analysed as formed from ὀρθός ([*orthós*] ‘straight’) and γωνία ([*gonía*] ‘angle, corner’). Therefore, it is possible to classify *orthogonal* in English as a borrowing from Ancient Greek. In the case of *orthopaedic*, there is no corresponding word in Ancient Greek. However, in Ancient Greek we have the words ὀρθός ([*orthós*] ‘straight’) and παῖς (stem παιδ- [*paid-*], ‘child’) which can be analysed as the basis for *orthopaedic*. This is the first realization of the issue of borrowing.

However, there is a second form in which this question arises. In fact, the non-existence of a corresponding Ancient Greek word does not necessarily show that the word was formed in English. In this case, it is possible to trace the origin to French. The French physician Nicolas Andry de Boisregard (1658-1742) created the term *orthopédique* in 1741. From French it was subsequently adopted in other European languages. In this case, it is possible to trace the formation of the word to a single person and a single publication, because the concept was invented by this person. In many cases it is much more difficult to trace the origin so precisely. If a concept is ‘in the air’ and arises in a context in which different people might think of the same name independently, or spread it very rapidly once someone starts using it, it is almost impossible to reconstruct which language is at the origin of a particular word.

In order to determine the significance in lexicography of the question of borrowing or formation in any particular case, we can choose between two perspectives. One is to aim for a description of the historical development of language use. In this perspective it is important to distinguish the origin of *orthogonal* and *orthopaedic*, stating that the former is borrowed from Ancient Greek and the latter from French where it was constructed from Ancient Greek components. The second possible perspective concentrates rather on the vocabulary as known by the speakers. Here the question is whether these words are structured or not. Most speakers will not be aware of the Ancient Greek origin of *orthogonal* or the French origin of *orthopaedic*. The difference between words formed in Ancient Greek and those formed in modern languages is not one that determines the structure of the mental lexicon of a contemporary speaker of English.

Following ten Hacken (2012), we will assume that in the case of genuine neoclassical word formation, there must be a system of neoclassical formatives in the lexicon. Such a system can emerge as the result of the reanalysis of borrowings. After a sufficient number of words with similar components have been borrowed, speakers may notice the regularity. In the first instance, both a new system and a new set of formatives have to be set up. It is quite likely that *ortho-* belonged to the initial set of formatives, because there are quite a number of old loanwords from Ancient Greek that contain it, e.g. *orthogonal*, *orthodox*, *orthography*. Once the system exists, it is easier to extend it. A single item may be sufficient. Thus, for *paedo-*, the origin may be a word such as *paedagogical*. We even find cases in which a neoclassical formative is borrowed directly from Ancient Greek, e.g. *psepho-* in *psephological*. This was not a formative in English when the word *psephological* was created in the 1950s. The first attestation in the OED is from 1952.

Systems of neoclassical word formation have come into existence in a number of languages, probably in the course of the 18th century. In fact, *orthopédique* is a very early example of a neoclassical formation that cannot have been borrowed directly from Ancient Greek. A much larger set of such words start appearing in the 19th century. Because languages such as French, English, and German all went through the same process, it is often difficult to determine which of these languages is at the origin of a particular word. However, the question of which modern language is at the origin of a particular formation is only relevant if we want to describe the history of a individual words. Speakers of such languages who have the neoclassical subsystem of word formation in their mental lexicon will analyse the relevant words independently of whether the original source is in their own language or in another one. Therefore, it is less relevant which of these languages is the historical origin of a particular word. Once it exists in one language it is quickly adopted in the others.

This is especially the case for words such as *orthopaedic*, because they belong to a very specialized register of language. Neoclassical words are part of the language of science. Not many speakers of English will have known them in the 18th or 19th century, but those who did are more likely to be in contact with and competent in other languages, such as French and German. As languages such as English, French and German only exist as entities when they are rational constructions associated with a conscious aim for codification, the decision to attribute the origin of a particular word to one of them is ultimately arbitrary.

## 2 Word Formation in Dictionaries

As argued in ten Hacken (2009), dictionaries do not describe the vocabulary of a language, but provide information about words to dictionary users. The aim of a description is not realistic because there is no suitable empirical object to describe. What we have is a number of speakers and a set of texts and utterances. The idea of a language to be described in a dictionary requires the classification of the spe-

akers and texts/utterances and is in this sense a constructed entity, not one empirically found as an entity.

The role of word formation in dictionaries is significantly enhanced by the insight that dictionaries are no more or less than sources of information for their users. Ulsamer (2013) gives an overview of current practice and insights as to the representation of word formation in dictionaries of various types. Booij's (2003: 254) suggestion that the place of word formation in dictionaries is mainly to state "if a possible morphologically complex word actually exists" is unnecessarily restrictive. First of all, Booij's wording suggests that there is an independent sense in which a word should exist. In the same way as for languages, there is no empirical entity corresponding to the word. What the lexicographer has is a number of speakers and a corpus of texts and utterances. In addition, Booij's suggestion reduces the role the representation of word formation can play in supporting the dictionary user.

On the basis of the available corpus and linguistic knowledge, for each morphologically complex word, lexicographers have to take decisions such as the following:

- whether a particular word should be represented in the dictionary,
- whether it should be represented as morphologically complex,
- which information should be given about the word, and
- how this information should be presented.

It is crucial to understand that these are decisions based on judgements. The available data in a corpus can serve to support the judgement, but they do not result in a decision in an algorithmic sense. There is also no objective truth to be discovered as to the existence or the nature of the words considered. Svensén (2009: 131-133) sketches some of the traditional techniques that have been used, including the so-called *run-on* entry that Booij seems to allude to.

In electronic dictionaries, the possibilities for representing and accessing information are greatly enhanced compared to paper dictionaries. However, it is by now well-known that these additional possibilities depend on an appropriate encoding of the information. It is not sufficient to present information in the way it is encoded in a print dictionary and just make it available in electronic form. This argument was made quite systematically by ten Hacken (1998) and has now become commonplace. Ulsamer (2013: 35-50) can refer to a number of dictionaries and lexical resources that were developed with this idea in mind and exploit the specific strengths of the electronic medium.

Ten Hacken (1998) distinguishes three types of lexicographic representation of word formation. One type focuses on the word formation analysis of individual words. If we take as an example *happiness*, this means that the information that this word was formed by adding to the suffix *-ness* to the adjective *happy* is given in the dictionary entry for *happiness*. Giving this information is not as commonplace in paper dictionaries as one might expect. Most general dictionaries only give a run-on entry and in learners' dictionaries we often find a separate entry without this information. In both cases, this

leaves it to the user to work out the structure of the word. In electronic dictionaries, there is of course no excuse based on space limitations to motivate such a decision.

A second type of information that dictionaries can give about word formation is the grouping of words that belong to the same word formation class. Here a word formation class is interpreted as a set of words that have some word formation properties in common. Ideally, it would be up to the user to choose which properties are most relevant for their question. In the case of *happiness*, the most obvious class would be that of nouns formed by suffixation of *-ness* to an adjective. This is a very large class and in a paper dictionary it is not normally presented, but in an electronic dictionary it is feasible to do so. Most users will only want to consider some examples, but linguistic researchers (a minority user group, but heavy dictionary users) may well be interested in the entire set. Other word formation classes that *happiness* is a member of include the class of words formed with *happy* as a basis, the class of nouns formed from adjectives, or the class of nouns formed by suffixation. Especially for the larger classes in these examples, any representation in an alphabetically ordered print dictionary will be too unwieldy, both to print it and to use it, but as Ulsamer (2013) shows, in electronic dictionaries there is an advantage in the proper representation of some of the possible classes.

The third type of representation concentrates on the rule as such. In the case of *happiness*, this may mean, for instance, an entry for *-ness* describing the word formation process. This is actually a common type of information also in print dictionaries. COED (2011) gives an entry for *-ness* with the different senses and an example for each sense. Atkins & Rundell (2008: 180) mention such entries as possible lemmata under the category of “partial words” and Svensén (2009: 132-133) presents it as one of the main representations of word formation in the dictionary. Of course, the representation in an electronic dictionary can be much improved by making a hyperlink available between the entries for the individual words (e.g. *happiness*) and the entry explaining the word formation rule, as well as by linking entries such as for *-ness* to the corresponding word formation class (i.e. all nouns formed with *-ness*).

The questions we want to address in relation to neoclassical word formation are then the following:

- which of these possible representations are used in current dictionaries?
- which improvements could be made in these representations, especially in electronic dictionaries? and
- to what extent do the choices depend on properties of different languages?

In order to provide a proper basis for the discussion of the last of these questions, we will consider two languages, English and Russian, in which it can be argued that the mechanism of neoclassical word formation does not work in the same way. First, however, we have to find a way to collect data for the first of these questions.

### 3 The Lexicographic Representation of Neoclassical Word Formation

In general, the decision how to treat a particular word formation process in a dictionary is determined at least in part by the productivity of the process. As shown by Bauer (2001), *productivity* is a notion that can be interpreted in different ways and has therefore raised a lot of discussion. A very useful analysis of the different notions of productivity is the one by Corbin (1987: 176-178). She distinguishes three aspects of productivity that in obvious cases will point in the same direction, but also give the conceptual vocabulary for a discussion about the choice of relevant properties in more controversial cases. The first is *régularité* ('regularity'). This is the extent to which the form and meaning of a particular formation can be predicted on the basis of the input (base) and the rule. The second is *disponibilité* ('availability'). This is a binary feature, indicating whether or not the rule is available for application to new bases. Finally, there is *rentabilité* ('profitability'). This is the degree to which the rule is applied to many new bases, yielding new formations. In her own work, Corbin (1987) concentrates only on availability. This is understandable because it is the underlying condition that has to be met in the linguistic competence before the other aspects can apply at all. However, in lexicography, the other two concepts are also relevant. Thus, regularity of new formations clearly influences to what extent run-on entries can be used and the profitability of a process will determine how important it is to treat it in the dictionary at all.

Much of the discussion of word formation in the context of lexicography concerns affixation. In the case of affixation it makes sense to consider each particular affix as a rule for which productivity (in its different aspects) can be calculated. In the case of *-ness*, the availability for new formations and the high degree of profitability make it a good choice for a separate entry. The regularity of many of the individual formations, such as *happiness*, makes it attractive to treat these as run-on entries.

If we want to extend this approach to productivity from affixation to neoclassical word formation, we encounter the problem that many neoclassical formations are more like compounding than like affixation. In the example of *orthopaedic*, there is a suffix *-ic*, but the central piece of the formation of this word is the combination of *ortho* with *paed(o)*. That neoclassical formatives such as *ortho* and *paedo* are not affixes is obvious from their distribution as well as from their contribution to the meaning of the resulting word. In lexicography, they are often called *combining forms*, e.g. by Svensén (2009:133). The variation in form is sometimes accounted for by different entries for the initial combining form and the final combining form. Thus, COED (2011) has different entries for *-phone* and *phono-*. This is not optimal for the insight that they represent the same formative, because there is no link from one entry to the other and a user not actively looking for the two variants will not find the connection, as they are separated by nine entries in the alphabetic order.

Returning now to productivity as a criterion to determine how to treat neoclassical word formation in a dictionary, we are faced by the situation that we have to decide for individual neoclassical formatives whether they deserve an entry and in which other ways they might be referred to. In this deci-

on, affixation is not a good parallel, because neoclassical formatives are not affixes, but compounding is not a good model either. In compounding, the lexicographer has to decide whether or not to devote an entry to the compound, but not whether the components of the compound should be treated. In the case of *apple juice*, the only question is whether it deserves an entry of its own, not whether *apple* and *juice* should have an entry. We do not attempt to determine the productivity of *apple* in compounding as the basis for any lexicographic decision. In the case of *ortho* and *paedo*, we have to consider the number of different formations they appear in and the frequency of those formations in some way to determine whether they deserve being treated in a separate entry.

In order to explore the way neoclassical formations are covered in English dictionaries, we selected a sample of neoclassical formations. In many discussions of neoclassical word formation, a very limited set of examples is discussed time and again. However, there is no indication to what extent these examples are representative of the phenomenon. We considered that a sample based on any specific English dictionary would be biased, in particular when we want to compare the coverage in English dictionaries with the coverage in dictionaries of another language. Therefore we used a source from a language that is not in the scope of the study, Bruguera i Talleda's (2006) Catalan dictionary of word formation.

There are several reasons why Bruguera i Talleda (2006) is a good source for our study of neoclassical word formation. First, Catalan has neoclassical word formation in a way similar to other European languages. As a Romance language, it is not biased to English or Russian and it has a more direct link to Latin than either of these languages, so that we can expect that the set of neoclassical formations tends to be larger. Secondly, the dictionary offers a type of access to neoclassical word formation that is convenient for our purposes. The lemmata of the dictionary are affixes and neoclassical formatives. The entries contain basic information about the use of the formatives and a full list of words formed with them. Where appropriate, entries are divided into different senses. In addition, the introduction (2006: 9-50) contains a detailed discussion of the different types of word formation and a classification of the lemmata. Therefore, it was easy to select a randomized sample of relevant formatives. Specialized dictionaries of this type are quite rare, in particular published as paper dictionaries.

Bruguera i Talleda (2006: 31-48) treats neoclassical word formation as *formació culta* ('learned formation'), which connects with Bloomfield's (1933) category of learned word formation as treated in section 1. Neoclassical formatives are listed in the section on neoclassical compounding. This is not ideal in principle, because neoclassical word formation is not restricted to compounding, as evidenced by formations such as *ethnic* or *morpheme*. However, compounding is much more prominent in neoclassical word formation than derivation, so it can be expected that in practice any formative that appears in derivation will also appear in compounding. Neoclassical word formation as it is used, for instance, in medicine is based on Greek formatives that in many cases passed through Latin. Bruguera i Talleda (2006: 38-47) gives separate tables for formatives of Greek and of Latin origin. We only considered the former. In addition, the formatives are divided into initial and final combining forms. There is a large degree of overlap between the two, but the former list is significantly longer. An additional practical

advantage of initial combining forms is that it is immediately evident when looking them up in an alphabetically ordered dictionary which and how many entries use them. Therefore we based our sample of neoclassical formatives on a random selection from the list of initial combining forms of Greek origin. The only adjustment we had to make is to adapt the spelling from Catalan to English and Russian.

## 4 Neoclassical Word Formation in English Dictionaries

For English, we took as our dictionaries CED (2000) and COED (2011). In Béjoint's (2000: 42-91) overview of dictionaries of English, it is obvious that British and American dictionaries follow different patterns. Therefore it would not be possible to generalize from one type of dictionary to the other. However, within British dictionaries, CED (2000) and COED (2011) fall into different subtypes. Béjoint (2000: 57-58) classifies the Oxford dictionaries as traditional, where the Collins dictionary belongs to an innovative trend that started in the 1970s. Both belong to the type van Sterkenburg (2003) identifies as 'the' dictionary, i.e. general-purpose dictionaries of a size big enough to give a fairly comprehensive overview of the vocabulary without giving a full scholarly account of its development. Therefore, especially when the findings of the two dictionaries coincide, we can safely draw conclusions for British dictionaries of this type in general.

The first sample we took was a randomized set of items from Bruguera i Talleda's (2006: 40-44) list of initial combining forms of Greek origin. We found that almost all of them had at least two examples of formations that were described in both English dictionaries. This means that the basic condition for identifying the item as a neoclassical formative would be fulfilled. However, only about a third of the formatives is described in separate entries. Thus, CED (2000) gives *thanatology* and *thanatopsis*, but no separate entry for *thanato-*. The structure of the words is only addressed in the section on etymology, where we find the following:

thanatology: [C19: from Greek *thanatos* death + -LOGY]

thanatopsis: [C19: from Greek *thanatos* death + *opsis* a view]

The difference in presentation suggests that *-logy* and *-opsis* are treated differently in the dictionary. In fact, both have an entry as a combining form. The difference between CED (2000) and COED (2011) is small. There are a few cases where COED (2011) gives the etymology and CED (2000) does not, e.g. *mammography*. Conversely, CED (2000) gives slightly more separate entries for combining forms. Thus, only CED (2000) has an entry for *noso-*. However, on the whole the two dictionaries have a remarkably similar treatment of the formatives in the sample.

We were not fully satisfied with our first sample, because for many of the formatives there were so few items that were listed in the dictionary that it was not clear whether lexicographic decisions concerning the inclusion of particular words or the lexicographic approach to neoclassical formati-

ves were responsible for the treatment we found. Thus, for *thanato-*, CED (2000) gives the two entries listed above, but COED (2011) gives only the first. Therefore, we took a second sample, which only included formatives that were relatively profitable in Corbin's (1987) sense.

For this second sample, we took as a selection criterion the length of the entry in Bruguera i Talleda (2006). As this dictionary gives all attested words with the relevant formative in Catalan, the length of an entry gives a measure of the profitability of the formative. From our first sample, we discarded each item for which the entry in Bruguera i Talleda (2006) is shorter than ten lines. Where the same formative occurs as an initial as well as final combining form, we combined the length of both entries. We replaced the discarded items with formatives that fulfilled this length criterion. The results of looking up this second sample in CED (2000) and COED (2011) showed that these formatives are generally more often described in a separate entry. In COED (2011) more than two thirds of the second sample appears as an entry and in CED (2000) we found almost all of them. This difference is in line with the tendency we observed in the first sample that CED (2000) gives more separate entries than COED (2011). As an example of an entry, COED (2011) gives the following for *diplo*:

**diplo-** ▶ **comb. form 1** double: *diplococcus*. **2** diploid: *diplotene*.

- ORIGIN from Gk *diplous* 'double'.

The corresponding entry from CED (2000) is as follows:

**diplo-** or before a vowel **diplo-** combining form. double: *diplococcus*. [from Greek, from *diploos*, from  $\text{DI}^{-1}$  + *-ploos* -fold]

The entries in the two dictionaries are very similar, but it is interesting to note the differences. Only CED gives the variant form. Only COED gives the second sense. This second sense is the result of a shortening of one of the complex forms the formative appears in. It is similar to the use of *gastro-* in the sense of 'gastronomic' rather than 'stomach'. In the etymological information, COED gives the classical form as it is likely to have been at the origin of the borrowing, whereas CED gives the oldest attested form in Greek as well as its word formation origin.

On the basis of the samples of neoclassical formatives we considered, we can conclude that British general dictionaries tend to include separate entries for neoclassical formatives when they appear in many different words. In the entries for these formatives, but also in the ones for neoclassical formations, the connection to the Ancient Greek forms is made explicit. There is a slight difference between CED (2000) and COED (2011) in the sense that the former has more separate entries for neoclassical formatives whereas the latter is more systematic in giving etymologies for neoclassical formations.

## 5 Neoclassical Word Formation in Russian Dictionaries

The position of neoclassical word formation in Russian is not directly comparable to that in English or other Germanic and Romance languages. Neoclassical forms are generally much rarer in Russian

and in many cases, a competing form with Slavic roots is preferred. Therefore, our hypothesis is that Russian dictionaries will give fewer neoclassical formations and will not analyse them as readily as English dictionaries.

In exploring this hypothesis, we used the dictionaries by Ušakov (1946-47) and Efremova (2000). They are roughly comparable in size to the dictionaries we used for English. In the Russian lexicographic tradition, the monolingual dictionary by Ušakov (1946-47) is categorized as a normative dictionary of contemporary standard Russian (Šanskij, 1972: 286; Ožegov, 1974: 171; Germanovič, 1979: 264). The dictionary includes nearly 90,000 entries. Germanovič (1979: 265) points out it is the first dictionary which gives separate entries for productive word formation elements such as prefixes or affixoids. The dictionary by Efremova (2000) is more recent and lists around 140,000 entries. The introductory material to the online version of the dictionary gives the information that prefixes, suffixes, initial elements of complex words and final elements of complex words are described in separate entries.

On the basis of this policy description, one might expect that the findings of checking our samples of formatives in Russian dictionaries would be closer to the results in CED (2000) and COED (2011) for English than originally thought. However, a closer look at the coverage of our first sample of formatives from Bruguera i Talleda (2006) in these two dictionaries confirmed our original hypothesis that only few neoclassical formatives are described in separate entries. The two dictionaries often differ in the treatment of the formatives in the sample. For instance, the dictionary by Ušakov (1946-47) gives этнолог ([etnolog] 'ethnologist'), этнологический ([etnologičeskij] 'ethnological'), этнология ([etnologija] 'ethnology') but does not have a separate entry for этно- ([etno-] 'ethno-'). A large part of the formatives in the first sample follow a similar pattern. For some neoclassical formations, Ušakov (1946-47) gives etymologies explaining the components of words as based on Ancient Greek.

ЭТНОГРАФИЯ, этнографии, мн. нет, ж. (от греч. ethnos - народ и grapho - описываю)

[etnografija, etnografii, mn. net, ž. (ot greč. ethnos - narod i grafo - opisivaju)]

'ethnography, ethnography<sub>GEN</sub>, no plur., fem. (from Greek ethnos - nation and grapho - writing)'

In principle, such etymological information would enable the user to add a system of neoclassical word formation to their mental lexicon, but there is no indication that many speakers of Russian do this. On the other hand, the dictionary by Efremova (2000) classifies the same formative as an initial part of complex words:

этно-

Начальная часть сложных слов, вносящая значение сл.: народ (этногенез, этнолингвистика, этнопсихология и т.п.).

[etno-]

[načalnaja čast' složnyx slov, vnosjaščaja značenie sl.: narod (etnogenenez, etnolingvistika, etnopsichologija i t.d.)]

'ethno-'

'initial part of complex words having the meaning nation (ethnogenesis, ethno-linguistics, ethnopsychology, etc.)'

Efremova (2000) does not give information about the etymology of neoclassical formatives, but the introductory material mentions that entries are included for around 900 combining forms. Of course not all combining forms are neoclassical formatives.

Checking our second sample of formatives from Bruguera i Talleda (2006) in these two dictionaries, we found that our initial expectations were largely confirmed again. Only for just over half of the formatives do the Russian dictionaries give any examples of formations, the number of formations per formative is lower than in English dictionaries, sometimes just one example, and only very rarely is the neoclassical formative described in a separate entry. Despite the generous coverage of combining forms announced in Efremova's (2000) introductory material, we only found very few separate entries for formatives that are part of our second sample. For *phago-*, Efremova (2000) gives the following entry:

фаго

Начальная часть сложных слов, вносящая значения: 1) поедание, пожирание чего-л. (фагоциты); 2) связанный с бактериофагом (фагодиагностика, фагопрофилактика).

[fago]

[Načal'naja časť složnych slov, vnosjaščaja značenija: 1) pojedanije, požiranije čego-l- (fagocity); 2) svjazannyj s bakteriofagom (fagodiagnostika, fagoprofilaktika).]

'phago'

'initial part of complex words having the meanings: 1) eating of something (phagocytes); 2) related to bacteriophage (phagodiagnosis, phagoprophylaxy)'

The second sense given in this entry is similar in nature to the second sense of *diplo* in section 4. It results from the shortening of a full neoclassical formation. Interestingly, Efremova (2000) lists separately also the formative *-phag* occurring in final position. The entry provides the following information:

фаг

Конечная часть сложных существительных, вносящая значение: поедающий, поглощающий то, что указывается в первой части слова (ихтиофаг, фитофаг и т.п.).

[fag]

[Konečnaja časť složnych suščestvitel'nych, vnosjaščaja značenie: pojedajuščij, pogloščajuščij to, čto ukazyvajetsja v pervoj časti slova (ichtiofag, fitofag i t.p.)]

'phag'

'final part of complex nouns having the meaning: eating, eating what is denoted by the initial part of the word (ichtiophag, phytophag, etc.)'

Ušakov (1946-47) does not give a separate entry for the initial neoclassical formative *phago*, but gives фагоцит ([fagocit] 'phagocyte'), and фагоцитоз ([fagocitoz] 'phagocytosis'). Similarly, the final neoclassical formative *-phag* is not given as an independent entry, but can only be traced, at least in principle, in entries such as фитофаг ([fitofag] 'phytophag').

An interesting case is represented by the formative *diplo*. The meaning corresponding to the examples from English in section 4 is found in the formations диплоккок ([diplokok] ‘diplococcus’), but the formative itself does not appear in a separate entry. In Efremova (2000), an initial part of complex words дип ([dip] ‘dip’), is treated as an independent listed item. The meaning is, however, related to дипломатический ([diplomatičeskij] ‘diplomatic’). It is a clipping found in formations such as дипкупе ([dipkupe] ‘diplomatic compartment’), дипкорпус ([dipkorpus] ‘diplomatic corpus’) or диппочта ([dippočta] ‘diplomatic mail’). This pattern of forming words is very productive in Russian and it is referred to as *stump compounds* or *abbreviated compounds* (cf. Benigni and Masini, 2009, Comrie and Stone, 1978, Molinsky, 1973).

On the basis of our sampling, we may conclude that in Russian monolingual general dictionaries the coverage of neoclassical word formation is much less extensive than in English. First of all there are far fewer entries for neoclassical formations. Moreover the way they are treated reflects to a much smaller extent a system of neoclassical word formation. Ušakov (1946-47) only gives etymologies for neoclassical formations, which is hardly sufficient to retrieve the use of neoclassical formatives. Efremova (2000) gives many entries for combining forms, but from our samples only few neoclassical formatives are actually covered in them.

## 6 Conclusion

In this paper we investigated the treatment of neoclassical formatives in English and Russian dictionaries. The basis of our research was a sample of initial combining forms. In order to exclude a direct bias towards English or Russian in our sampling, we used a word formation dictionary for a third language, Catalan, as the basis for our samples. The choice of initial combining forms was based on the consideration that where neoclassical formatives appear as the base of a derivation, they also occur as initial combining forms in a neoclassical compound. Almost all final combining forms also appear as initial combining forms. Initial combining forms are a much larger set and they can be easily retrieved also in an alphabetically ordered paper dictionary.

As our dictionaries, we selected CED (2000) and COED (2011) for English and Ušakov (1946-47) and Efremova (2000) for Russian. These dictionaries represent more traditional and more modern trends in British and Russian lexicography in such a way that we can consider our results typical of British and Russian dictionaries in general.

We found that English dictionaries give more neoclassical formations than their Russian counterparts, which suggests that they are more numerous in English. In addition, individual formatives are in many cases described in a separate entry in English dictionaries, but this is very rare in Russian dictionaries. These findings are in line with the hypothesis that for a significant proportion of speakers of English there is a system of neoclassical word formation, whereas this is not the case for speakers of Russian.

To the extent that this hypothesis is correct, the lexicographic policy on the inclusion of entries for neoclassical formatives adopted in English and Russian dictionaries can be justified by properties of the languages they cover. In English, having the information as to what a particular formative means will enable the user to decode new formations that are not in the dictionary, help the user build up a system of neoclassical word formation as part of their mental lexicon, and thus support the acquisition and retention of new neoclassical formations.

In Russian, the situation is different. New neoclassical formations will be borrowings, not the result of applying a neoclassical formation rule. If such a borrowing combines two neoclassical components, an entry for one of these components will often be of little help. In many cases, the other component will not exist yet, so that there is no obvious background structure into which a new formation could be incorporated.

In electronic dictionaries, representing neoclassical formatives is to be recommended for English. In an optimal representation, an entry for a formative will give access to the class of all formations it is part of. This includes both the use as an initial and as a final combining form, because a formative that occurs in both roles (e.g. *paedo* in *orthopaedic* and in *paedagogical*) is basically the same formative. For Russian, there is no similar level of support for setting up such a system.

## 7 References

### 7.1 Dictionaries

Bruguera i Talleda, Jordi (2006), *Diccionari de la formació de mots*, Barcelona: Enciclopèdia Catalana.

CED (2000), *Collins Dictionary of the English Language*, 5th edition, Glasgow: Collins.

COED (2011), *Concise Oxford English Dictionary*, 12th edition, Angus Stevenson & Maurice Waite (eds.), Oxford: Oxford University Press.

Efremova, Tatjana F. (2000), *Новый словарь русского языка* [New dictionary of the Russian language], Moscow: Russkij Jezik, <http://www.efremova.info>.

OED (2014), *Oxford English Dictionary*, Third edition, edited by John Simpson, [www.oed.com](http://www.oed.com).

Ušakov, Dmitrij N. (1946-47), *Толковый словарь русского языка* [Explanatory dictionary of the Russian language], online edition <http://www.dict.t-mm.ru/ushakov>.

### 7.2 Other works

Atkins, B.T. Sue & Rundell, Michael (2008), *The Oxford Guide to Practical Lexicography*, Oxford: Oxford University Press.

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

Béjoint, Henri (2000), *Modern Lexicography: An Introduction*, Oxford: Oxford University Press.

Benigni, Valentina and Masini, Francesca, (2009), 'Compounds in Russian', in *Lingue e linguaggio* 2/2009, pp. 171-194.

Bloomfield, Leonard (1933), *Language*, London: Allen & Unwin.

- 
- Booij, Geert (2003), 'The codification of phonological, morphological, and syntactic information', in van Sterkenburg, Piet (ed.), *A Practical Guide to Lexicography*, Amsterdam: Benjamins, pp. 251-259.
- Comrie, Bernard and Stone, Gerald (1978), *The Russian language since the revolution*, Oxford: Clarendon Press.
- Corbin, Danielle (1987), *Morphologie dérivationnelle et structuration du lexique*, Tübingen: Niemeyer (2 vol.).
- Germanovič, Ivan Klimovič (1979), 'Лексикография' [Lexicography], in Šuba, Pavel Pavlovič, Современный русский язык I [The contemporary Russian language I], Minsk: BGU, pp. 256-312.
- ten Hacken, Pius (1998), 'Word Formation in Electronic Dictionaries', *Dictionaries* 19:158-187.
- ten Hacken, Pius (2009), 'What is a Dictionary? A View from Chomskyan Linguistics', *International Journal of Lexicography* 22:399-421.
- ten Hacken (2012), 'Neoclassical word formation in English and the organization of the lexicon', in Gavrilidou, Zoe; Efthymiou, Angeliki; Thomadaki, Evangelia & Kambakis-Vougiouklis, Penelope (eds.), *Selected papers of the 10th International Conference of Greek Linguistics*, Komotini: Democritus University of Thrace, pp. 78-88.
- Molinsky, Steven J. (1973). Patterns of ellipsis in Russian compound noun formations. The Hague/Paris: Mouton.
- Ožegov, Sergej Ivanovič (1974), Лексикология, лексикография, культура речи [Lexicology, lexicography, speech culture], Moskva: Vyššaja škola.
- van Sterkenburg, Piet (2003), 'The' dictionary: Definition and history', in van Sterkenburg (ed.), *A Practical Guide to Lexicography*, Amsterdam: Benjamins, pp. 3-17.
- Svensén, Bo (2009), *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*, Cambridge: Cambridge University Press.
- Šanskij, Nikolaj Maksimovič (1972), Лексикология современного русского языка [Lexicology of the contemporary Russian language], Moskva: Prosveščeniye.
- Ulsamer, Sabina (2013), 'Wortbildung in Wörterbüchern – Zwischen Anspruch und Wirklichkeit', in Klosa, Annette (ed.), *Wortbildung im elektronischen Wörterbuch*, Tübingen: Narr, pp. 13-59.