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

\textsuperscript{1}Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
\textsuperscript{2}Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland
\textsuperscript{A}ludmila.dim2@yahoo.com , \textsuperscript{B}amaz@inetia.pl , \textsuperscript{C}joannasat@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 project\textsuperscript{1} 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.

\textsuperscript{1}Semantics 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: погостя, 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 project\(^2\) and described in (Tufis et al. 1999). The procedure for selecting the headwords take into account word frequency, word class, and

\(^2\)PL96-1142 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 project (Dimitrova et al. 1998). The POS
composition of the selected headwords set has to reflect the corresponding distrib-
ution 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. Approx-
imately 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 encod-
ing 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 form 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

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-изчислителна машина “EIM” 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 “EMC”, 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 EMC 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 высокоговорител 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 време и the Polish zabijać czas are loan translations of the French tuer le temps), see (Polański 1999, page 284), the Bulgarian свядинето 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:

аэробика f aerobik m; записах се на ~ записалах си на aerobik
блог m blog m; защо не сложиш снимки в ~া си? dlaczego nie wrzucisz zdjęć na swój blog?
банкомат m bankomat m; ще изтегли пари от най-близкия ~ ве знам то от най-близкия bankomat
гигабайт m gigabajt m; колко ~å са свободни? ile jest wolnych gigabajtów?
диск m dysk m; ~т е пълен dysk jest zapełniony
компютър m komputer m; изключи ~å! wyłącz komputer!
лаптоп m laptop m; имам 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
- джисем m komórka f; забравих си ~а вкъщи zapomniałem w domu komórkę
- икон(a), -и f obraz święty, ikona f; pot. komp. komputerowa ~a ikona (ikonka komputerowa)
- колон(a), -и f kolumna f; pot. tech. wzmacniacz m; включи новите ~и подłącz nowe kolumny
- купон, -и m kupon m; bon m; с ~ и на bony, na talony, na kartki; pot.impreza f, zabawa f; отивам на ~ idę na imprezę; у тях пак има ~ u nich znowu jest impreza
- маймунско “а” pot. inform. małpa
- миш(a), -и f zool. mysz f (Mus); polska ~a mysz polna; pot. komp. komputerowa ~a mysz komputerowa
- мреж(a), -и f sieć f, siatka f; ~а за покупки siatka na zakupy; железопътна ~a sieć kolejowa; телени ~и woj. zasieki; pot. inform. internet ~a sieć internetowa; нямам ~a nie mam Internetu; сърфим в ~та serfowa 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 в тъм autokomisie znajdują się takie ładne samochody
- антивирусен, -на, -но adi. antywirusowy; ~на ваксина antywirusowa szczepionka; ~на програ’am 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, Satola, 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:

* влизам, -иш vi. state, intransitive; входим intransitive; входит в университет
* влизам, влизаш vp. event, intransitive; войти intransitive; да ~ ли? czy mogę wejść?
* разглеждам, -иш vp. event, transitive; наблюдать, присматриваться
* разглежда вързано ~а всеки проблем Internet szybko globalizuje każdy problem
* доближавам, -иш vi. state, intransitive; приблизить

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

* глобализирам, -аш, vi. state, transitive; globalizować transitive; интернет
* съфирям, -аш, vi. state, intransitive; сферовать intransitive; ~ в мрежата серфовать в 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-режда́м (от), копи́рам (от), пре-връщам (от) but its usage in Bulgarian as
Neologisms in Bilingual Digital Dictionaries ...

daунлоадвам (downloadvam) 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


L. Dimitrova, V. Koseska-Toszewa, J. Satoła-Staśkowiak

Koseska-Toszewa, V., Satoła-Staśkowiak, J., Duszkin, M. (2012). Polish-Bulgarian-Russian, Bulgarian-Polish-Russian or Russian-Bulgarian-Polish Dictionary? (In this volume)


