

LUDMILA DIMITROVA^{1,A}, VIOLETTA KOSESKA-TOSZEWA^{2,B}
JOANNA SATOŁA-STAŚKOWIAK^{2,C}

¹Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

²Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland

^Aludmila_dim2@yahoo.com , ^Bamaz@inetia.pl , ^Cjoannasat@op.pl

NEOLOGISMS IN BILINGUAL DIGITAL DICTIONARIES (ON THE EXAMPLE OF BULGARIAN-POLISH DICTIONARY)

Abstract

The paper discusses the presentation of neologisms in the recent version of the Bulgarian-Polish digital dictionary. We also continue the discussion of important problems related to the classifiers of the verbs as headwords of the digital dictionary entries. We analyze some examples from ongoing experimental version of the Bulgarian-Polish digital dictionary.

Keywords: Bulgarian, Polish, classifiers, dictionary entry, digital dictionary, headword, neologism, online dictionary.

1. Introduction

We have started to develop the first Bulgarian-Polish digital dictionary under the joint research project¹ between the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences and the Institute of Slavic Studies, Polish Academy of Sciences, coordinated by L. Dimitrova and V. Koseska-Toszewa. No bilingual digital resources with Bulgarian and Polish existed ever before.

In our previous works (Dimitrova, Koseska-Toszewa 2008a), (Dimitrova, Koseska-Toszewa 2008b), (Dimitrova, Koseska-Toszewa 2009a), (Dimitrova, Koseska-Toszewa, Satoła-Staśkowiak 2009) we discussed some problems related to the classifiers of headwords in the digital dictionary entries, especially with the focus on their harmonization and standardization in accordance with the semantics features of both languages. In this article we discuss the presentation of neologisms in the Bulgarian-Polish digital dictionary.

¹Semantics and Contrastive linguistics with a focus on a bilingual electronic dictionary.

1.1. Why we choose Bulgarian and Polish for development of bilingual digital dictionary

Bulgarian and Polish have been purposefully chosen for the following reasons: 1) there are no digital dictionaries for these languages, 2) there are no parallel corpora for these languages for supporting such dictionaries, and 3) the language pair [Bulgarian, Polish] is representative — each language represents sub-group of the Slavic language families, Bulgarian belongs to the South-Slavic and Polish to the West-Slavic language family.

1.2. Short overview of printed Bulgarian-Polish dictionaries

Firstly, we would like to note that in the past 25 years neither Bulgarian-Polish nor Polish-Bulgarian dictionaries have been published. The market in both countries is saturated with English-Polish or Polish-English and English-Bulgarian or Bulgarian-English printed dictionaries. The available Bulgarian-Polish/Polish-Bulgarian printed dictionaries are only few, let's mention them here. The first printed Polish-Bulgarian dictionary has been prepared by Ivan Lekov and published in 1944 (Lekov 1944). The second one (Lekov, Sławski, Eds. 1961) has been published 52 years ago. Both aforementioned dictionaries are a bibliographic rarity.

Two printed dictionaries were available recently: Bulgarian-Polish dictionary by Franciszek Sławski (Sławski 1987) and Polish-Bulgarian dictionary by Sabina Radeva (Radeva 1988). Their circulation is approx. 6700 each, their volume is approximately 50 000 words, and they are more or less equivalent in terms of lexical content. For our purposes, however, both dictionaries have several disadvantages. First, as they were published 25 years ago. Second, the dictionaries were created for a specific audience: students of Slavistics and translators of fiction, which is reflected in their contents. Both dictionaries do not always contain the translated correspondences, i.e. sometimes instead of a Polish word, there is a definition, which interprets the meaning of a Bulgarian word. Furthermore, both dictionaries contain many outdated words and expressions, which are no longer used in Bulgarian language, e.g. dialectisms or **loan-words** from Turkish. Since prefixing in Bulgarian is still a productive method to create new words, there are many synthetic words, which could formally belong to a certain word-formation group, for example: *позостя, погощавам, позабързам, подвзема, подвземам, подгордея се, подгордявам се*, and others. However, some of these words may not have been in use 40 years ago or at present.

That is why we would use the above-mentioned dictionaries not as a primary source, but as reference, although one could say that the sources for a digital dictionary with Bulgarian and Polish are the printed dictionaries.

2. Headwords Selection Procedure

One of the initial tasks in our project was the selection of Bulgarian headwords. The applied method follows the method, statistical and linguistic at the same time, developed for CONCEDE project² and described in (Tufis et al. 1999). The procedure for selecting the headwords take into account word frequency, word class, and

²PL96-1142 INCO-COPERNICUS project CONCEDE *Consortium for Central European Dictionary Encoding*.

the number of words there were in a given word-class and word-frequency band. The text used was encoded as CES ANA, (Ide 1998), which specifies for each word-form its associated lemma and grammatical information. Such Bulgarian text was developed in the MULTEXT-East project³ (Dimitrova et al. 1998). The POS composition of the selected headwords set has to reflect the corresponding distribution of the different POS in the Bulgarian MULTEXT-East corpus. First 500 lemmas were chosen for the relevant ten grammatical categories identified in the MULTEXT-East project, according to the frequency of their occurrence in corpus. Selected Bulgarian headwords were verified for correspondence with the frequency list of the Bulgarian-Polish corpus. Next Bulgarian headwords were chosen from Bulgarian-Polish corpus (Dimitrova, Koseska 2009b) by the same method. Approximately 2000 Bulgarian neologisms are also added to the set of the headwords. The Bulgarian and Polish use different character sets: Cyrillic — for Bulgarian, and Latin with some special diacritic symbols — for Polish. That’s why the lexical database of the digital dictionary (Dimitrova, Panova, Dutsova 2009) uses encoding scheme defined in Unicode 8. For more detailed description of the headwords selection we refer to the paper of Dimitrova, Dutsova, (in this volume).

3. Coined New Word in Bulgarian-Polish Digital Dictionary

Here we’ll concentrate our attention on the **presentation of newly coined words (neologisms)** in the bilingual Bulgarian-Polish digital dictionary.

3.1. Which words are treated as “neologisms”?

One could find in literature the next explanations of a “neologism”: (1) a new (a newly coined) word, meaning, usage, or phrase; (2) the introduction or use of new words or new senses of existing words (e.g. familiar word used in a new sense).

So, a neologism is a new word for a notion describing abstract or concrete objects in a logical sense. Since neologisms are most often loans from a foreign language, it is necessary to distinguish between a loanword and a neologism. A loanword is considered a neologism only if it has been embedded in the language system so far as to be commonly accepted by the language users or by the users from a certain area (professional slang). Many studies on neologisms do not make this distinction and loanwords are “labeled” neologisms, which we consider a methodological and theoretical error of scholars of neologisms.

There are different ways for neologisms to appear. For example in Kashubian (or Cassubian) neologisms are coined in a way to stand apart from Polish lexis. This is due to the relatively recent and ongoing process of creation of a Kashubian literary language (Popowska-Taborska 2007, 2011).

3.2. In our study we consider the following types of neologisms:

3.2.1. New words, related to new abstract or concrete notions for objects in general use or professional slang

The introduction of new technical machinery in science, technology or everyday life

³EU COP Project 106 MULTEXT-East *Multilingual Text Tools and Corpora for Central and Eastern European Languages*.

is one way for new words to enter a language. For example, the originally accepted (from the **loan translation** of the English *automatic computing machine* “**ACM**”) Bulgarian name *електронно-изчислителната машина* “**ЕИМ**” is first used in professional lexis serving the computing technology and computational mathematics. However, it was soon replaced by the word “**компютър**” /**computer**/. Similarly, its original Polish name was *elektroniczna maszyna cyfrowa* “**ЕМС**”, also a **loan translation** of *automatic computing machine* “**ACM**”, but in the spoken language the shorter word “**komputer**” /**computer**/ was preferred, which replaces “elektroniczna maszyna cyfrowa”.

This example reveals **two facts** about loaning foreign words in a given language:

First, in Bulgarian and in Polish **ЕИМ** и **ЕМС** are loan-translations of the English **ACM**, but both languages use the foreign word “computer”. As well known a **loan translation** means an expression or combination of expressions created by means of the native language but based on a loaned semantic model of a foreign language. Loan-translations can be:

- “lexical” (an exact translation of the foreign lexical model, for example the Polish *listonosz* is a loan translation of the German *Briefträger*, the Bulgarian *високоговорител* is a loan translation of the German *Lautsprecher*);
- “grammatical” or “syntactic” (for example the Polish *Wydaje się być* is a loan translation of the English *Seem to be*);
- “phrasal” (for example, the Bulgarian *убих времето* and the Polish *zabijać czas* are loan translations of the French *tuer le temps*), see (Polański 1999, page 284), the Bulgarian *съединението прави силата* is a loan translation of the French *l’union fait la force*.

Second, the process of neologism creation in a given language confirms the observation of many linguists that shorter forms prevail in the naming of new terminology (like **компютър** in Bulgarian and Polish, **T-shirt** in Polish) (Stieber 1973, 1974). Since the term “bawełniany podkoszulek z krótkim rękawem” is expressed in Polish with too many words, the English word T-shirt, which is shorter, is widely used in Polish and so we may consider it a neologism instead of a loan-word.

Some examples of neologisms in our dictionary follow:

аеробика *f* aerobik *m*; **записах се на** ~ записаłam się na aerobik
блог *m* blog *m*; **защо не сложиш снимки в ~а си?** dlaczego nie wrzucisz zdjęć na swój blog?
банкомат *m* bankomat *m*; **ще изтеглим пари от най-близкия** ~ weźmiemy pieniądze z najbliższego bankomatu
гигабайт *m* gigabajt *m*; **колко ~а са свободни?** ile jest wolnych gigabajtów?
диск *m* dysk *m*; **~ът е пълен** dysk jest wypełniony
компютър *m* komputer *m*; **изключи ~а!** wyłącz komputer!
лаптоп *m* laptop *m*; **имама нов** ~ mam nowy laptop

3.2.2. Neosemantisms

In this group belong known words which have acquired a new meaning. Some examples of neosemantisms in our dictionary are:

- блистер** m listek m, **ще купя само един ~ от тези таблетки** kupię tylko jeden listek tych tabletek
джиесем m komórka *f*; **забравих си ~а вкъщи** zapomniałam wziąć z domu komórke
икон|а, -и *f* obraz święty, ikona *f*; *pot. komp.* **компютърна ~а** ikona (ikonka komputerowa)
колон|а, -и *f* kolumna *f*; *pot. tech.* **включи новите ~и** podłącz nowe kolumny
купон, -и m kupon m; bon m; с ~ и na bony, na talony, na kartki; *pot.* **импреза** *f*, zabawa *f*; **отивам на ~** idę na imprezę; **у тях пак има ~** u nich znowu jest impreza
маймунско “а” *pot. inform.* małpa
мишк|а, -и *f* zool. mysz *f* (Mus); **полска ~а** mysz polna; *pot. komp.* **компютърна ~а** mysz komputerowa
мреж|а, -и *f* sieć *f*, siatka *f*; **~а за покупки** siatka na zakupy; **железопътна ~а** sieć kolejowa; **телени ~и** woj. zasięki; *pot. inform.* **интернет-~а** sieć internetowa; **нямам ~а** nie mam Internetu ; **сърфирам в ~та** serfować w sieci

Listek and *komórka* are examples of neosemantisms in Polish. Neosemantisms can also be complex words formed from known words, for example:

- автокъща** *f* autokomis *m*; **в тази ~ се намират и хубави коли** w tym autokomisie znajdują się także ładne samochody
антивирус|ен, -на, -но *adi.* antywirusowy; **~на вакцина** antywirusowa szczepionka; **~на програ'ма** antywirusowy program
ветрогенератор m generator wiatru; **тук няма ~и, защото са опасни за птиците** tutaj nie ma generatorów wiatru, ponieważ są niebezpieczne dla ptaków

4. Syntactic and semantic classifiers in the Bulgarian-Polish electronic dictionary

As mentioned in (Dimitrova, Koseska, Satoła, 2009), transitive and intransitive *syntactic classifiers* have been introduced. In this case *transitivity* refers to the usage of nouns as direct objects following the verbal form.

The Polish transitive verbs are always followed by the accusative case of nouns or adjectives. In Bulgarian transitive verbs are always followed by a direct object, Bulgarian lacks a nominal declination. In Bulgarian intransitive verbs are followed only by an indirect object, but in Polish intransitive verbs are followed by any case of nouns or adjectives except the accusative case.

The *semantic classifiers* introduced in the dictionary are related to the aspect forms of the verb. Traditional dictionaries employ the aspect classifier /**aspect**/ with values “*несвършен вид*“ and „*свършен вид*“ without distinguishing form from

meaning; this problem has been addressed in (Dimitrova, Koseska 2009a). The choice of these semantic classifiers is motivated by the well-known fact that the aspect category is formalized by a paradigm only in the Slavic languages (Koseska, Satoła, Duszkin 2012, in this volume).

The verbal form is perfect or imperfect but its meaning can be *state* or *event*, where a *state* is a state or sequence of states and events concluding with a *state*, while an *event* is an event or a sequence of events concluding with an *event*. The notions event and state are based on works where the description of tense and aspect are based on Petri net theory, adapted to the natural language by A. Mazurkiewicz (Mazurkiewicz 1986, 2008), (Koseska, Mazurkiewicz 2009, 2010), (Koseska 2006), (Satoła 2010).

We have introduced in our dictionary a new semantic classifier to mark the meaning of the verbal form with values **state** and **event**. The main characteristic differentiating between these notions is the temporal continuity of states and the instantaneity of events. In other words, **states** „last”, whereas **events** can only „happen”.

The correspondence between **state** and **event** can be visualized by an abstract comparison between a segment from the number axis (**state**) and a point lying on the segment (**event**).

The following are some examples illustrating the new verbal classifiers:

влиза|м, -ш *vi. state, intransitive*; wchodzić *intransitive*; **~м в университета** wstępuję na uniwersytet; **~м в грях** popełniam grzech; **~ м в сила** nabierać ważności, nabierać mocy

вляза, влезеш *vp. event, intransitive*; wejść *intransitive*; **да ~ ли?** czy mogę wejść?

разгледа|м -ш *vp.event, transitive*; obejrzeć, przejrzeć, ropatrzeć *transitive* v. **разглеждам**

разглежда|м, -ш *vi. state, transitive*; oglądać, przeglądać; rozpatrywać *transitive* **доближ|а, -иш** *vp. event, intransitive*; zbliżyć *intransitive*; **~а се** *aux. zbliżyć się* v. **доближавам се**

доближава|м, -ш *vi. state, intransitive*; zbliżać, przybliżać *intransitive*; **~м се** *aux. zbliżać się, przен. upodabniać się*

The next two examples show verbs — neologisms from both languages:

глобализир|ам, -аш, *vi. state, transitive*; globalizować *transitive*; **интернет бързо ~а всеки проблем** Internet szybko globalizuje każdy problem

сърфир|ам-аш, *vi. state, intransitive*; serfować *intransitive*; **~ в мрежата** serfować w sieci

5. Conclusion

In this paper we have discussed the problems of creation of new words in Bulgarian and Polish and their usage in Bulgarian-Polish digital dictionary. Some English verbs have exact Bulgarian correspondences, for example, **download** means *за-реждам (от), копирам (от), прехвърлям (от)* but its usage in Bulgarian as

даунлоадвам (download**вам**) is not accepted as neologisms. The usage of such forms is not acceptable neither in written, nor in spoken language.

We have also discussed new dictionary entry classifiers we have proposed for more adequate description of the verbs. Such new classifiers must reflect the specific characteristics of the compared languages, for example the aspect classifier (with values transitivity or intransitivity) is important for the syntax of both languages, but is much more important on the morphologic-syntactic level for Polish, a synthetic language, in contrast to Bulgarian, an analytic language like English.

References

- Dimitrova et al. (1998)**. Dimitrova, L., Erjavec, T., Ide, N., Kaalep, H.-J., Petkevic, V., and Tufis, D. Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages. In: *Proceedings of COLING-ACL '98*. Montréal, Québec, Canada, 315–319.
- Dimitrova, L., Koseska-Toszewa, V. (2008a)**. The Significance of Entry Classifiers in Digital Dictionaries. In: *Proceedings of the MONDILEX First Open Workshop*. Moscow, Russia, 3–4 October 2008, 89–97, Russian Academy of Sciences, IITP.
- Dimitrova, L., Koseska-Toszewa, V. (2008b)**. Some problems in multilingual digital dictionaries. *Cognitive Studies / Études Cognitives*. Vol. 8, SOW, Warsaw, 237–254.
- Dimitrova, L., Koseska, V. (2009a)**. Classifiers and Digital Dictionaries. *Cognitive Studies / Études Cognitives*. Vol. 9, SOW, Warsaw, 117–131.
- Dimitrova, L., Koseska, V. (2009b)**. Bulgarian-Polish Corpus. *Cognitive Studies / Études Cognitives*. Vol. 9, SOW, Warsaw, 2009, 133–141.
- Dimitrova, L., Koseska, V., Satola-Staskowiak, J. (2009)**. Towards a Unification of the Classifiers in Dictionary Entries. In: Garabík, Radovan (Editor, 2009). *Metalinguage and Encoding Scheme Design for Digital Lexicography. Proceedings of the MONDILEX Third Open Workshop, Bratislava, Slovakia, 14–15 April 2009*. Tribun, Brno, 48–58.
- Dimitrova, L., Dutsova, R. (2012)**. Implementation of the Bulgarian-Polish Online Dictionary. (In this volume)
- Dimitrova, L., Panova, R., Dutsova, R. (2009)**. Lexical Database of the Experimental Bulgarian-Polish online Dictionary. In: Garabík, Radovan (Editor, 2009). *Metalinguage and Encoding Scheme Design for Digital Lexicography. Proceedings of the MONDILEX Third Open Workshop, Bratislava, Slovak Republic, 15–16 April 2009*. Tribun, Brno, 36–47.
- Ide, N. (1998)**. Corpus Encoding Standard: SGML guidelines for encoding linguistic corpora. In: *Proceedings of First International Conference on Language Resources and Evaluation, LREC'98*, Granada, ELRA, 463–470.
- Koseska-Toszewa, V. (2006)**. *Gramatyka konfrontatywna bułgarsko-polska*. vol. 7, *Semantyczna kategoria czasu*, Warszawa, SOW, 210 pages. (In Polish)
- Koseska, V., Mazurkiewicz, A. (2009)**. Net-Based Description of Modality in Natural Language (on the Example of Conditional Modality). In: Shyrov, Dimitrova (Eds. 2009), *Organization and Development of Digital Lexical Resources. Proceedings of the MONDILEX Second Open Workshop, Kiev, 2–4 February, 2009*, 98–105.
- Koseska-Toszewa, V., Mazurkiewicz, A. (2010)**. *Time Flow and Tenses*. SOW, Warsaw, 223 pages.

- Koseska-Toszewa, V., Satoła-Staškowiak, J., Duszkin, M. (2012).** Polish-Bulgarian-Russian, Bulgarian-Polish-Russian or Russian-Bulgarian-Polish Dictionary? (In this volume)
- Lekov, Ivan (1944).** *Полско-български речник*. BAS Publishing House, Sofia, Bulgaria.
- Lekov, I., Sławski, F. (Eds.) (1961).** *Полско-български речник*. BAS Publishing House, Sofia, Bulgaria.
- Mazurkiewicz, A. (1986).** Zdarzenia i stany: elementy temporalności, In: *Studia grammatyczne bułgarsko-polskie*, vol. I, *Temporalność*, Wrocław, 7–21. (In Polish)
- Mazurkiewicz, A. (2008).** A Formal Description of Temporality (Petri net approach). In: Iomdin, Dimitrova (Eds. 2008), *Lexicographic Tools and Techniques. Proceedings of the MONDILEX First Open Workshop, Moscow, 3–4 October, 2008*. 98–108.
- Polański, K. (1999).** *Encyklopedia językoznawstwa ogólnego*. Ossolineum, Wrocław (In Polish)
- Popowska-Taborska H. (2007).** *Leksyka powstającego kaszubskiego języka literackiego* In: *Z polskich studiów slawistycznych*, seria XI, *Językoznawstwo*, Warszawa, 173–180. (In Polish)
- Popowska-Taborska H. (2011).** Gdyby kiedykolwiek w przyszłości miał powstać Słownik etymologiczny kaszubskiego języka literackiego... *Studia z Filologii Polskiej i Słowiańskiej*, vol. 46. (In Polish) (to appear)
- Radeva S., (1988).** *Podręczny słownik Polsko-Bułgarski z suplementem*. 2nd edition, Warszawa, Polska.
- Rozko R. (2008).** Innowacje podlegają unifikacji, In *Zofia Rudnik-Karwatowa (Ed.), Współczesna komunikacja językowa*, *Prace Slawistyczne* 125, Instytut Slawistyki Polskiej Akademii Nauk, Slawistyczny Ośrodek Wydawniczy, Warszawa: 239–251.
- Satoła-Staškowiak J. (2010).** *Polsko-bułgarskie odpowiedniości przekładowe czasów przeszłych*, Instytut Slawistyki PAN, SOW, Warszawa, 144. (In Polish)
- Sławski F., (1987).** *Podręczny słownik Bułgarsko-Polski z suplementem*. 2nd edition, Warszawa, Polska.
- Stieber, Z. (1974).** *Świat językowy Słowiań*. Warszawa, PWN, 145–146. (In Polish)
- Stieber, Z. (1973).** *Zarys gramatyki porównawczej języków słowiańskich. Fleksja werbalna*, Warszawa (In Polish).
- Tufis, D., Rotariu, G., Barbu, A.-M. (1999).** Data sampling, lemma selection and a core explanatory dictionary of Romanian. In: *Proceedings of COMPLEX'99*, Pecs, Hungary, 219–228.