DOI: 10.11649/cs.2014.006

ANNA KISIEL A , JOANNA SATOŁA-STAŚKOWIAK B & WOJCIECH SOSNOWSKI C

Institute of Slavic Studies, Polish Academy of Sciences Warszawa, Poland A ania-kisiel@tlen.pl ; B joannasat@op.pl ; C wpsosnow@uw.edu.pl

THE NEED FOR AN ELECTRONIC MULTILINGUAL DICTIONARY

Abstract

The paper analyses the issue of providing adequate equivalents in multilingual dictionaries. If equivalents are adequate, it means that: (1) the scope of meaning of one item is identical to its equivalent (cf. drive: drive a nail vs. drive a car); and (2) the collocations of the equivalents overlap. Two significant problems arise when searching for adequate equivalents: the lack of equivalents whose meanings are identical (narrower/wider meanings, partial overlap of meanings, more than equally good equivalent), and equivalents with homographs in a given language. Because such issues are difficult to resolve in a printed dictionary, we put forward some methods of addressing the problems in an electronic dictionary. The paper offers an example entry from such a dictionary, which presents a suggestion of a layout. We also took into consideration the potential problems which may appear if the entry is presented in this manner: first, one must set a limit for the description (a defined number of lexical units); second, one must avoid circularity, but at the same time also strive for an exhaustive description. Electronic dictionaries offer greater possibilities of presenting modern vocabulary and adding new classifiers (e.g. a classifier of politeness).

Keywords: lexicography, multilingual dictionaries, interlingual equivalence, multiple meanings.

The idea of a tri-lingual dictionary stems from the contrastive studies by V. Koseska-Toszewa, J. Satoła-Staśkowiak (2013) and V. Koseska-Toszewa, J. Satoła-Staśkowiak, M. Duszkin (2012a-b). The work on the enterprise is continued by W. Sosnowski, V. Koseska-Toszewa and A. Kisiel. To date, 4 volumes have been prepared — the first volume will be published in 2015.

Because our dictionary contains three languages, it differs significantly from other traditional bilingual dictionaries. The dictionary's uniqueness does not consist only in the fact that three languages were combined — the dictionary's authors carefully follow the ongoing trends in the languages in question and keep track of

the vocabulary reflecting these trends. Owing to such an approach, the dictionary is in line with the latest research in semantics and contrastive linguistics. The lexical material for the dictionary is based on a Parallel Polish-Bulgarian-Russian corpus, which is being developed in Zespół lingwistyki korpusowej i semantyki (The Group for Corpus Linguistics and Semantics in the Institute of Slavic Studies of the Polish Academy of Sciences)¹. None of the languages of the dictionary has a primary status; regardless of the possible number of forms and meanings in a given language, we endeavour to accurately present all languages. Such a principle of design entails a need for an electronic version of the dictionary.

When developing a multilingual dictionary, the greatest problem that one encounters is how to provide adequate equivalents. If the equivalents are adequate, it means that:

- the equivalents are indentical as regards the scope of their meanings (cf. Pol. jazda, Bul. es∂a and Rus. es∂á 'drive, ride' as a description of movement; vs. Pol. jazda 'travel', Bul. nomeweemeue, Rus. nymewéemeue as a word denoting a journey; vs. Pol. jazda 'trip', denoting something nice and thrilling, especially the experience of taking drugs, and the similar Rus. καῦφ, Bul. usəнсueáвane, which may convey both positive and negative emotions)
- the equivalents belong to the same grammatical class (cf. the Polish noun jazda 'go!', which is used to order people to hurry up, vs. the Bulgarian exclamation $x\acute{a}\acute{u}\acute{d}e$ or the adverbial $\acute{o}\acute{e}pso$, and the Russian adverbial $m\acute{u}\acute{e}om$)
- the collocations for all equivalents overlap (cf. Pol. jazda pociągiem 'train ride' and Rus. eзда́ на поезде, vs. Pol. jazda na tyżwach and Rus. ката́ние)
- the equivalents should not only be identical in terms of their semantic and grammatical properties, but also in terms of their pragmatics (cf. the analysis of neologisms in this paper).

Obviously, it would be quite naïve to assume that a given unit in one language has an equivalent in another language that is semantically, grammatically and pragmatically identical. Therefore, one must make an assumption that in a multilingual dictionary, one unit may have two or more equivalents in another language; or that an equivalent in another language will not be adequate (i.e. it will diverge from the base word in one of the aspects given above). The information about the non-adequacy of the equivalent ought to be given expressis verbis. Marking equally good equivalents would be problematic in a printed dictionary (which is linear), but in an electronic version, we are able to overcome this difficulty.

Of course, the electronic form is not perfect — two problems arise in the development of electronic dictionaries:

• It is essential that one sets limits for the description of lexemes. Since the form constitutes the starting point for a description, one must set the number

¹ Joanna Satoła-Staśkowiak and Violetta Koseska-Toszewa edited another dictionary based on the Polish-Bulgarian-Russian corpus: the bilingual Współczesny bułgarsko-polski słownik. The first volume of the dictionary is currently in press.

Table 1

LC 1 N E	jazda 1 N. Fem.	πετύρομο 1 Ν	
езд а́ 1 N. Fem. 'travel by means of transport, usually on the ground' езда́ на по́езде; езда́ на велосипе́де ката́ние N. Neut. 'travel by means of transport; for pleasure, or as a form of training' ката́ние на конъка́х; ката́ние верхо́м; ката́ние с горы́	'change of place of residence with the use of ground transportation (including ani- mals) or sports equipment' jazda pociągiem; jazda samocho- dem; jazda kon- na; jazda na ro- werze; jazda figu- rowa na lodzie; brawurowa jazda	пъту́ване 1 N. Neut. 'travel by means of transport, usually on the ground' (cf. also каране) пътуване с влак; каране на кола; каране на велосипед	
кавале́рия N. Fem. hist. 'equestrian military unit' лёгкая кавале́рия	jazda 2 N. Fem. hist. 'equestrian military unit' pułk jazdy	кавале́рия N. Fem. singulare tantum hist. 'equestrian military unit' Служа в кавале- рията.	kawaleria N. Fem. hist. 'equestrian military unit' oddział kawalerii; ciężka, lekka ka- waleria
καμφ N. Masc. coll. 'a pleasant and thrilling experience, usually induced by drugs' лοσάπο καμφ; ε καμφ; πο κάμφy	thrilling experience,	изживя́ване N. Neut. 'experience, either positive or negative' Готино изживяване, нали?	doświadczenie 1 N. Neut. 'what we know as a result of our experience' zdobyć doświadczenie; nabyć doświadczenie; bogate doświadczenie doświadczenie 2 N. Neut. 'actions undertaken to prove whether a theory is correct or not; experiment' przeprowadzić doświadczenie
		а́дски ADV. 'fantastic, splendid, marvellous; bloody'	and a season service

	сканда́л N. Masc. 'an argument involving shouting and physical fighting' Устро́ить сканда́л cf. выволочка, разборка	jazda 4 N. Fem. coll. 'a situation when someone is angry at someone else and expresses anger in an aggres- sive manner, but the other person is disinterested' Zrobiła mi jazdę, że za późno wróciłem do domu.	сканда́л N. Masc. 'quarrel' Ди́гам сканда́л.; Пра́вя сце́ни.	skandal N. Masc. 'an action or behaviour that causes a public feeling of shock and disapproval' wywołać skandal; uniknąć skandallu; skandal towarzyski; głośny skandal
migiem	ми́гом part.	jazda 5 N. Fem.	ха́йде part.	
ADV. coll. 'szybko, natych- miast' zrobić coś migiem	coll. 'very fast, immediately' Μύσοм οπκρώπιο οκπό.	coll. 'an exclamation used to make somebody do something faster' Jazda mi stąd!; Jazda, zabieraj te graty.	, , ,	
			бъ́рзо ADV.	szybko 1 ADV.
			'quickly, fast'	'at a fast speed'
				toczyć się szybko iść szybko
				szybko 2 ADV.
				'immediately, at once'
				zrobić coś szybko cf. migiem

of units that are characterised in advance. If a unit x in language A has an equivalent y in language B, we can assume that there are other units in the form of y, which then has its own equivalents in languages B and C. We previously assumed that languages A, B and C have the equal status; this entails that every unit identical in form to the equivalent in question must also be characterised and given an equivalent in each of the remaining languages. By using such a method, we may easily run into the problem of infiniteness, hence, it is necessary that a limit of units is set.

• One needs to avoid circularity in description. At the same time, one must strive for an exhaustive description: every unit must have an equivalent in the other languages; units must not be defined by units that have already been defined; if a unit in language A has an equivalent in language B, the unit in language B should offer a link to the unit in language A.

Modern vocabulary in the dictionary

The methodology established for the construction of the dictionary enables us to record the latest vocabulary. It also ensures that the vocabulary we include faithfully reflects the recent trends in the languages, and that the frequency of use is based on corpus data. Corpus studies give measurable frequencies for neologisms and neosemanticisms, which might later be included in the dictionary.

The authors of the dictionary feel responsible for the vocabulary they include in their work. Therefore, we adopted a methodology that enables us to work on the latest vocabulary. The methodology allows the analysis of the acceptability of a given form, its frequency of use, as well as its stylistic features².

There are a number of indicators that show whether a given lexeme already has a permanent place in a lexical system of its language:

In Russian, as well as in Polish, a lexeme fully accepted by the grammatical system receives inflectional endings, and/or when it fully enters the declension system of the language. Additionally, new words proliferate within the word family of the lexeme, and these words also receive case inflection:

- баннер, баннера, баннеры, etc. 'banner';
- оцифровка, оцифровку, оцифровке, оцифровки, etc. 'digitisation';
- сканировать, сканированию, сканирования, скан, etc. 'scan';
- клонировать клонированию клонирования клон клонирование, etc. 'clone'.

In Bulgarian, a new word (or a borrowing accepted by language users) receives articles and often, but not always, a plural form:

- банер, банери, банерите, банерен, etc. 'banner';
- цифризация, цифровизирам, цифров, цифровизиране, etc. 'digitisation';
- сканирам, сканиране, сканираш, скан, etc. 'scan';
- клонирам клониране клонираш клонирате клониран клонирането клонинг, etc. 'clone'.

In Polish, there are three indicators showing that a word has been accepted in the system: it receives the Polish inflectional suffices, it enters the Polish inflectional paradigm, and new words proliferate in the word family based on the word:

- baner, banerem, banery, banerach, etc. 'banner';
- cyfryzacja, cyfryzacji, cyfryzacja, etc. 'digitisation';
- skanować, skanuję, skanowanie, skan, etc. 'scan';
- klon klony klonować klonujemy klonujecie, etc. 'clone'.

²For more information on the methodology, cf. Satoła-Staśkowiak (2014).

The above criteria proved to be successful in determining whether a given lexeme is accepted in the system or not. Nevertheless, they must not be regarded as a perfect or universal — there exist lexemes that are accepted in the systems of these languages, even though they do not exhibit the above features. In such case (i.e. when a lexeme does not fulfil these criteria), we may treat the frequency of use of a given lexeme as the indicator of its acceptability. In Polish, for guru 'guru' and kiwi 'kiwi' have been acceptable for a long time, even though they have neither given rise to a word family, nor have they entered the inflectional paradigm of Polish (cf. Satoła-Staśkowiak, 2014).

Because each of the languages in the dictionary is developed separately, we can present additional meanings for a given lexeme. These meanings may arise as a result of such processes as colloquialisation, vulgarisation and neosemanticisation. The possibility of describing compared forms in one of the languages offers a wide range of applications, and allows us to explore the ongoing processes that take place in the language described. However, as far as neologisms are concerned, it is difficult to choose the best equivalent, because neologisms are inherently colloquial, hence stylistically marked. Due to the colloquial nature of neologisms, lexicographers must perform a very subtle and detailed analysis, before they choose an equivalent which not only has the same meaning, but also the same stylistic profile (cf. zon vs. przecietniak, średniak, szarak 'a poor man').

Table 2

Russian	Bulgarian	Polish
бана́н, -ы N. Masc.	бана́н, -и <i>N. Masc.</i>	banan, -y N. Masc.
 'tropical plant, giving elongated yellow fruit' 'the fruit of this plant' 	 'tropical plant, giving elongated yellow fruit' 'the fruit of this plant' 	 'tropical plant, giving elongated yellow fruit' 'the fruit of this plant'
мегало́л N. Masc.	зъ́бене N. Neut.	3. coll. 'wide grin'
slang. 'smile, wide grin' 'exclamation used to convey delight'	coll. 'wide grin, showing your teeth'	
Я в восто́рге! Мегало́л!	През целия си живот не съм виждал толкова много зъбене.	
го́лый ADJ. Masc.	I гол, -и ADJ. Masc.	goły, goli ADJ. Masc.
1. 'not covered with clothing; bare' 2. 'about plants and animals: devoid of fur, bristle, feather, foliage' 3. 'not covered with anything, devoid of furniture, adornments, etc.' 4. 'with no additives' 5. 'devoid of money' 6. 'a poor person' 7no meaning- 8no meaning-	1. 'not covered with clothing; bare' 2. 'about plants and animals: devoid of fur, bristle, feather, foliage' 3. 'not covered with anything, devoid of furniture, adornments, etc.' 4. 'with no additives' 5. 'devoid of money' 6. 'a poor person' 7. coll. 'an average, mediocre person' 8. coll. 'an unexperienced person'	1. 'not covered with clothing; bare' 2. 'about plants and animals: devoid of fur, bristle, feather, foliage' 3. 'not covered with anything, devoid of furniture, adornments, etc.' 4. 'with no additives' 5. 'devoid of money' 6. 'a poor person' 7. cf. coll. man in the street, Joe Bloggs 8. cf. greenhorn

Neologisation, neosemanticisation, vulgarisation, and colloquialisation indicate the nature and the scope of the processes in different (due to the divisions within the Slavdom) Slavic languages. Internal neosemanticisation is usually based on metonymy and metaphor; less frequently its source lies in clipping, generalisation and creating opposites. Let us analyse the internal neosemanticisation of the Polish word *stoik* (lit. 'jar').

Table 3

Russian	Bulgarian	Polish
ба́нка, -и N. Fem.	бурка́н, -и N. Masc.	słoik, -i N. Masc.
'a glass container used for stor- ing various substances'	'a glass container used for stor- ing various substances'	1. 'a glass container used for storing various substances'
dat. coll. лими́тчик 'a person who came to a different city or town and obtained the right to register their residence and to work in that place, because there was a need for people of particular profession'		2. coll. 'a person working in Warsaw who came to the city usually from a smaller town, for financial reasons'

External neosemanticisation is almost exclusively rooted in words of Englis origin. Let us take the example of a word, used in all three languages, i.e. Pol. kultowy, Rus. ку́льтовый, Bul. култов ('cult').

Table 4

Russian	Bulgarian	Polish
ку́льтовый ADJ.	култов АДЛ.	kultowy ADJ.
new meaning: 'an object of	new meaning: 'an object of	1. 'related to a religious cult' 2. new meaning : 'an object of cult-like veneration, also: popular'

When developing the electronic dictionary, we decided to include also addressative forms, regarded as lexical units. The Russian-Bulgarian-Polish dictionary contains a few dozen items which perform addressative functions in the languages, e.g. brat ('brother'), babcia ('grandmother'), mama ('mum'), ojciec ('father'), siostra ('sister'), or kolega ('friend').

The analysis of monolingual and bilingual dictionaries of Polish, Russian and Bulgarian yielded the following results:

- In monolingual dictionaries, the descriptions of addressative forms usually include some notes about the patterns of use of the lexeme in a given culture. Nevertheless, these notes are insufficient to use the addressative function fully in accordance with the rules accepted in the speech community.
- In bilingual dictionaries, the meanings of addressative forms are usually not included; hence, the reader is not provided with any information on the proper use of addressative forms in a given speech community.

We will now exemplify these findings with the lexeme Pol. babcia / Rus. δάδυμικα 'grandmother'. An electronic dictionary of Polish (www.sjp.pwn.pl) gives the following description for the lexeme babcia:

babcia

- 1. the mother of one's mother or father
- 2. an elderly woman

In comparison, the definition in www.gramota.ru reads as follows:

бабушка, -и; мн. род. -шек, дат. -шкам; ж.

1. The mather of one's father or mother (this word is used by grandchildren to refer to their grandmother). *Omnpasums сына на лето κ бабушке. Стать бабушкой* (when a woman becomes a grandmother) *Это твоя б.?* / Used to refer to an elder female family member.

2. Ласк.

Old, elderly woman. Cadumec, δ .

The dictionaries of Polish (items 1 and 2), and Russian (item 1) indicate a possible addressative function of the lexeme, but neither dictionary gives any information about the addressative use of the lexeme. Only item 2 in the Russian dictionary provides an example of addressative use of the lexeme.

Both languages and cultures use the lexeme as a form that one uses to address one's own grandmother. The dictionaries do not inform the reader that both cultures allow their speakers to use the form to address also the grandmother of one's husband or wife. In the Russian culture, the lexeme can be used interchangeably with the name and the patronymic ($\delta\acute{a}\acute{b}yu\kappa a$ / Japuca Heahoeha).

In conclusion, monolingual dictionaries tend to indicate addressative functions of lexemes, but they are inadequate for a language user to use the lexeme properly.

We suggest that an example entry in a dictionary, with a classifier of politeness, should look as follows:

Table 5

Russian	Bulgarian	Polish
ба́бушк а <i>N. Fem.</i>	ба́ба, -и N. Fem.	babci a, ø, -e N. Fem.
1. 'the mother of one's father or mother'	1. 'the mother of one's father or mother' (cf. стара майка)	1. 'the mother of one's father or mother' (cf. babunia)
Моя бабушка Даниела про- сто чудесна!	Моята баба Даниела е чудес- на!	Moja babcia Daniela jest wspa- niała!
	2. 'form used to address the mother of one's father or mother' neutral form of address	
3. 'the mother of one's spouse's (wife's or husband's) mother or father'	3. 'the mother of one's spouse's (wife's or husband's) mother or father'	3. 'the mother of one's spouse's (wife's or husband's) mother or father'

4. 'form used to address the 4. 'form used to address the 4. 'form used to address the mother of one's spouse's (wife's mother of one's spouse's (wife's mother of one's spouse's (wife's or husband's) mother or father' or husband's) mother or father' or husband's) mother or father? neutral form of address (cf. neutral form of address neutral form of address (voc. name and patronymic cf. Bebabciu) ра Павловна) neutral form of address used interchangeably with a name and patronymic 5. 'an old woman' (cf. crapy-5. 'an old woman' (cf. стари-5. 'an old woman' (cf. staruха, старушка) ца, стара жена; бабичка) szka, starowinka) У ключа на лавочке сидят На пейката седяха две баби Na ławce siadły dwie babcie две бабушки. уморени от разходката. zmęczone spacerem. 'form used to address an 'form used to address an 'dat. form used to address an older woman in the older woman' (cf. ctapyxa, older woman' старушка) countryside. (Today only Бабо, как да хвана автобуса colloquial; might be regarded Бабушка, далеко ли до гороза Силвен? as pejorative.)' (see pani) ∂a ? Babciu, a daleko stad do miasta?

The lack of classifiers of politeness, especially in bilingual dictionaries, results in the poor quality of translations, as well as problems in teaching foreign languages. If we create a reliable set of principles for indicating addressative meanings and politness in dictionaries, we will be able to improve the quality intercultural communication, translations and foreign language teaching. Such goal will be best achieved by creating an electronic dictionary linked to a parallel corpus.

References

- Brown, P., & Levinson, S. C. (2008). *Politeness. Some universals in language usage* (17th Ed.). Cambridge: Cambridge University Press.
- Dimitrowa, L., & Koseska, V. (2009). Bulgarian-Polish Corpus. Cognitive Studies | Études cognitives, 9, 17–132.
- Dimitrova, L., Koseska, V., Roszko, D., & Roszko, R. (2010). Application of Multilingual Corpus in Contrastive Studies (on the Example of the Bulgarian-Polish-Lithuanian Parallel Corpus). *Cognitive Studies | Études cognitives, 10, 217–240.*
- Dimitrova, L., & Koseska, V. (2012). Bulgarian-Polish Parallel Digital Corpus and Quantification of Time. Cognitive studies | Études cognitives, 12, 199–208.
- Duszkin, M., & Satoła-Staśkowiak, J. (2011). The Bulgarian-Polish-Russian Parallel Corpus. Cognitive Studies | Études cognitives, 11, 241–254.
- Gol'din, V. E. (2009). Obrashchenie. Teoreticheskie problemy. Moskva.
- Huszcza, R. (2006). Honoryfikatywność. Gramatyka. Pragmatyka. Typologia. Warszawa: Wydawnictwo Naukowe PWN.
- Kisiel, A. (2008). Synonimia wyrażeń funkcyjnych na przykładzie partykuł wyróżniających. In *Nowe Studia Leksykograficzne* (Vol. II, pp. 167–182). Kraków: Lexis.
- Kita, M. (2005). *Językowe rytuały grzecznościowe*. Katowice: Wyższa Szkoła Zarządzania Marketingowego i Języków Obcych.

- Koseska-Toszewa, V., & Satoła-Staśkowiak, J. (2013). O pracy nad słownikami książkowymi i elektronicznymi z językiem bułgarskim. In L. Krumova-TSvetkova, D. Blagoeva, & S. Kolkovska (Eds.) 70 godini bŭlgarska akademichna leksikografiia. Dokladi ot Shestata natsionalna konferentsiia s mezhdunarodno uchastie po leksikografiia i leksikologiia (pp. 98–108). Sofiia: Akademichno izdatelstvo "Prof. Marin Drinov".
- Koseska-Toszewa, V., Satoła-Staśkowiak, J., & Duszkin, M. (2012a). Polish-Bulgarian-Russian, Bulgarian-Polish-Russian or Russian-Bulgarian-Polish Dictionary? *Cognitive Studies | Études cognitives*, 12, 27–38.
- Koseska-Toszewa, V., Satoła-Staśkowiak, J., & Duszkin, M. (2012b). Teoreticheskie i metodologicheskie aspekty sozdaniia pol'sko-bolgarsko-russkogo slovaria. In Prykladna linhvistyka ta linhvistychni tekhnologii, MegaLing-2011 (pp. 213–222). Kyïv: NAN Ukraïny, Ukr. movno-inform. fond.
- Koseska-Toszewa, V., Satoła-Staśkowiak, J., & Sosnowski, W. (2013). From the problems of dicionaries and multi-lingual corpora. *Cognitive Studies | Études cognitives, 13,* 113–122. doi: 10.11649/cs.2013.007.
- Marcjanik, M. (2002). Polska grzeczność językowa. Kielce: Wydawnictwo Akademii Świętokrzyskiej.
- Marcjanik, M. (2007). Grzeczność w komunikacji językowej. Warszawa: Wydawnictwo Naukowe PWN.
- Roszko, D. (2012). Semantic Contrastive Linguistics Theory and Dialectological Studies. Cognitive Studies/Études cognitives, 12, 79–92.
- Roszko, R. (2012). Lexical Exponents of Hypothetical Modality in Polish and Lithuanian. Cognitive Studies | Études cognitives, 12, 15–25.
- Satoła-Staśkowiak, J. (2013a). Neologizmy bułgarskie, polskie i rosyjskie w ujęciu konfrontatywnym. In D. Blagoeva, S. Kolkovska, & M. Likosa (Eds.), *Problemi na neologiiata v slavianskite ezitsi* (pp. 21–30). Sofiia: AI "Prof. Marin Drinov".
- Satoła-Staśkowiak, J. (2013b). Polskie i bułgarskie neologizmy znaczeniowe. In D. Blagoeva, S. Kolkovska, & M. Likosa (Eds.), *Problemi na neologiiata v slavianskite ezitsi* (pp. 218–230). Sofiia: AI "Prof. Marin Drinov".
- Sosnowski, W. (2013). Forms of Address and their Meaning in Contrast in Polish and Russian Languages. *Cognitive Studies | Études cognitives*, 13, 225–235. doi: 10.11649/cs.2013.015.
- Tomiczek, E. (1983). System adresatywny współczesnego języka polskiego i niemieckiego. Socjolingwistyczne studium konfrontatywne. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Tomiczek, E. (1992). Z badań nad istotą grzeczności językowej. In J. Anusiewicz, & M. Marcjanik (Eds.) *Język a kultura* (Vol. 6: Polska etykieta językowa). Wrocław: Wiedza o Kulturze.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (http://creativecommons.org/licenses/by/3.0/pl/), which permits redistribution, commercial and non-commercial, provided that the article is properly cited. © The Author(s) 2014.

Publisher: Institute of Slavic Studies PAS & University of Silesia in Katowice