

# **Entry Selection for a Bilingual Phraseological Dictionary**

**Vasyl Starko**

Volyn State University  
Applied Linguistics Department  
3 Robitnycha Str. – 43010 Lutsk  
Ukraine  
vstarko@operamail.com

## **Abstract**

The paper discusses the factors that influence the lexicographer's choice of entries for a bilingual phraseological dictionary: theoretical considerations, dictionary type, possible uses of the dictionary, expectations of the users, and other dictionaries, both existing and projected. The author argues for a robust classification of phraseological items included in the dictionary that would serve as a set of reference points for both the compiler and the user. An example of a such classification involving English idioms is provided along with the explanation of criteria and considerations for each class. It is suggested that with the necessary changes, the same approach can be adopted for compiling other phraseological dictionaries. Another argument concerns the need for secure areas of overlap in entry selection between various dictionaries for the same pair of languages.

## **1 Introduction**

One of the first issues that a lexicographer working on a project has to tackle is selecting appropriate entries for the projected dictionary. It is also one of the first questions potential users ask from their own perspective: What will we be able to find in this dictionary? A satisfactory answer to this question for both the compiler and the user goes a long way in ensuring the success of a dictionary. The compiler has to have a robust set of criteria for entry selection and the user needs clear guidance as to what the dictionary contains.

While the problem of entry selection requires careful study with respect to any kind of dictionary, it is especially acute for phraseological dictionaries. There are several factors that have an effect on the lexicographer's choice: 1) varying theoretical approaches to the defining of the scope and essence of phraseology; 2) dictionary type; 3) possible uses of the information presented in the dictionary; 4) expectations on the part of the intended user groups; 5) the state of affairs in lexicography in the targeted geographical area. Below we will discuss each of these problems in turn. The discussion will use the example of the English-Ukrainian dictionary of idioms we are compiling now and will set the scene for the subsequent explanation of entry selection for this dictionary.

## **2 Factors influencing entry selection**

### ***2.1 Problems with the theory***

There is general agreement among linguists about the lack of consistency of terminology

in the field of phraseology. Both the scope and the internal categories of phraseology are defined by different scholars in often non-complementary ways. Moon (1998: 19) writes that "there is no generally agreed set of categories, as well as no generally agreed set of terms. Moreover, no clear classifications are possible." This makes it all more difficult for a pragmatically oriented lexicographer to categorize phraseological units for the purposes of dictionary-making. Ideally, one would have clear-cut coherent categories with a finite set of distinguishing features. In practice, however, boundaries between different groups are blurred and groups themselves are amorphous.

As an example of the complexities involved Moon (1998) indicates that in her database approximately 25% of phraseological units have been assigned to two categories simultaneously. She points out that a theorist's approach to classification is different in nature from a lexicographer's. Whereas the former can employ a flexible system with multiple category assignment, the latter is forced into making dichotomic choices of inclusion/non-inclusion. This results in regrettable misrepresentation of the "continuum of idiomaticity or compositionality" (Moon 1998: 18).

## **2.2 Dictionary type**

This aspect is rather straightforward. The lexicographer needs to decide whether the dictionary is intended for human or machine use, whether it is going to be monolingual or bilingual, synchronic or diachronic and what group of users it will target. In what follows our discussion will focus on the principles for compiling a synchronical bilingual (English-Ukrainian) phraseological dictionary for human use that incorporates features of learners' and translators' dictionaries. The targeted user group ranges from advanced Ukrainian learners of English to teachers of English and professional translators.

## **2.3 Possible uses of the dictionary**

Dictionaries are used for two basic purposes: encoding (active use) and decoding (passive use). It would not be an exaggeration to say that the majority of existing dictionaries are more suitable for decoding than for active use. It is only recently that truly active dictionaries began to spring up (Apresjan et al. 2000, Longman 1993). Active dictionaries are more difficult to compile as the user needs to be provided with much more detailed information and the requirements for lexicographical presentation are strict (Apresjan 2000).

There are other choices to be made: is the dictionary going to be used for reference purposes, language learning, translation, and/or linguistic research? In our project we envisage that the dictionary under compilation can be used for all of these purposes. It should combine the features of learners' and translators' dictionaries and provide detailed and varied information needed for active use, as well as an extensive range of translation equivalents.

## **2.4 Users' expectations**

Users' expectations of a new dictionary are shaped by earlier commonly-used major dictionaries of the same type. These dictionaries thus establish a tradition in their field from which a radical breakaway is hardly welcome for any lexicographical work aspiring to be

economically viable. On the other hand, certain modernization is almost a must, as well as improvement over the deficiencies of similar reference works.

In Ukraine two bilingual phraseological dictionaries have enjoyed widespread popularity – Barantsev (2005) and Kunin (1998). At the time of its first publication in 1969, the former was a great lexicographic accomplishment, even though it suffered from somewhat indiscriminate entry selection. Its second edition was motivated by virtual unavailability of the dictionary on the market. However, only cosmetic modifications were made, with the editors admitting a great need for an up-to-date work.

Kunin (1998), especially since the fourth 1984 edition, became arguably the most authoritative English-non-English phraseological dictionary in the former Soviet Union. The dictionary is based on rigorous theoretical principles, in particular in its choice of entries. It represents the English language from the middle of the nineteenth century which, by today's standards, is not altogether contemporary. The author chose not to include archaic, jargon and obscene words, as well as the majority of scientific terms and such periphrastic expressions as *to take a walk*. The categories of lexical items selected for inclusion in Barantsev (2005) and Kunin (1998) will be discussed in section 3.

According to our personal observations, expectations of Ukrainian users have not been influenced by monolingual dictionaries of English idioms to a significant extent. These dictionaries are much less popular among Ukrainians than general-language monolingual dictionaries of English, primarily due to such detrimental factors as relatively high prices and absence of Ukrainian equivalents, which are felt to be a very important component in a phraseological dictionary.

### **2.5 The state of affairs in English-Ukrainian lexicography today**

In section 2.1 we mentioned the problem of misrepresenting the continuum of idiomaticity when it comes to binary choices of inclusion/non-inclusion. The problem can be resolved if we approach it from a wider perspective. The ultimate goal of lexicographical study of any particular language is to offer a description of all its coded, or institutionalized, lexical units. Assuming that all such units have been identified, the problem boils down to their distribution among different types of dictionaries. Theoretically, this is not mandatory as one can conceive of one all-encompassing comprehensive dictionary of the language in question (in the domain of phraseology such an attempt has been made by Czech lexicographers (Čermák 1994) working on a five-volume dictionary of Czech idioms). In practice, however, a dictionary of that kind is extremely difficult to compile, and even if written, it cannot be expected to describe all the lexical items in great depth. Thus, lexicographers work on separate dictionaries, often focusing only on a part of the whole vocabulary of a language. Ideally, their well-coordinated effort would eventually yield a lexicographic description of the language in its entirety. What happens in reality, though, is that coordination is not always present, with some areas of vocabulary enjoying lexicographers' (and publishers') unflagging interest and others being neglected for various reasons. Whatever the case is for a given language (or a pair of languages), a lexicographer undertaking a new dictionary has to take into account the existing dictionaries and, if there are gaps in their coverage, make certain assumptions about

future dictionaries, in particular about their entry selection. In these circumstances it is important to understand that since many items stubbornly resist unequivocal classification, they will need to be simultaneously included in two or even more dictionaries. Thus, the scope of these dictionaries should be constructed in such a way as to provide for an effective overlap. This is necessary for yet another reason. With comfortable areas of overlap in place, the users can have greater confidence that their reference needs will be met.

This approach can be applied to both the lexical continuum in general and to its specific areas, for example phraseology. Two of its success factors are a robust categorization of lexical items to be described, which we will discuss later, and understanding the state of affairs in a given lexicographical area. With this in mind, we will now proceed to an overview of relevant dictionaries in Ukraine.

As far as English-Ukrainian phraseological dictionaries are concerned, besides Barantsev (2005) there is one semi-bilingual dictionary of idioms (Chambers 2002), which was compiled by way of adding Ukrainian equivalents to the existing monolingual dictionary. There is also an English-Ukrainian handbook of proverbs and sayings (Dubenko 2004) in which a typical entry consists only of the English expression and its Ukrainian equivalent(s), without any examples or lexicographical labels. Two other dictionaries (Medvedeva, Dajneko 1994; Medvedeva, Holden 2003) are focused on certain types of expressions: word pairs, speech idioms and interjections. All of these are predominantly passive dictionaries and not as comprehensive as Barantsev (2005) or Kunin (1998).

There are several general-language English-Ukrainian dictionaries of which the two most notable are Balla (1996) and Zubkov (2003), each claiming over 110,000 entries. These dictionaries include only the most common phraseological units and provide the user with little more information on a specific expression than just a Ukrainian equivalent. This is partly due to restrictions of space since they were published as paper dictionaries.

To conclude, there is a general lack of up-to-date English-Ukrainian phraseological dictionaries. Thus, addressing the issue of entry selection for our dictionary, we will need to make a number of forward-looking assumptions as to what the future bilingual dictionaries should include and what types of dictionaries will be needed.

### **3 Entry selection in Barantsev (2005) and Kunin (1998)**

Barantsev (2005) covers a very broad spectrum of expressions including sayings, proverbs and winged words. The author's view on phraseology also encompasses restricted collocations (e.g. *to lose flesh*), terms (e.g. *allegation of faculties*), certain abbreviations (e.g. *ABC*), some free word combinations (e.g. *to have a full realization of something*), paraphrastic light-verb constructions (e.g. *to have a glance*), foreign phrases (e.g. *ab initio*) and certain archaic, dialectal and jargon expressions that occur in literature.

The dictionary (Kunin 1998) is distinguished, in A. Cowie's words (Cowie 1998a: 219), by "the principled selection and systematic analysis of entries. Items are chosen for inclusion strictly according to their membership of categories recognized by the compiler as phraseological." Let us now consider these categories.

Kunin (1998) divides the expressions included in the dictionary into three main classes: idioms, semi-idioms and phraseomatic units. Idioms are multi-word units that have partially

or fully figurative meaning and come in various structural types: noun, verbal, adjectival, prepositional and adverbial phrases, interjections, parentheses, constructions involving subordinate clauses (e.g. *fiddle while Rome is burning*) and full sentences (proverbs, sayings and pragmatic formulae such as greetings). All of these correspond to the types of units included in many other contemporary dictionaries of idioms. Most of the items of the second type (semi-idioms) are terms or professional expressions that have acquired figurative meaning (e.g. *a chain reaction*). In this latter meaning they are idiomatic expressions. The items in the third category (phraseomatic units) have either literal or "phraseomatically bound" meaning. This category is a bit of a ragbag. It comprises six subcategories: 1) phrases in which one of the components is uniquely combined, e.g. *all told* and *a boon companion*; 2) phrases with restrictive meaning, e.g. *in a hurry* in the sense "quickly, willingly"; 3) non-figurative fixed expressions, e.g. *again and again*, *safe and sound*, *to be conspicuous by one's absence* and *the beginning of the end*; 4) prepositional phrases with literal meaning, e.g. *at best* and *at most*; 5) terminological combinations that have gained currency in literary usage, e.g. *general ticket*; 6) phrases with phraseomatically bound meanings, e.g. *to pay attention*. In the appendix the author also includes foreign phrases (e.g. *per aspera ad astra*). It should be noted that type 5 expressions are now generally confined to terminological dictionaries regardless of their currency outside scientific contexts, whereas type 6 expressions are placed in dictionaries of collocations. The other types appear in different dictionaries of idioms with varying degrees of consistency.

To sum up, Ukrainian users may expect a contemporary English-Ukrainian dictionary to address a very broad scope of phraseological units including idioms proper, proverbs, sayings, certain literal expressions and possibly foreign phrases. They are less likely to look for a systematic treatment of restricted collocations (which have been represented sporadically) and pragmatic formulae. Finally, they would not in all likelihood expect to find phrasal verbs, grammatical phraseologisms (such as prepositions, conjunctions and particles), free word combinations, single-word items, as well as archaic, dialectal and jargon expressions.

#### 4 Classification of the included phraseological units

Before any specific categories are mentioned, let it be stated that we have used the orthographic criterion which requires that an entry unit be written as two words. Hyphenated, and sometimes unhyphenated, compounds are found in the dictionary only as run-ons, e.g. *trail-blazing* and *a trail-blazer* are placed under *to blaze a trail*, *a penny-pincher* under *to pinch pennies*, and *nit(-)pick(ing)* under *to pick nits*. In those cases where the spelling is not fully established we have included a lexical item, even if its hyphenated spelling is used more often, e.g. *dyed-in-the-wool*, which is less frequently spelled *dyed in the wool*, has a separate entry under the keyword *wool*.

The classes of phraseological units selected for inclusion in our dictionary are specified below. Also provided are some points of argumentation which explain the criteria we have used in the selection process.

1. Traditional, core units of phraseology, which are characterized by idiomaticity, semantic integrity and have the status of prefabricated units of the lexicon, e.g. *to spill the beans*, *to cut corners*, *to run smb. into the ground*, *a shot in the dark*, *a free hand*, *old hat*, *at death's*

*door, from hand to mouth, for the life of me, by leaps and bounds, to beat the band, hand over fist, if the worst comes to the worst, when the band begins to play, one's stock is high and the coast is clear.* In addition to this group the following types of fixed expressions are included.

2. Conventional similes, e.g. *as a doornail* (as in *as dead as a doornail*), *like the back of one's hand* (as in *to know smth. like the back of one's hand*), and *like a sponge* (as in *a memory like a sponge*). Also included are such expression as *the patience of Job* in which the surface structure disguises the underlying comparison (as patient as Job). Similes have been a traditional part of numerous phraseological dictionaries both in Ukraine and in other countries.

3. Intensifiers that cannot be interpreted literally, e.g. *to a tee, to death, (win/beat) hands down* (as in *win/beat hands down*), and (eat/drink smth.) *till it comes out of one's ears*. Idiomatic in character, these units have also enjoyed consistent treatment in dictionaries.

4. Strings containing items that are not found in other collocations (either as whole lexical units or in one of their meanings), e.g. *short shrift* and *all told*. Some of these expressions, most notably noun phrases, may be classified as restricted collocations, e.g. *a boon friend/pal/buddy*. Their equivalents are typically also restricted collocations and thus they are best included in dictionaries of collocations.

5. Pragmatic formulae, e.g. *Don't mention it!* and *Never mind*. Some of the most common expressions were included in (Barantsev 2005) and Kunin (1998). For the full coverage of pragmatic formulae a separate dictionary is needed of the kind similar to Gorodnikova, Dobrovol'skij (2000). In our project we have limited ourselves to those formulae that exhibit a certain degree of idiomaticity (as in the examples above) and may present problems for Ukrainian students of English.

6. Interjections that are idiomatic in nature, e.g. *Far out!*; *Give me a break!*; *Out of sight!* (meaning "Wonderful!") and others. Again, from a broad range of items in this category we have selected for inclusion only idiomatic ones.

7. Ordered word pairs and triples (irreversible binomials and trinomials) that are fixed in the language and have varying degrees of idiomaticity, e.g. *alarms and excursions, smb.'s pride and joy* and *slowly but surely*. These expressions are prefabricated, often redundant units that often pose special difficulties in translation. Kunin (1998) and Barantsev (2005) quite consistently include binomials in their treatment of English phraseology. Neither general-purpose dictionaries nor dictionaries of collocations can be expected to include all or the majority of these expressions. Once again, we have excluded clearly non-idiomatic items of this type, e.g. *boys and girls* and *knife and fork*.

8. Common proverbs and sayings, e.g. *you can't have your cake and eat it* and *enough is enough*. This category has been a standard feature of phraseological dictionaries in Ukraine. However, examples of usage were difficult to collect due to relatively low frequency of use. Thus, despite such modern reference works as Dubenko (2004), there is a need for a more detailed treatment that would involve definitions and examples.

9. Expressions which are not fully idiomatic because at least one of their constituent elements preserves its literal meaning, but which have traditionally been included in phraseo-

logical dictionaries, e.g. *to press/push one's luck* and *to cross one's mind*. It may be argued that this group of expressions belongs in a dictionary of collocations. In our opinion, in a phraseological dictionary it would be desirable to include only those items that have a high degree of fixedness in the language and were given treatment in earlier dictionaries (which causes the user to expect them also in contemporary sources). However, this category is difficult to delineate and inconsistencies exist within and across phraseological dictionaries. For example, they tend to include such expressions as *to speak one's mind* and omit *to pay/capture/catch/direct one's attention*, even though both cases contain, strictly speaking, a combination of metaphorical and literal components and both can be expected in a phraseological dictionary. As an example of a somewhat different nature, it may be argued that the expression *in the nick of time*, which is traditionally included in dictionaries of idioms, contains a non-metaphorical component *time* since it does refer to time. Strict exclusion of semi-metaphorical expressions would entail omission of this idiom. Thus, we are inclined to include only a small amount of such expressions using comprehensibility and translatability as additional criteria, i.e. preference is given to items that are more difficult to understand and/or translate and require more extensive treatment in the dictionary.

It should be noted that certain idioms can belong to more than one of the above categories, while other expressions can be said to belong to only one of them and only to a certain degree due to significant differences they exhibit. The categorization we have offered above is not intended to provide unequivocal category assignment. Rather it is meant to serve as a set of reference points helping the compiler in entry selection process and giving the user guidance as to what types of items the dictionary describes.

### 5 Expressions not included

It is equally important to outline those classes of expressions that are excluded from the dictionary:

1. Compound words of non-idiomatic nature, e.g. *pocket money* and *politically correct*. These ought to appear as separate entries in general-language dictionaries.

2. Compound terms and taxons, including those based on metaphor, e.g. *rabbit eye* "blueberry".

3. Restricted collocations, understood here as word combinations typically consisting of two or three words in which at least one element preserves its conventional meaning, e.g. *to throw a party* and *awkward silence*. This group borders on types 4 and 9 in the previous section.

Closely connected and even overlapping with restricted collocations are incompletely metaphorical expressions (in which, again, one element preserves its literal meaning), e.g. *to excite smb.'s anger* and *one's temper rises*. There is admittedly a fine line between expressions in this group and those mentioned in category 9 above. The deciding factors are fixedness, ready-made recall and tradition of lexicographical treatment (coupled with the users' expectations) versus compositionality, construction rather than reproduction and non-inclusion in earlier phraseological dictionaries. Additional criteria, as has been mentioned above, are difficulty for understanding and translation.

4. Phrasal verbs, including those that contain *it* and *oneself/yourself*, irrespective of the degree of metaphoricity. Here are some examples: *to put up with smth.*, *to go for it* and *to pick oneself up* "to recover from a fall or other mishap". Lexicographical description of his class requires a separate dictionary and it has been traditionally excluded from phraseological dictionaries in Ukraine.

5. Free word combinations, e.g. *to break an arm/leg/bone*. However, we do include fully idiomatic interpretations of such word combinations, e.g. *break a leg* in the sense "wish of success to an actor before a performance".

7. Grammatical or auxiliary phraseologisms performing ancillary functions in the sentence, e.g. *in spite of*, *instead of* and *so as to*. Due to frequency of their use these items are mandatory for inclusion in general-purpose dictionaries and are learned by Ukrainian speakers of English at early stages.

8. Hyphenated words including those based on metaphor, e.g. *chock-a-block* (full of smb./smth.), *dog-tired* and *trigger-happy*.

9. Separate words that acquire a special meaning when used in a specific grammatical form or syntactic construction, e.g. *on the cheap*, *to level with smb.*, and *one with smb.* in the sense "united".

Items in categories 7, 8 and 9 do not generally present special difficulties for students of English and translators and are to be described in general-language dictionaries.

10. Foreign phrases occurring in English writing and speech, e.g. *status quo* and *Sturm und Drang*. A separate lexicographical work needs to be dedicated to this group.

Four more classes of expressions deserve mention here: clichés, catchphrases, euphemisms and winged words. Classification of lexis into these groups is based on what can be called secondary features such as overuse and staleness (for clichés). This results in the use of these labels as umbrella terms for heterogeneous collections of lexical items and therefore we have refrained from using such classification for our purposes. For comprehensive treatment, special English-Ukrainian dictionaries need to be compiled for each group.

## 6 Additional guidelines

Additional criteria that we have used in selecting appropriate phraseological units for inclusion in the dictionary are summarized in the table below.

For inclusion	Against inclusion
British and American English as the two major varieties of English and models used in TEFL	Other varieties of English
Broad geographical and social distribution	Expressions limited to certain geographical areas or social dialects
Contemporary language (most citations starting from 1990)	Largely outdated items not surviving in the contemporary language
Items in the stylistic range from formal to highly informal	Items in the range from vulgar to taboo, as well as blasphemies

Figure 1. Table summarizing additional criteria

As the table suggests, we have decided to represent in our dictionary only British and American English as the two major varieties that serve as models in TEFL. The dictionary is going to comprise phraseological units known to the majority of speakers of these varieties of English as opposed to dialect and jargon expressions. In so far as possible, citations are drawn from contemporary sources (1990 till present). Preference will naturally be given to the most recent examples of use. A separate constraint is placed on the stylistic range of items in that vulgar and taboo expressions, as well as blasphemies are not included.

It should be noted that in other phraseological dictionaries these parameters may vary depending on the lexicographer's design.

### 7 Conclusion

A phraseological dictionary must be finely tuned to the users' needs and expectations, as well as to other dictionaries in current use between the same languages. Where there is lack of dictionaries describing certain classes of lexis, the compiler has to make predictions as to what these dictionaries will or should include once published.

A robust classification of lexical items included in the dictionary serves as a set of reference points for both the lexicographer and the user and provides helpful guidance with respect to the content of the dictionary. Equally important is a description of items not to be found in the dictionary and an explanation of overall criteria employed in entry selection.

We have used the example of the English-Ukrainian phraseological dictionary that is currently under compilation to demonstrate how this can be done in practice. We have suggested that as far as entry selection is concerned, there should be secure areas of overlap between this dictionary and English-Ukrainian dictionaries of other kinds, both existing and projected. With the necessary changes, the classification and principles set out in this article can be utilized to compile phraseological dictionaries of other types and between other languages.

## References

### A. Dictionaries

- Apresjan, Yu.D., Botjakova, V.V., Latysheva, T.E. (2000), *English-Russian Dictionary of Synonyms* (First edition 1979.), Moscow, Russkii jazyk.
- Balla, M.I. (1996), *English-Ukrainian Dictionary*. (In 2 vols.) Kyiv, Osvita.
- Barantsev, K.T. (2005), *English-Ukrainian Phrase-Book*. (First edition 1969.), Kyiv, Znannya.
- Chambers (2002), *Chambers English-Ukrainian Dictionary of Idioms*. Kyiv, Vseuvyto.
- Dubenko, O.Yu. (2004), *British and American Proverbs and Sayings*. Vinnytsya, Nova knyha.
- Gorodnikova, M.D., Dobrovol'skij, D.O. (2000), *Nemecko-russkii slovar' riechevogo obshcheniia*. Moscow, Russkii jazyk.
- Kunin, A.V. (1998), *English-Russian Phraseological Dictionary*. (Third edition 1967, fourth edition 1984.), Moscow, Zhyvoi yazyk.
- Longman (1993), *Longman Language Activator*. Harlow, Longman.
- Medvedeva, L., Dajneko, V. (1994), *English-Ukrainian Dictionary of Word Pairs*. Kyiv, Ukrainska encyklopedia.
- Medvedeva, L., Holden N. (2003), *English-Ukrainian Dictionary: Speech Idioms, Interjections, Onomatopoeia*. Kyiv, Dnipro.
- Zubkov, M.H. (2003), *A Comprehensive English-Ukrainian Dictionary*. Kharkiv, Folio.

### B. Other Literature

- Apresjan, Yu.D. (2000), 'English Synonyms and a Dictionary of Synonyms', in Apresjan, Yu.D., *Systematic Lexicography*, Oxford, Oxford University Press, pp. 3-54.
- Čermák, F. (1994), 'Czech Idiom Dictionary', in Martin, W., Meijs, W., Moerland, M., ten Pas, E., van Sterkenburg, P., Vossen, P. (eds.), *Euralex '94 Proceedings*. Amsterdam, 1994, pp. 426-431.
- Cowie, A.P. (1998a), 'Introduction', in Cowie, A.P. (ed.), *Phraseology: Theory, Analysis and Applications*. Oxford, Oxford University Press, pp. 1-20.
- Moon, R. (1998), *Fixed Expressions and Idioms in English*, Oxford, Clarendon Press.