A Modern Bilingual Dictionary — Results and Prospects

Valerij P. Berkov

Humanity has been creating bilingual dictionaries for four thousand years at least: we are familiar with fragments of bi- and trilingual dictionaries compiled about 2000 BC. This means that wide experience has been gained in the field of bilingual lexicography. But paradoxical though it may be, the number of theoretical papers which draw on this experience in order to improve bilingual lexicography even today unadequately reflects the real importance of such linguistic activity.

I realize that speaking about results and prospects of bilingual lexicography is a rather responsible task, and that it will hardly be possible to deal comprehensively with all aspects of the complexity of problems. Such a task is beyond the resources of one person. The purpose of my lecture is more modest: to draw your attention to some important aspects of bilingual dictionaries, to review some time-honoured traditional decisions, and to discuss possible alternatives. Some of these considerations are to a certain degree implemented already in my large RUSSIAN-NORWEGIAN DICTIONARY. Bilingual dictionaries have, among other things, the peculiarity that many theoretically reasonable and practically advantageous decisions used in them go often unnoticed not only by the ordinary reader but also by the experienced lexicographer. Explication of such new methods in prefaces or introductions to dictionaries does not help much: firstly, this new information is often hard to find in the large amount of rather trivial material; secondly, one knows by experience that prefaces and introductions are not given due attention by readers. All this results in many interesting new lexicographical decisions known but to a relatively narrow circle.

Also relevant here is the fact that the majority of lexicographers are not polyglot and are therefore simply not able to acquaint themselves with many new lexicographical methods if they do not know or do not know well the source and the target languages of the dictionary. For this reason the compilers of bilingual dictionaries often invent, independently of each other, new ways of presenting material, not knowing that some very clever decisions are to be found in dictionaries already published, or in other words, using a Russian phrase, "invent a bicycle" (= reinvent the wheel).

I think that this situation must be improved. The following is one possible solution to this problem.

It would be appropriate to help lexicographers in different countries to join forces and to appoint a commission which would analyze the best specimens of bilingual dictionaries and on this basis devise an international standard. Standardization is the watchword of our time, and would give a number of advantages:

(1) Such a standard would sum up the valuable experience of many lexicographers in many different countries and make it available to all. It would raise the level of bilingual lexicography throughout the world. Such recom-

mendations are of particular importance for those countries where bilingual lexicography is at an early stage of development.

(2) The standardization of bilingual dictionaries would make it easier for people in different countries to use such dictionaries, whose common design would exempt the reader from having to master different, sometimes rather divergent, ways of presenting similar material.

One concrete proposal which might lead to the solution of the problem might be the following: that our Congress should elect a Commission to evolve standards for bilingual dictionaries.

The Commission would include representatives of different countries, it must be representative, but at the same time not too numerous, because what it would gain in coverage it would lose in efficiency. An optimal number would seem to be a body of 10—15 persons. Needless to say, it must consist of persons who have a wide experience in compiling bilingual dictionaries and have studied seriously also the theoretical aspects of lexicography. It ought to be a matter of course that the Commission should include persons who have worked with various languages, including those of Asia, principally South-East Asia, and Africa.

When we think of the importance of international relations in the modern world, and of the study of foreign languages increasing world-wide, and attaining global dimensions in the optimization of foreign language learning, we can conclude that the work of further improvement in bilingual lexicography could and should be done under the aegis of UNESCO. If the participants of our Congress share my opinion, the Congress could approach UNESCO with this problem. Such work, done under the aegis of UNESCO, would acquire international authority. To be sure, the directives of the commission could only be recommendatory, but the authority of an organization which is a specialized institute of the UN would give them great weight.

Without going into details concerning the work of the proposed commission, I would only emphasize the fact that its task would be to work out parameters and structures of several dictionary types ranging from the largest to the smallest ones intended for beginners.

It is obvious that the task the Commission would have to solve would be difficult enough. The greatest difficulty would clearly be to take account of the peculiarities of languages with different structures and different writing systems (special mention must be made of Arabic, Vietnamese, Chinese, Japanese etc.).

Finally, on this question of bilingual dictionaries standardization, I would like to draw your attention to a specific and important aspect of the problem. Purely economic considerations show that a colossal number of languages, probably the greatest majority, support only one or two bilingual dictionaries, for example, in the People's Republic of Angola the inhabitants speak the following languages: Owimbundu (2.4. millions), Mbundu (1.5 millions), Kongo (850 thousands) etc.; it is self-evident that—at least in the foreseeable future—the only bilingual dictionary possibility for these languages will be with Portuguese as source and target language. Therefore anyone who wishes to study these languages will be compelled to use Portuguese bilingual dictionaries, resulting in a situation where a great number of language students have to use dictionaries in which neither the source nor the target language is their mother tongue. If the present considerable diversity of bilingual dictionaries is eliminated and the material arranged according to

academically acceptable and pedagogically effective principles, it will facilitate their use for many readers.

Let us now discuss the features that in my opinion a modern large bilingual dictionary must have. I emphasize the word "large": I shall further discuss large dictionaries that have 6 or more million characters (in the USSR it is more than 150 printed sheets, i.e. more than 2,400 pages with 2,500 characters on each).

The main principle of a bilingual dictionary as of every reference book was formulated long ago: the maximum of information in the minimum of space. Practical experience of translation, however, and theoretical considerations show that a translation—both from the mother tongue and from a foreign language—can be successful only when the reader can find in the dictionary a large amount of information of various types. Therefore the above-mentioned considerable dimensions of the dictionary ought to be used very economically. In other words, the modern bilingual dictionary must be characterized by a certain amount of compression, but not so much as to have an adverse effect on the reader's interests.

The other basic premise of the bilingual dictionary is the observance of the frequency principle. To be sure, this is impeded—for the time being—by the lack of satisfactory frequency dictionaries (not to mention the fact that most languages have no frequency dictionaries at all). In some cases, however, it is justifiable to include low-frequency words — those which designate concepts which are important for communication between two groups speaking different languages; for instance, in a situation where the equivalent in the target language has a high frequency (e.g. some fish names in a Russian-Icelandic dictionary).

The third principle—one unfortunately not duly respected—is a clear and thorough assessment of the minimum of information on the phonetics and morphology of the foreign language, indispensable for using the dictionary successfully. This determines the distribution of phonetic and morphological information between the "corpus" and the grammar outline. Thus the main body of dictionary and the grammar outline must be coordinated. Moreover, it is possible to cross-refer from the corpus to the grammar outline and vice versa, a process which, in addition to saving a considerable amount of space, will have an incontrovertibly beneficial pedagogical effect.

The fourth principle runs as follows: aside from some special cases, bilingual dictionaries should in one volume be intended for users speaking both the source and the target languages, as either L1 or L2.

In the fifth place, and very importantly, a bilingual dictionary must provide (as far as humanly possible) working equivalents rather than glosses and explanations. Further, a dictionary must prevent (as far as humanly possible) mistakes before the reader makes them.

The sixth principle is that a bilingual dictionary is essentially a confrontation of two cultures. Those lexicographers are right who think that a bilingual dictionary must also be an ethnographical reference book.

The seventh principle is that a bilingual dictionary viewed as a whole ought to be an unified translational tool. This means that its task is not only to provide the reader with equivalents of lexical items of different kinds but also provide him or her with information about the means by which the target language can express the grammatical information on the source language.

Each of these principles involves a number of more specific ones. It is impossible to discuss all of them in this lecture, only some of them will be considered.

1. The systematic economizing of space in a large dictionary results in considerable savings. The simple elimination of the semicolon before the figures separating meanings of a polysemous word, and of the period at the end of the entry, which was carried out in the USSR some years ago, saved on each large dictionary at least 150.000 characters. Below I mention other space-saving methods in a dictionary.

1.1 For languages which, owing to their orthography, need phonetic transcription, only those words are transcribed where deviation from the regular is observed, wherever possible, partial transcription is preferred. For example, the English words *black, blaze, bleed, cup, few, miss, paddle, sender, rupture, rustic, shabbiness, willing* require no transcription; with German words such as *Adler, Arzt, Mond, Wüste*, only the long vowels are to be marked; in the case of words like *ab, Grammatik*, on the other hand, the shortness of the vowel is indicated. It is enough to transcribe the English words *misery* as ['mɪz-], *northern* as [-ð], *river* as ['rɪ-] etc.

1.2 In many dictionaries the morphological characteristics given for entry words are rather wasteful of space. Nor is the other extreme, the printing of a mere symbol, e.g. a two-figure code for the concrete information a fortunate solution. It is appropriate to balance this information between the main body of the dictionary and the 'end matter', as it is done for example in the new series of the large Russian-... dictionaries (index system by A.A. Zaliznyak). This will enable the reader to acquire the basic morphological information with the minimal waste of time consulting the 'end matter' (i.e. grammar tables) in the outline only when necessary.

1.3 Elimination of regular derivatives from the entry list is a tradition which results in considerable space saving. It is however worth paying attention to two points:

1.3.1 Where the omission of a derived form means that this form is translated regularly, i.e. that this omission is significant, the derivative must in each case be tested for regularity in translation; for instance, the omission of the word *vrijwilligheid* in the latest DUTCH-RUSSIAN DICTIONARY (Moscow, 1987) is justified, but this is not the case with the word *scheefheid*, because its equivalent in Russian is not regular (кривизна, косина).

b) The reader should be informed how regular equivalents are built. At present, the best that he or she can find is information in the grammar outline (or in the section on "word-formation") on the meaning of different derivation formatives, without any explanation of how they are rendered in the other language. Therefore it seems to be necessary to include in the corpus as separate entry units different derivation formatives (prefixes, suffixes, transfixes, circumfixes, etc.) with a description of their meaning, giving examples indicating (the most important of) equivalents (if possible with morphological comments).

1.4 All languages have series of homogeneous lexemes that can be used in homogeneous constructions, e.g. names of the days of the week, cf. *on Friday/Tuesday..., on Fridays/Tuesdays..., last Friday/Tuesday...,* etc. The same is true of the names of months. Much space is saved when such constructions are given only once, e.g. under the headword which comes alphabetically first, e.g. under the headwords *Friday* and *August*, resp., with cross-references in the other

entries as, for example, in the RUSSIAN-NORWEGIAN DICTIONARY, to the headword where the constructions in question are to be found. The same method can be used when a series of source language words is translated by means of a series of homogeneous constructions, cf. the entries норвежка and француженка, немка, etc. in the RUSSIAN-NORWEGIAN DICTIONARY.

1.5 It is a rather frequent phenomenon to find that a word group in the source language has two equivalents: a regular one and an irregular one. In such cases all one can do is to include the irregular translation preceded by the label *also*.

1.6 Finally, in many dictionaries we may find the same group of words or multi-word lexical units appearing in different entries (sometimes, which is unforgivable, with different translations). Such units must be listed at only one headword, to which the other elements are cross-referred. The technique can vary, but this problem and the problem of how to identify the key word cannot be discussed here.

2. Simple calculations show that in order to create a frequency dictionary which might form the basis of the entry list of a large dictionary, the sample has to be about $1·10^{11}$ for words with the frequency $1·10^{-7}$ ("not very frequent words"), the relative statistical error being $\delta \approx 0.03$, i.e. for unsatisfactory precision of the results. At this stage frequency dictionaries for lexicography are far from attaining even these parameters. Therefore the deductive approach is of special importance.

The frequency principle, if consistently observed, requires the inclusion of many types of lexical units on which lexicographers' opinions are divided. We shall discuss only some of them.

2.1 The dictionary should include all those types of lexical units which lie above the frequency threshold accepted by the compiler, including substandard words and even taboo words (of course with the required labels).

2.2 The problem of the amount of terminology in a general dictionary calls for some discussion. Many correct statements have been made about this. It is sufficient to stress the importance of including terminology that current conditions have brought into common use by non-specialists, e.g. many terms in the fields of sociology, medicine, motoring, photography, radio, TV, sound recording, etc.

2.3 A characteristic of many dictionaries is their unsatisfactory treatment of proper names. Only some of the numerous problems these pose will be discussed here.

2.3.1 For many languages it is important to include anthroponyms—surnames, given names and their derivatives—providing the user of the dictionary with information on pronunciation, stress, or sex of the owner, etc. For instance, the RUSSIAN-NORWEGIAN DICTIONARY contains a list of the most widespread Russian surnames with indications of stress; it is the first time in Soviet lexicography that such a list is included. It is evident that the need for such information depends on the language: Finnish surnames do not need any indication of pronunciation, Finnish given names, on the other hand, must be provided with the indication of sex, e.g. *Toivo* is a male name, *Aino* a female one. It is obvious that dictionaries should include diminutives of Russian given names, cf. Саша, Шура — diminutives of Александр(а), Гоша of Георгий or Игорь, etc.

Similarly, the names of historical persons are not properly represented in most dictionaries, though it is evident that many of them are frequent enough to justify inclusion and that it is often impossible "to guess" their equivalents, cf. the Russian Иван Грозный, Александр Македонский, Сократ, Гейне, Теккерей, Гегель

and their English counterparts *Ivan the Terrible, Alexander of Macedonia, Socrates, Heine, Thackeray, Hegel*.

2.3.2 Dictionaries usually include current place names, but for historical texts many old names are equally important, cf. *St. Petersburg, Gallia, Ceylon, Persia, British Honduras, The Gold Coast*, etc.; and so are names of places which no longer exist as such, e.g. *Carthage, Troy*, and names of places which derive their importance from their association with important historical events, e.g. *Borodino, Thermopylae, Hastings, Waterloo*, etc.

As a rule dictionaries do not list microtoponyms such as *Red Square, the Champs-Élysées, Hyde Park, Wall Street*, etc. It is also useful to include names of world-famous buildings, monuments, etc., e.g. the *Coliseum, the Leaning Tower, Westminster Abbey, White House, the Great Wall of China*, etc.

Important also are periphrastic names of cities and countries, e.g. *The Northern Palmyra, the Eternal City*, etc.

2.3.3 It is likewise necessary to include in the dictionary the most important mythological names, for example *Zeus, Jupiter, Athena, Odin*, etc., for we all know that mythological motifs are used in many books and works of art. The same must be said for Biblical names such as *Abraham, Noah, Methuselah, Ruth, Judas*, etc. (By the way, the titles of he principal books of the Bible ought also to be included, e.g. *Genesis, I (II) Book of the Kings, The Gospel according to St. Matthew*, etc.).

Inadequate also is the present practice with regard to the names of well-known characters in folklore and literature, e.g. *Cinderella, Manilov, Don Quixote, Hamlet, Gargantua, Sherlock Holmes*, etc.

2.3.4 In my opinion the inclusion of a fairly large list of titles of the best known literary works is fully justified: it is an established fact that the same work has often different titles in different languages. Such a list is included in RUSSIAN-NORWEGIAN DICTIONARY. Equally useful would be a list of the most outstanding works of art (operas, ballets, films, pictures, sculptures, etc.).

I am convinced that all proper names ought to be placed in the corpus, but I have no time to give further grounds for this opinion.

2.4 Winged words (literary quotations) are as a rule treated unsystematically and inadequately in bilingual dictionaries.

To be sure, full justice can be done to them only in specialist dictionaries (I dare say, the first bilingual dictionary of winged words was my Russian-Norwegian one, Moscow, 1980), but it is possible, and indeed justifiable to include some hundreds of winged words in a general dictionary.

3. If the compiler determines clearly the level of the competence in the phonetics and morphology of the foreign language which the reader must have in order to be able to use the dictionary successfully, this will ensure, on the one hand, that all necessary information is included, and it will permit the exclusion of redundant material, on the other; in other words, it will make the dictionary more informative. It can be achieved by including an explicit description of the competence required in the grammar outline in the dictionary (we cannot go into detail here: it may be boldface, etc.). It is essential that the grammar outline and the corpus are interdependent. Let us discuss some aspects of this statement.

3.1 On pronunciation marking: in the grammar outline are indicated the principal sounds which correspond to each letter (and if required those instantiating digraphs and trigraphs, e.g. German *sch*). All deviations from these

rules (but only these deviations) are registered in the corpus, e.g. Engl. live [lɪv] ("according to the rule" *[laiv] is expected), Germ. Obst ['o:pst] (instead of the "expected" *[ɔpst]), Norw. bonde ['bun:ə] (instead of *['bɔŋ:ə]), etc.

This is a complex mass of problems, connected as it is with a number of more specific questions: establishing how much information is to be given about pronunciation in the dictionary as a whole, assigning this information to the corpus or the outline, deciding on the type of transcription, and so on. This network of problems is dealt with at length in my book "Slovo v dvujazychnom slovare" [A Word in a Bilingual Dictionary] (Tallin, 1977).

3.2 A certain amount of explicit grammar information about the language in question makes it possible to solve optimally a number of problems.

3.2.1 Grammatical indications about the words can be more compact since they will not contain information already known to the user.

3.2.2 Such a method gives the compiler a clear criterion of which noncanonical word forms (i.e. forms that differ from the form of the entry-word) are to be included in the dictionary as an entry-word and which are not. Although this principle is self-evident, dictionaries often ignore it. For more about this see my books "Voprosy dvujazychnoj leksikografii" [Problems in Bilingual Lexicography] (Leningrad, 1973) and "Slovo..."., cf. above.

In this connection it may be noted that since in many languages such irregular forms form alternations (cf. Russ. рыть — рою, роешь, роет, etc., разбить — разобью, разобьёшь, разобьёт, etc., дочь — дочери, дочерью, etc., Germ. *nehmen — nimmst, nimmt, nimm!*), a generalized presentation is justified, e.g. разобь fut. of разбить, etc.; and one is used in the RUSSIAN-NORWEGIAN DICTIONARY.

3.2.3 It is appropriate to arrange the dictionary, and especially the grammar outline, in such a way that it may be used in ad hoc translations, i.e. by a person without knowledge of the language in question, and interested in translation of one concrete text.

4. For purely economic reasons, the bulk of bilingual dictionaries must be aimed at both the source and the target language speakers, i.e. they have to give both categories of user all essential information. This statement seems to contradict the well-known thesis about the necessity of four dictionaries for each pair of languages, but this contradiction is imaginary: the point I wish to make is that each category of user has its own specific needs, as well as requiring much information or help to the other group as well.

5. The problem of adequate equivalents is certainly the central problem of the bilingual dictionary. For these equivalents general principles are easiest formulated, yet for these very equivalents inaccuracies and mistakes are often made. The reason seems to be that the selection of equivalents is largely as much an art as a science, in the same way that literary translation is an art. At the same time it is evident that a dictionary can only give *langue* correspondences while in reality one often has to translate *parole*, which frequently demands the substitution of another less accurate item for the exact equivalent, less accurate from the point of view of equivalence but more adequate from the point of view of its function in the translated text.

Time does not permit me to discuss the whole complex of problems connected with equivalents. I shall consider briefly a selection of them, reflecting typical errors found in many dictionaries.

5.1 The equivalence problem is closely related to the filiation problem, i.e. with the problem of the arrangement of the meanings in an entry. As a rule this arrangement must not depend on the number and the character of equivalents (cf. the views of the Prague and the Bratislava lexicographers).

A frequent mistake, caused by what E. Nida calls the difference in specialization level of languages (cf. E. Nida Nauka perevoda (The Science of Translation). In: "Voprosy jazykoznanija" [Problems in Linguistics], 1970, Nr. 4, p. 11.), is an unjustified splitting up of meaning, without an overall translation[1] being offered, in spite of the fact that this problem has several times been discussed in theoretical lexicography. A detailed exemplification can help in such cases.

5.2 In the early stages of the development of bilingual lexicography, one source language word was usually given one equivalent. A frequent drawback of many modern dictionaries is just the opposite: they provide the reader with a lot of equivalents without differentiating between them. For instance, if the source language word means 'brave' it is also given the following "equivalents": *courageous, fearless, unafraid, bold*, etc. This is not an advantage, as many think, but a drawback. Such presentation veils the difference between the source language synonyms (the synonym series of two languages, of course, not always being symmetric). Such an *embarras de richesse* does not simplify the use of the dictionary, but on the contrary, complicates it unduly. It is true that very often a meaning of the source language word is covered only by some partial equivalents, but a list of such equivalents must contain explanatory comment. This rule was formulated long ago, but unfortunately it is not always observed.

5.3 Less attention has been paid by lexicographers to quite the opposite case, i.e. to the situation when a given equivalent has a wider meaning than the source-language word, cf. Russ. прийти — Germ. *kommen*, while the correct translation is ... (*zu Fuß*). A human being internalizes the information found in a dictionary, and remembering the equation прийти = *kommen* will be inclined to translate the word *kommen* by прийти also in cases where it should be translated as приехать, прилететь, etc.

5.4 There is another inaccuracy that is often met with in bilingual dictionaries: the confusion of equivalents and explanations (the latter having a metalinguistic function). This happens especially often in definitions of culture-bound lexical units and those which mark the object of a process, e.g. in the "NORWEGIAN-RUSSIAN DICTIONARY" we find "*gane* gut a herring" instead of "... gut (a fish)".

5.5 In connection with definitions of meaning of culture-bound lexical units, many problems arise. Some of them will be discussed below. Here I will point out that in some cases pictorial illustrations are necessary, a treatment of such lexical items frequently mentioned in the literature.

5.6 It is self-evident that an equivalent must correspond to the word to be translated in all its connotations. Among the most widespread errors, the rendering of a stylistically-marked word by a neutral one must be noted first of all. Labelling is often neglected too. Finally, an exact equivalent may be such a rare word that the most of the speakers of the target language do not know it, and the user must be informed of this.

5.7 The compiler of a bilingual dictionary has to foresee and prevent mistakes a reader is apt to make. The Russian word фронтовик in the RUSSIAN-NORWEGIAN DICTIONARY does not list the equivalent "frontkjemper" as it signifies the Norwegian who fought against the USSR on the side of Nazi Germany; the reader is informed of this by a parenthetical explanation with the sign NB. In Russian the word интермеццо is only a musical term, but in Norwegian the word *intermesso* means also "incident, episode"; the parenthetical explanation is: (NB *mus.; not fig.*).

6. Probably no existing bilingual dictionary gives all the various types of cultural information which the reader needs in order fully and exactly to comprehend a text in a foreign language. This is a different and complicated subject[2]. I will here dwell only on some of the main aspects. It is obvious that dictionaries have to describe cultural phenomena that do not exist in the "other" culture, and to show the possible differentials between such cross-cultural phenomena. Less evident are two other discrepancies:

6.1 Discrepancy in the function of similar phenomena: e.g. яхтклуб in the USSR is a democratic sports organization while a *yacht club* in Great Britain is an association of yacht owners.

6.2 Discrepancy in symbolic meaning. In Russia the birch is a symbol of Russian nature, of the fatherland, while the equivalents of this word in other languages lack this connotation.

It is a task for the theory of bilingual lexicography to devise an optimal presentation of such cultural information. For the time being the practice is ahead of the theory.

7. I have now come to the last aspect which I wish to consider. If the dictionary is to be a universal translation tool, it should enable the reader to translate grammatical units too. In other words it should contain rules for establishing a correspondence between the source language grammemes and the target language grammemes (of course, to the extent that this is possible). For instance, a process that began in the past and is still going on in the present is rendered in Russian by the present tense (я жду тебя два часа), whereas in English by the present perfect or the present perfect continuous (*I have been waiting for you for two hours*).

Such information in a bilingual dictionary, arranged systematically and clearly presented so that it is easy to grasp, is very useful, especially if it is linked in some way with the material in the corpus.

* * *

In this talk I have made an attempt to sum up some of the achievements of practical and theoretical bilingual lexicography. Many problems have of necessity found themselves outside the scope of the lecture. The main object of this discussion was to show how difficult it is to compile a really informative bilingual dictionary and to urge lexicographers from various countries to join together and coordinate their efforts to work out a standard which would make it possible to raise the level of bilingual lexicography throughout the world.

Notes

[1] An example of such incorrect presentation from the INDONESIAN-RUSSIAN DICTIONARY: *punjangga* "1) author; writer; poet; 2) scholar (linguist, philologist, philosopher)" while the right translation is 'humanist, representative of Enlightenment'.

[2] See, e.g. "Slovo v dvujazychnom slovare" [A Word in a Bilingual Dictionary], Tallin, 1977. pp 84—109.

Cited Dictionaries

NORVEZHSKO-RUSSKIJ SLOVAR' [NORWEGIAN-RUSSIAN DICTIONARY]. 1963. V.D. Arakin. Moscow.

RUSSKO-NORVEZHSKIJ SLOVAR' [RUSSIAN-NORWEGIAN DICTIONARY]. 1987. V. P. Berkov. Moscow: Russkij Jazyk.