

76. Bilingual phraseological dictionaries

1. Introduction
2. Corpus-informed bilingual lexicography
3. Outstanding challenges and lacunae in bilingual phraseological dictionaries
4. Conclusion
5. Select bibliography

1. Introduction

Bilingual phraseology has recently been the subject of much practical and theoretical work, driven by advances in general lexicography, language pedagogy, and corpus linguistics. As work continues on defining the optimal content and structure of entries in various kinds of lexicons, and as demands on the scope and quality of language-learning resources intensify, the availability of monolingual and parallel corpora put a new and valuable resource in the hands of lexicographers.

In this paper we highlight a number of the continuing challenges facing compilers of bilingual phraseological dictionaries as well as the new tools that are pushing to redefine, but also support the description of, phraseological entities. We focus on the presentation of phrases in bilingual phraseological dictionaries – as opposed to monolingual dictionaries or general bilingual dictionaries – since the breadth and depth of description differ greatly among these types of sources. We will call phraseological units “phrasemes” in order to circumvent the more semantically restricted label of “idiom”. (See Moon (1998, 2–5) for a summary of the wide variety of terms applied by linguists and lexicographers to the units in this field, and Cowie (1998, 1–20) for a detailed categorization of phraseological word combinations.) The rationale for discussing phrasemes instead of idioms per se is that pedagogical and practical (e.g., in natural language processing) experience has shown that knowledge of a broad range of (semi-)fixed entities is advantageous. The intersection between phrases, idioms and formulaic language has been discussed, e.g., by Oakey (2002, 86), and numerous recent works have supported Cowie’s prediction that “studies of collocations have pushed the boundary that roughly demarcates the ‘phraseological’ more and more into the zone formerly thought of as free, and it should not surprise us if future dictionaries consist of a higher proportion than before of collocations,

idioms, and formulae” (Cowie 1998, 20). In addition, the past decade has witnessed an increase in attention to lexical phrases as a productive approach to English language teaching (Nattinger/DeCarrico 1992). Thus, our broad definition of phraseme is in keeping with a practically-oriented, more inclusive trend in phraseological description.

2. Corpus-informed bilingual lexicography

During the past decade, lexicographic research has increasingly concentrated on dictionaries whose content is informed by corpus-based methods (see Atkins 2002, 1–29 for bilingual dictionary applications and Botley/McEnery/Wilson 2000 for an overview of relevant projects). The increasing reliance on corpora has been supported by a widely shared belief that lexicographer’s introspection should be complemented by external evidence. Moon (1998, 44) remarks, “I take it as axiomatic that effective and robust descriptions of any kind of lexical item must be based on evidence, not intuition, and that corpora provide evidence of a suitable type and quality”.

Both monolingual and parallel (i.e., multilingual aligned) corpora can assist in the decision-making of a lexicographer working on a bilingual phraseological dictionary. A monolingual corpus can:

- (a) detect common phrasemes, even if they do not fit into typical lists of idioms, defined as entities in which the sum of the parts does not equal the whole; for example, age is expressed in idiosyncratic patterns cross-linguistically: French *J’ai 4 ans* [lit.: I have 4 years], Russian *Mne 4 goda* [lit.: to-me 4 years], English *I am 4 years old*;
- (b) point out unexpected variants of phrasemes: e.g., *kick the bucket* is by far the most common form of this phraseme, but it permits a limited number of modifiers, like *kick the proverbial bucket*;
- (c) show a range of syntactic patterns in which a phraseme can participate, and with what frequency each such pattern occurs: e.g., the Russian idiom *sxvatit’ <pojmat’> za ruku* ‘catch red-handed’ [lit. grab <catch> by arm/hand] is often used in the indefinite personal form: *Ego sxvatili za ruku* ‘He was caught red-handed’, or in the negative perfective future generic 2nd person form: *Za ruku ego ne*

svxatis' 'You won't catch him red-handed' (this information is presented in Lubensky 1995, 577);

- (d) reveal the most typical ordering of words in a phraseme, as well as other possible although less typical orderings;
- (e) compare the frequency of phrasemes in various types of corpora – e.g., newswires versus literary prose;
- (f) give a truly rich picture of the semantic and pragmatic contexts in which a phraseme can be used, as well as its combinatorial properties and co-occurrence patterns.

Parallel corpora of the languages being described can highlight the degree of syntactic correlation between phrasemes with a given meaning in different languages, and indicate to what extent phraseme correlations between the given languages are fixed or, conversely, to what extent context-sensitive descriptive turns of phrase are employed.

The increased availability of parallel corpora has impacted not only lexicographic practice but has also stimulated interest in using parallel texts in language instruction, whose proponents emphasize their advantages over bilingual dictionaries:

"Dictionaries [...] lack the richness of context that occurs in parallel texts; and furthermore they lack the flexibility afforded by using parallel texts, where any number of patterns can be searched for. The student/investigator is not limited to the words and phrases that happen to have been chosen by the dictionary maker. In addition, dictionaries vary greatly in how well they deal with collocational information" (Barlow 2000, 114).

Barlow's argument, however, creates an unnecessary strawman: there is no reason to put dictionaries and parallel corpora in competition since they have different strengths and weaknesses. Whereas it is true that parallel corpora can show more contexts than are possible in dictionaries, they are also full of noise, including incorrect and imprecise translations, and they do not provide the detailed description possible through the introspection of a highly-trained lexicographer. Thus it is toward convergence rather than dominance of one genre or the other that the field should seek to move.

However, despite the advantages that new corpus-based methods afford, they do not solve the most difficult lexicographic challenges; they are just a tool like any other. In the next section we discuss some of the outstanding challenges of bilingual phraseologi-

cal dictionaries with an emphasis on those aspects that we believe deserve continued attention.

3. Outstanding challenges and lacunae in bilingual phraseological dictionaries

Fillmore and Atkins summarize the difficult decisions lexicographers have to make as follows:

"Making dictionaries requires making choices. From the mass of data about a headword, assembled during the analysis process, the lexicographer selects those facts which best suit the requirements that have been defined for the dictionary being compiled. What will the intended user of this dictionary be looking for? What will he or she be able to understand? What facts about the word are so important that they must be included in any account of the word, regardless of the intended user? But, above all, what can be left out?" (Fillmore/Atkins 1994, 350–51).

Compilers of bilingual dictionaries face these problems on both sides, the source language (L1) side and the target language (L2) side.

Different interpretations of the abovementioned decision space leads to a diverse array of bilingual phraseological dictionaries. The most important differences include: the size and degree of comprehensiveness; the selection of headphrases, depending on how the compiler defines phraseme; the ordering of entries; their structure; the presentation (or omission) of grammatical, syntactic, and pragmatic information; the selection of target language equivalents; and the nature, quantity, and presentation of illustrative material. Some of these differences are explained by the fact that different dictionaries aim to serve different audiences, e.g., speakers of L1 vs. speakers of L2, or intermediate level learners vs. advanced speakers (e.g., translators and scholars). In the subsections below we will address those of the abovementioned issues, which, in our judgment, require continued attention in print dictionaries.

3.1. Inventory of headphrases

Users must know what they can expect to find in a dictionary, and it is the compiler's job to explain the principles of phraseme selection succinctly and clearly *on a level accessible to the user*. First, it must be explained which broad categories of phrasemes are included: "pure" idioms of *kick the bucket* type, sayings

and proverbs, clichés and formulae, etc. Second, it must be indicated whether only current usage is included or whether obsolete usage is covered as well – as would be necessary for a user reading historical or literary texts. In most bilingual phraseological dictionaries (BPDs) the principles of selection (along with a compiler's approach to the ordering and presentation of items), are clarified in the front matter, but some BPDs do not have any front matter, leaving the user to figure out the principles of selection by trial and error (e.g., Gualtieri 1995).

As concerns coverage of headphrases, it seems clear that a comprehensive phraseeme dictionary should include all idioms as narrowly defined, multi-word interjections, conjunctions and particles, and however many fixed turns of phrase can be incorporated considering the usual space restrictions imposed on print resources. Both in language pedagogy and in the machine processing of text, the practical benefits of truly broad coverage are proving to eclipse the theoretical rationale behind a more restricted definition of lexicalizable phrasemes.

3.2. Contents of a headphrase

If orthographic, morphological, or lexical variants are included in the headphrase, it is preferable to present them typographically in such a way that they do not affect alphabetical ordering. For languages with unpredictable word stress, headphrases should be supplemented with accent marks. In addition, the headphrases should be presented together with their contextual partners, i.e., words with which they typically collocate. If any of these words may be elided, this should also be indicated. For example, the Russian verbal idiom *brat' nogi v ruki* [lit: take feet into hands/arms] has an elliptical variant, in which the verb itself is elided. So, *beri nogi v ruki i v magazin* 'take_{IMPERATIVE} legs into hands/arms and to the store' and *nogi v ruki i v magazin* 'legs into hands/arms and to the store', are equivalent, both meaning 'Get your butt in gear and run to the store!'

3.3. Usage labels

In unidirectional dictionaries, intended primarily for L2 speakers, usage labels are given predominantly on the L1 side, whereas in bidirectional dictionaries they need to be provided on both sides. Temporal labels, which show an idiom's standing in relation to con-

temporary usage, and the frequency label *rare*, are typically borrowed from the most authoritative, current monolingual dictionaries. Other commonly used categories of usage labels, such as "style and status" (Landau 2001, 240), may be liable to a higher dose of subjectivity, depending on the compiler's starting point: what is *neutral* for one, may be *informal* for another. It is useful if such labels are clearly explained and supported with examples, as is done, for example in the monolingual dictionaries LDOEI 1979 and CO-BUILD Idioms 2002, as well as the bilingual dictionary Lubensky 1995.

The inventory of labels characterizing attitude – especially when the label indicates both the attitude itself and its degree – can raise so many questions as to make such labeling counterindicated. The terminological problem is particularly acute when labels are presented in the source language for a target-language audience. For example, the compilers of the *Russisch-deutsches phraseologisches Wörterbuch* chose to give all usage labels in Russian, even though the target audience includes speakers of German, who might find the labels opaque. Moreover, among the Russian stylistic labels are *prezr(itel'noe)* and *prenebr(ežitel'noe)*, which might be glossed as 'derisive (slightly more pejorative)' and 'deriding (slightly less pejorative)', respectively. The semantic difference between these labels is barely perceptible, and the need for both is not motivated by the dictionary compilers (Petermann/Hansen-Kokoruš/Bill 1995 XXVII).

The traditional inventory of usage labels would benefit from the incorporation of some newly coined labels, which have been devised for hitherto uncovered territory. A number of such labels appear in the *New Explanatory Dictionary of Synonyms of the Russian Language*, and some can be productively applied to phrasemes: e.g., the label *narrat(ivnoe)* 'narrative' points out that the appropriate type of discourse for the given entity is literary narrative, and alerts the reader not to use it in colloquial speech; and the label *stil(izovannoe)* 'stylized' marks obsolete lexical items that could be used in the contemporary language to achieve a stylized effect (Apresjan/Boguslavskaja/Levontina et al. 1997, vii).

3.4. Ordering of headphrases

There are several approaches to ordering headphrases, which differ significantly with

regard to user-friendliness. These approaches are found both in monolingual and bilingual phraseological dictionaries, so we draw examples from both. Since the method of ordering is less important for speakers of L1, we will focus on ordering for L2 speakers who may differ in their command of L1.

The two most well-represented approaches are ordering by the first word of the phraseme and ordering by its main content word. Alphabetical ordering by the first word in Cowie/Mackin/McCaig (1983) (a dictionary of English) is justified by the fact that the volume has a comprehensive index of head-phrases. This approach would not, however, work as well for languages with non-restricted word order, although a comprehensive index or a well-developed cross-reference system might help. The other prominent option, ordering by the main content word, has its own disadvantages. Discussing accessibility of multi-word expressions in the *Macmillan English Dictionary for Advanced Learners*, Bogaards (2003, 45) poses the question: "The first [problem] is: which element is to be considered as the main one? In cases like **pocket the difference** or **all's fair in love and war**, this question cannot be answered without some hesitation." Obviously, even expert opinions can sometimes differ on which element of a given idiom should be considered the main one – not to mention that it would be unrealistic to expect that the users' judgments will always jibe with those of the compilers.

Even in relatively straightforward cases, such as noun phrases and binomials – where the noun in the former and the first noun in the latter are generally considered the main word – alphabetizing by the main word may not work well for speakers of L2 if L1 has complex inflectional paradigms. E.g., if a non-Russian speaker comes across the phrase *sinim čulkom* 'bluestocking' (instrumental sing.) or *sinix čulkov* (genitive plural) in a text, he or she must know that the nominative form is *sinij čulok* in order to find the phraseme in a dictionary (that is, the nominative singular form of the noun has the fleeting vowel 'o'). In other words, effective lexicon use typically requires that users know morpheme-boundary alternations, irregular plurals, case form variants, suppletive forms, and suchlike. This problem might be alleviated in corpus-based BPDs if, for example, a search program

were supplied with a morphological analyzer; however, it persists in print dictionaries.

When the verb in verb phrases or clauses is considered the main content word, the matter of tracing it back to its base form may be even more complex. Whereas deriving base forms from inflected ones is not a serious problem when L1 is morphologically impoverished, like, English (e.g., in Baldini 1985, all the phrasemes that contain the verb 'to pay' are placed under the heading **TO PAY**), in morphologically rich languages it can be a daunting problem. For example, languages like Russian and Polish have morphologically rich verbal paradigms that include different forms of the infinitive for the perfective and imperfective aspects, extensive alternations in conjugation, and developed systems of prefixation that impede any straightforward approach to alphabetical ordering. In short, all these prefixal puzzles, suffixal riddles, and occasional suppletive mysteries will certainly confound an inexperienced user who sees in a text the masculine past singular perfective form of the verb 'buy': "**Kupil** kota v meške!" 'He bought a pig in a poke!' and must trace it back to the infinitival imperfective form: **po-kupat'** kota v meške (e.g., Arsent'eva 1999, 204).

It seems more efficient to put practicality ahead of theory and order entries by the content words that are *the least changeable* part of the phraseme. For example, in Slavic languages, like Russian, Ukrainian, Polish and Czech, the noun components in idiomatic VPs and in some idioms functioning as sentences generally do not change, with the exception of an occasional change in number; thus indexing the phraseme under the fixed nominal form should make it easier to track down for less experienced users. However, two additional ordering conventions must be systematically followed. First, even when verbal phrasemes containing a fixed noun are ordered alphabetically by the noun, it is still necessary to be consistent in presenting aspectual forms (i.e., consistently place the same aspect first) in order to facilitate the search in case there are numerous verbal idioms with the same head noun. Second, the head nouns should be ordered according to the forms in which they are used in the phrasemes, rather than in their "base" form (usually nominative singular). This is particularly relevant for languages with well-developed declension paradigms.

Regardless of how dictionary entries are ordered, users who do not have a linguistic background (and most likely would not bother to read the front matter), still must be able to find every idiom without getting frustrated by the process. The main two options to assist such users are cross-referencing and including an index – with the maximum benefit derived from including as many content words (and, in the case of multi-word interjections and particles, non-content words) as possible. Such features are absolutely indispensable for languages with non-fixed word order, where phrasemes may vary their word order in different contexts.

While both cross-referencing and indexing, if compiled expertly, fulfill their functions, indexing has certain advantages over cross-referencing. Indexing usually saves space in comparison with cross-referencing – not a minor consideration for print dictionaries. Even more importantly, indexing is more attractive aesthetically, which is not a frivolous matter, as described by de Schryver: “Some aspects of the traditional paper dictionaries that, as of today, remain unbeatable [...] they are easy to browse, can truly be read recreationally, [...] provide pleasure when physically handled, and do not stress the eyes as much as computer displays” (de Schryver 2003, 152). It is much more enjoyable to browse a dictionary or read it for fun when dictionary entries do not mingle with long stretches of cross-referencing – in other words, it is nice for a dictionary to look and feel like a book.

3.5. Definitions

A sampling of bilingual phraseme dictionaries indicates that most do not provide definitions for L1 entries (Kunin 1984, Levintova/Wolf/Movshovich et al. 1985, Petermann/Hansen-Kokoruš/Bill 1995). Yet in order to fully understand the meaning of the headphrase, L2 speakers – especially translators – actually need more information than just an L2 equivalent: they need explanatory definitions of the L1 phrasemes written in their own native language (L2). For example, it is doubtful that the brief English definitions provided in the *English-Hungarian Dictionary of Idioms* are sufficient for Hungarian-speaking learners of English, who would benefit immeasurably more from extensive definitions written in Hungarian. The compiler of that dictionary actually sought to equally serve the

L1 and L2 audiences, saying: “Special consideration has been given to the needs of Hungarian students studying English, but no less important is the emphasis put on the needs of foreign students learning Hungarian” (Nagy 1996, 23).“ However, since English speakers already know the definitions of the head phrases, and Hungarian speakers need more help than a brief definition, this attempt to find a middle ground seems non-optimal. We have found no convincing argumentation in favor of limiting the size or vocabulary of L2 definitions if the audience is, indeed, L2 speakers, and agree with Fillmore and Atkins (1994, 364) that a “substitutability requirement” – by which the definition must be roughly substitutable in contexts where the word occurs – is superfluous.

3.6. Grammar

In a bilingual dictionary for L2 speakers, some grammatical, stylistic, and pragmatic information is needed on the L1 side. Scrupulously selected grammatical information supports the efforts of L2 speakers who wish to master a given phraseme for active use. It also highlights standard use, thus helping users to recognize nonstandard occurrences in a text, play on a phraseme’s meaning, deconstruct a phraseme for stylistic purposes, and the like.

Because phrasemes are generally similar to non-phraseme entities in their structure and syntactic function, an indication of the type of phrase (NP, VP, AdjP, etc.) will be helpful to some users. However, some phrasemes show atypical properties of their respective phrase types, and these deviations must be indicated. For example, whereas verbs in fleective languages can generally be used in a range of inflectional forms, in some phrasemes the form is fixed, as in the following Russian examples, where the verb (in boldface) must be used in the infinitival form. Only in the last example can both aspects of the verb be used.

- *delat’ nečego* [lit.: do_{INFIN} nothing] ‘there is nothing to be done (about it)’
- *ot nečego delat’* [lit.: from nothing do_{INFIN}] ‘for want of anything better to do’
- *delit’ nečego* [lit.: share_{INFIN} nothing] ‘one has (got) no bone to pick with someone’
- *devat’/det’ nekuga* [lit.: put_{INFIN-IMPV}/put_{INFIN-PV} nowhere] ‘someone has more (of something) than he knows what to do with’

Apart from inflectional restrictions, other information that should be provided for verbal entities is the type(s) of subject and object(s)

it can select, in terms of both grammar and semantics. Some useful grammatical categories to describe verbal complements are count noun, non-count noun, collective noun, infinitive and clause. Some useful semantic categories are human noun, animal noun, personal or geographical entity, concrete noun and abstract noun. Of course, in some cases the selectional restrictions can be narrowed even further, and if the potential fillers are limited to just a few entities, they should be listed outright.

Users also benefit from seeing contrastive differences between the structure and function of L2 and L1 phrasemes. One such difference regards word order within a phraseme. Information on word order is indispensable when L1 on the whole has unrestricted word order because unrestricted word order often does not apply to phrasemes. Phrasemes in “free” word-order languages might have fixed word order, fixed word order apart from one movable element, fixed word order except when the phraseme is used in a given syntactic configuration, etc. All such restrictions should be indicated.

The presentation of equivalents for certain types of phrasemes – especially VPs, sentential phrasemes, and certain adverbial intensifiers – warrants special discussion. In our judgment, these types of phrasemes should be presented as sentence patterns for L1, with all L2 equivalents following the same pattern. Both the L1 pattern and its L2 equivalents can use variables, just as they are used in some monolingual dictionaries for semantic explications (e.g., Mel’čuk, I. A./N. Arbatchewsky-Jumarie/L. Elnitsky et al. 1984). Take, as an example, the Russian idiom *glaza razbegajutsja/razbežalis* [lit.: eyes run-all-over_{3PL PRES/PL PAST}], whose translations are presented in Lubensky 1995:106 in a pattern that reads literally ‘at X eyes are-running-all-over (from Ys)’ (we present a subset of translations),:

- u X-a *glaza razbegajutsja (ot Y-ov)* ≈ X doesn’t know <X scarcely knows> where to look (first <next>); X is dizzyed by the multitude of Ys; it’s more <there are more Ys> than the eyes can take in.

The use of patterns offers a number of advantages over traditional presentation styles. Patterns can show the most typical temporal/aspect frames for phrasemes and can underscore the correlation between verb forms in languages with different tense-aspect structures.

For example, the present tense form *X katit bočku na Y-a* [lit.: X rolls_{3RD SING PRES} barrel on Y] can be translated as ‘X is trying to lay the blame on Y’, ‘X is trying to make Y a scapegoat’ (Lubensky 1995, 31); thus, the imperfective aspect in Russian is conveyed by the modal auxiliary ‘try to’ in English. Patterns can also show different syntactic realizations of actants in each language, as in *glaza razbegajutsja*, presented above.

3.7. Equivalents

Linguists and lexicographers generally agree that bilingual dictionaries cannot equally serve two audiences – speakers of L1 and L2 – because even the best bilingual dictionaries do not provide the same amount of information on both the L1 and L2 sides. While there is no *principled* reason why they could not, there is a practical reason: space and cumbersome. L1 speakers need more information about L2 and vice versa, making it prudent to provide namely and only the information for the target audience. The matter of providing adequate information for target users is particularly evident in the case of the presentation of equivalents for phrasemes.

Some BPD’s provide a number of synonymic L2 equivalents without explaining any semantic and/or stylistic differences between them or illustrating such differences with examples or citations. Thus, an Italian user of Gualtieri (1995) looking for English equivalents of *Sentirsi molto abbattuto* (which is presented together with the less common *Sentirsi accasciato (depresso)* and the very common *Essere giù di spirito*), will find eleven English equivalents: ‘to feel flat (bad, cheap, rotten, seedy)’; ‘to be in low water’; ‘to be low-spirited’; ‘to have the hip’; ‘to be down in the mouth’; ‘to be sore at heart’; and ‘to have one’s heart in one’s boots’. Some of the equivalents are so different semantically and stylistically (not to mention their questionable status in standard English to begin with) that it is difficult to get a sense for the semantic core of these phrasemes; yet only ‘seedy’ and ‘down in the mouth’ are provided with brief usage notes.

A similar approach to bunching phrasemes and their equivalents is taken in Kuzmin (2001), which is a Russian-English phraseological dictionary published in Russia and intended primarily for native speakers of Russian. Kuzmin claims that his dictionary “is particularly true to the idea that its readers

must have *free choice* [Kuzmin's emphasis] for making adequate decisions" and consequently offers "thesaurus blocks of equivalents" (2001, v and vi) on the English side. However, these equivalents are not supported by a sufficient number of contexts, nor is there any guidance for Russian speakers in choosing which of the equivalents would be best in a given context. For example, for the Russian idiom *duši ne čajat* 'to love someone deeply, to be extremely fond of someone' Kuzmin offers the following equivalents: 'be the apple of one's eye'; 'to think the world of smb, smth'; 'to love smb with all one's heart and soul'; 'to have a soft spot for smb'; 'to care for smb, smth'; 'to fall for smb'; 'to worship the ground smb walks (or treads) on', etc. The Russian phraseme, however, is used in a wide variety of contexts and the equivalents are not mutually replaceable: e.g., how would a Russian user know that a child can be the apple of his parent's eye but not the other way round? Such coarse-grained bunching approaches do not, in our opinion, come close to the goal of bilingual phraseological dictionaries, which is to provide all the semantic, contextual, stylistic and usage information that a non-native speaker needs in order to appropriately use phrasemes in real-world situations.

A BPD for L2 speakers, especially for translators, must contain all of the above plus something else: equivalents that are catered to specific contexts. Translators need a milieu, a habitat for each idiom, that helps them not just to understand it, but to feel and sense it. This way, if a translator cannot use the equivalents offered in a dictionary, he can rely on his own creativity to produce a new equivalent. The technique developed in (Lubensky 1995) uses, where warranted, two types of equivalents: generic ones and ones labeled "in limited contexts," with all of the latter instances being exemplified by literary or invented examples.

One of the main considerations in the selection of equivalents is their functional equivalence in both languages. Selection of functional equivalents is based on a contrastive analysis of phrasemes in both languages, which involves a number of theoretical issues – cognitive factors, preferred types of metaphorization, culturally based differences, etc. (Dobrovolskij 2000; Altenberg/Granger 2002). We agree with Dobrovolskij that "in the absence of a total cross-linguistic equiva-

lence between a given L1-idiom and L2-translation (and this is mostly the case), the nearest equivalent is an expression which is based upon the same conceptual structure and, therefore, evokes mental images from the same conceptual domain" (Dobrovolskij 2000, 172–173).

Obviously, bilingual corpora are a powerful tool in helping lexicographers to find such equivalents. However, bilingual corpora by themselves are not a dictionary, and their data need to be carefully analyzed to allow for a selection of L2 functional equivalents. That is, while the correspondences in bilingual corpora can offer lexicographers excellent suggestions for functional equivalents, the degree to which parallel corpora are actually parallel is often far lower than would be desired for a research tool (that is, many parallel corpora reflect loose paraphrase rather than translation). Ideally, one would want, for example, all quality translations of literary works – which tend to stay quite close to the original – to be on-line supported by an interface that permitted search and comparison.

When a lexicographer is working with the originals and their translations (published or on-line), a caveat would be in order: translations should be put under a magnifying glass to ascertain that the rendering of a phraseme is the best possible choice for each given context. We repeat with Peters/Picchi/Biagini (2000, 73): "[...] the full meaning of any word or expression is dependent on and cannot be divorced from its context. Any work of translation must thus be not only linguistically correct but also appropriate to the frame of reference of the source, i.e. it must also achieve equivalence at the level of style, register, cultural and social context, etc." It is the phraseographer's job to carefully analyze the quality of translation along each of these parameters, since a user may not be competent to do so.

It seems that corpus organization along the lines suggested by Johansson and Hofland (2000, 134) for the English-Norwegian Parallel Corpus – i.e., having an equal number of English and Norwegian original works (fiction and non-fiction) along with their translations – contains a safety valve: it allows the lexicographer to navigate from the original to the translation in both directions and thus separate widely applicable equivalents from context-specific translations. It will also per-

mit the elimination of lapses, occasionally made even by professional translators.

One issue bilingual lexicographers have to deal with is selecting (and sometimes inventing) equivalents for phrasemes which, while not culturally specific in meaning, include proper or geographical names, such as the English *carry coals to Newcastle*, whose closest culture-specific Russian translation is *v Tulu so svoim samovarom ne ezdjat* (lit. 'you don't bring your own samovar to Tula'). The cultural clash in referring to Newcastle in, say, a translation of Russian literature makes the use of this equivalent impractical in most contexts; but its presentation in a dictionary is appropriate for general semantic and stylistic orientation. One way of alerting naive users to the undesirability of using culture-specific phrasemes in translations is by listing them not as equivalents per se but in a "cf." zone, which effectively says "this phrase is like X but don't use it as an equivalent". Another similar example is the English *Rome wasn't built in a day* as compared with the French *Paris ne s'est pas fait en un jour* 'Paris wasn't built in a day' and the Russian *Moskva ne srazu stroilas'* 'Moscow wasn't built all at once'.

It is particularly challenging to establish L2 equivalences for L1 phrasemes conveying emotions, especially for multi-word interjections that contain a modal element. It is this group of phrasemes that reflects differences in national temperaments and characteristic emotional reactions between L1 and L2 speakers. Because interjections are used mainly in direct (and occasionally in reported) speech, but not in narratives, their use is heavily conditioned by the pragmatic aspects of communication. Most such phrasemes (or meanings of polysemous phrasemes) can express different degrees or aspects of a given emotion and/or reflect the speaker's temper and idiolect; as such, they require an especially detailed presentation, which, ideally, should include: a usage note explaining typical contexts and social appropriateness; a particularly large number of equivalents in order to cover various contexts and degrees of emotion; and a particularly large number of typical, contextually sufficient examples. In other words, entries for such idioms should re-create each idiom's emotional milieu, thus enabling L2 speakers (especially translators) to develop a feel for the idiom and to come up with their own

equivalents for contexts unforeseen by the compiler.

A question arises: how should phraseographers handle "unused" L2 idioms, i.e., idioms that have no phraeological equivalents in L1 and therefore no natural home in a BPD? One could disregard them, but doing so would exclude some linguistically and culturally important phenomena. One solution would be for the compiler to check which L2 idioms ended up not being included and try to find an appropriate place for at least some of them. Additionally, some culturally important idioms, which cannot be recommended as equivalents, might be appropriately placed in the "cf." zone so that they help users grasp the cultural distinctiveness of L2.

3.8. Illustrations

The illustrations zone is where the two kinds of differences between BPDs are particularly noticeable: the *kind* of illustrative material presented and the *approach* taken to translating the examples. With regard to the kind(s) of illustrative material, BPDs show a range of possibilities:

- no illustrations at all (Petermann/Hansen-Kokoruš/Bill 1995)
- brief invented examples, which are frequently contextually insufficient to illustrate the use of the given phraseme (Arany-Makkai 1997, Baldini 1985, Benedito 1990)
- literary citations in both languages, which are occasionally supplemented by examples from media or other dictionaries, and are usually limited to one citation per entry (Čerdanceva/Recker/Zor'ko 1982, Levintova/Wolf/Movshovich et al. 1985, Andrejčina/Vlaxov/Dimitrova et al. 1980, Kunin 1984)
- a combination of literary citations and invented examples in both languages (Lubensky 1995).

Most learners' BPDs present invented examples, which is a simple and space-saving approach to illustration. However, invented examples often fail to provide sufficient context, especially preceding context, which is essential for understanding how the phraseme is used. Additionally, there is always a danger that both the invented examples and their L2 equivalents will reflect preferences related to the compiler's (or compilers') idiolect. Some dictionaries combine invented examples (both their own and those borrowed from other dictionaries) and literary and/or media citations. Comprehensive BPDs definitely benefit from literary citations, especially when they are

presented along with their published translations since the compiler can select a context that is sufficient to highlight the use of the given phraseme.

With regard to translating the illustrative material, various approaches have been taken. Invented examples are invariably translated by the phraseographers themselves, and the translations can, therefore, reflect too heavily the phraseographers' own idiolects (like invented examples themselves). The situation is more complex with literary citations, which in some dictionaries are translated for the most part by the dictionary compilers, and in other dictionaries are presented along with their published translations. In his introduction to (Čerdanceva/Recker/Zor'ko 1982, 12), Recker offers an explanation of why the compilers of this dictionary did not use some published translations of Italian literary works, even translations of high artistic quality: responding to the specificity of context, translators may replace one artistic image with another, use a stylistic device that is different from the original, or make up for a loss of a detail in another part of the context. Consequently, such translations would not illustrate the equivalents selected by the compilers. While complexities associated with inexact, context-affected equivalents certainly exist, the conclusion drawn by Recker is not fully convincing: there should be a place in a dictionary entry for creative context-specific translations. Lubensky (1995), for example, includes literary illustrations that reflect so-called "contextual translations", which show particularly felicitous context-specific translations. Contextual translations exemplify how appropriate a non-standard translation can be, conveying both the semantics and stylistic nuances intended by the author.

4. Conclusion

Phrasemes differ widely in their complexity and, accordingly, in the type and extent of information that need be provided about them in bilingual phraseological dictionaries. On one end of the spectrum are phrasemes like *outdoor pool* (not **outside pool*), whose multi-word status is basically arbitrary and whose description in a phrasal dictionary need include only a definition or an equivalent sufficient to convey to the L2 speaker what type of entity is being described. On the other end of the spectrum are the more complex types

of phrasemes described above: for these, no amount of information and no number of examples are superfluous to the motivated L2 learner – the more, the better. However, there are practical considerations. In print dictionaries, the inventory of entries and the amount of illustrative material tend to be especially hard hit by space constraints. Moreover, a truly unbridled inclusion of, say, examples, could make the dictionary too cumbersome to be easily used.

On-line dictionaries, by contrast, present some interesting possibilities for resolving these problems. For example, one could design an on-line bilingual phraseological dictionary using methods of "progressive disclosure", such that a basic entry were provided first with links to more information of specific kinds for interested users. In fact, such a dictionary could be connected to an aligned corpus in which lexicographers have selected the best, say, 10 illustrations (with "best" defined in terms of scope of idiom use, sufficiency of the translation, etc.), then the next best 10, etc., until some threshold of inventory size has been reached. Such guided corpus access would provide truly sufficient contexts for L2 learners trying to get the feel of using given phrasemes.

5. Select bibliography

5.1. Dictionaries

Andrejčina, K./Vlaxov, S./Dimitrova, S./Zaprjanova, K. (1980): *Russko-bolgarskij frazeologičeskij slovar'*. S. Vlaxov (ed.). Sofia & Moscow.

Apresjan, J.D./Boguslavskaja, O.J./Levontina, I.B./Uryson, E.V./Glovinskaja, M.J./Krylova, T.V. (1997): *Novyj ob'jasnitel'nyj slovar' sinonimov russkogo jazyka. Pervyj vypusk. Pod obščim rukovodstvom akademika Ju.D. Apresjana*. Moscow.

Arany-Makkai, A. (1997): *2001 Russian and English idioms*. Hauppauge, N.Y.

Arsent'eva, E.F. (1999): *Russko-anglijskij frazeologičeskij slovar'*. C. Carlson (ed.). Kazan'.

Baldini, A. (1985): *Idioms: Dizionario idiomatico Inglese-Italiano*. Reggio Emilia.

Benedito, F.S. (1990 [1977]): *Diccionario conciso de modismos Inglés-Español Español-Inglés*. Madrid.

Čerdanceva, T.Z./Recker, J.I./Zor'ko, G.F. (1982): *Ital'jansko-russkij frazeologičeskij slovar'*. Moscow.

COBUILD Idioms 2002 [1995, Founding Editor-in-Chief John Sinclair]: *Collins COBUILD dictionary of idioms*. Glasgow.

- Cowie, A.P./Mackin, R./McCaig, I.R. (1983): Oxford dictionary of current idiomatic English. Volume 2: Phrase, Clause & Sentence Idioms. Oxford.
- Gualtieri, F.M. (1995): Dizionario Italiano-Inglese della fraseologia idiomatica. Genova.
- Kunin, A.V. (1984 [1955]): English-Russian phraseological dictionary. 4th Edition, Revised and enlarged. Moscow.
- Kuzmin, S.S. (2001): Translators' Russian-English phraseological dictionary. Moscow.
- LDOEI 1979: Longman dictionary of English idioms. Editorial Director T.H. Long. Managing Editor D. Summers. Harlow.
- Levintova E./Wolf, E./Movshovich, N./Budnitskaya, I. (1985): Diccionario fraseológico Español-Ruso. Moscow.
- Lubensky, S. (1995): Random House Russian-English dictionary of idioms. Developmental Editor M. McShane. New York.
- Mel'čuk, I.A./Arbatchewsky-Jumarie, N./Elnitsky, L./Iordanskaja, L./Lessard, A. (1984): Dictionnaire explicatif et combinatoire du français contemporain. Volume I. Montreal.
- Nagy, G.L. (1996): English-Hungarian dictionary of idioms: British and American English. Budapest.
- Petermann, J./Hansen-Kokoruš, R./Bill, T. (1995): Russisch-deutsches phraseologisches Wörterbuch. Hrsg. von J. Matešić. Leipzig.
- Wheeler, M. (1992 [1972]): The Oxford Russian-English dictionary. General Editor B.O. Unbegaun. Oxford.
- 5.2. Literature
- Altenberg, B./Granger, S. (2002): Recent trends in cross-linguistic lexical studies. In: Altenberg, B./Granger, S. (eds.): Lexis in contrast: corpus-based approaches. Amsterdam, 3–48.
- Atkins, B.T.S. (2002): Bilingual dictionaries: past, present and future. In: Corréard, M.-H. (ed.): Lexicography and natural language processing. A Festschrift in honour of B.T.S. Atkins. EURALEX. [First delivered as plenary paper at EURALEX 1996 and published in *EURALEX '96 Proceedings*, 515–546]
- Barlow, M. (2000): Parallel texts in English teaching. In: Botley, S.P./McEnery, A.M./Wilson, A. (eds.), 106–115.
- Bogaards, P. (2003): MEDAL: A fifth dictionary for learners of English. Review article. [Review of: Rundell, M. (ed.) (2002): Macmillan English dictionary for advanced learners. Oxford.] In: International Journal of Lexicography, vol. 16, no. 1, 43–55.
- Botley, S.P./McEnery, A.M./Wilson, A. (eds.) (2000): Multilingual corpora in teaching and research. (Language and computers: studies in practical linguistics, No. 22). Amsterdam.
- Cowie, A.P. (1998): Introduction. In: Cowie, A.P. (ed.): Phraseology: theory, analysis, and applications. Oxford, 1–20.
- Dobrovolskij, D. (2000): Contrastive idiom analysis: Russian and German idioms in theory and in the bilingual dictionary. In: International Journal of Lexicography, vol. 13, no. 3, 169–186.
- Fillmore, C.J./Atkins, B.T.S. (1994): Starting where the dictionaries stop: the challenge of corpus lexicography. In: Atkins, B.T.S./Zampolli, A. (eds.): Computational approaches to the lexicon. Oxford, 349–393.
- Johansson, S./Hofland, K. (2000): The English-Norwegian parallel corpus: current work and new directions. In: Botley, S.P./McEnery, A.M./Wilson, A. (eds.), 134–147.
- Landau, S.I. (2001 [1984]): Dictionaries: the art and craft of lexicography. 2nd edition. Cambridge.
- Moon, R. (1998): Fixed expressions and idioms in English. Oxford.
- Nattinger, J./DeCarrico, J. (1992): Lexical phrases and language teaching. Oxford.
- Oakey, D. (2002): Lexical phrases for teaching academic writing in English: corpus evidence. In: Nucorini, S. (ed.): Phrases and phraseology – data and descriptions. Bern, 85–105.
- Peters, C./Picchi, E./Biagini, L. (2000): Parallel and comparable bilingual corpora in language teaching and learning. In: Botley, S.P./McEnery, A.M./Wilson, A. (eds.), 73–85.
- Schryver, G.-M. de (2003): Lexicographers' dreams in the electronic-dictionary age. In: International Journal of Lexicography, vol. 16, no. 2, 143–199.
- Tomaszczyk, J./Lewandowska-Tomaszczyk, B. (eds.) (1990): Meaning and lexicography. Amsterdam.

*Sophia Lubensky, Albany, NY (USA),
Marjorie McShane, Baltimore, MD (USA)*