WOJCIECH PAWEŁ SOSNOWSKI[1,A], **DIANA BLAGOEVA**[2,B], & **ROMAN TYMOSHUK**[1,C]

[1] Institute of Slavic Studies, Polish Academy of Sciences, Warsaw
[2] Institute for Bulgarian language, Bulgarian Academy of Sciences, Sofia, Bulgaria
[A]w.sosnowski@uw.edu.pl ; [B]d.blagoeva@ibl.bas.bg ; [C]roman.tymoshuk@gmail.com

NEW BULGARIAN, POLISH, AND UKRAINIAN PHRASEOLOGY AND LANGUAGE CORPORA

Abstract

This article examines phraseological innovations in the Bulgarian, Polish and Ukrainian languages. Particular attention is paid to trends in the development of phraseology and to the sources of the enrichment of the phraseology of the three studied languages. The role of corpus technologies in research on language dynamics is described.

Keywords: new phraseology; Bulgarian language; Polish language; Ukrainian language; electronic corpora

1 Introduction

Over recent decades the growth and expansion of the phraseological stock has played a prominent role in the lexis development dynamics of Slavic languages (Mokienko, 2003; Georgieva & Velichkova, 2008; Styshov, 2005, 2015; Pajdzińska, 2013; Nedkova, 2017 etc.). Neologisation phenomena in the field of phraseology have yet to be an object of research, both with respect to the individual languages and from a comparative perspective. Research into the sources and the mechanisms underlying the development of new phraseological units is crucial to uncovering the full extent of the developmental tendencies governing contemporary Slavic languages.

The study in question outlines some preliminary observations on phraseological innovations in three languages: Bulgarian, Polish and Ukrainian, which respectively represent the three major groups of the Slavic languages — Southern, Western, and Eastern. The objective of the study is to establish the most important inter-lingual similarities and differences between the current tendencies determining the development of the phraseological systems of the three languages. In doing so, the analysis adheres to the common theory of the inherent characteristics of phraseological units and draws on V. Mokienko's thesis that new phraseology consists of newly coined, newly borrowed or transformed set expressive lexical combinations, paroemias, catchwords and

phrasemes (Mokienko, 2003, p. XI). The article will also examine the opportunities for the observation and description of innovative phenomena in phraseology offered by contemporary language technology and resources, namely electronic language corpora.

Extensive linguistic material has been collected for the purposes of the research. It has been excerpted from various sources: contemporary monolingual, bilingual and multilingual dictionaries, dictionaries of neologisms, text material from electronic corpora, text material illustrating the language of electronic communication, advertising slogans, headlines and texts from periodicals, dialogues from contemporary television series and films, contemporary songs, etc.

2 Observations upon new Bulgarian, Polish and Ukrainian phraseology

Very much like all living organisms, language develops and it allows speakers to assign names to new phenomena. This can be done through the coinage or borrowing of new lexical units or the development of new meanings for existing words, but it can also be achieved through the expansion of the phraseological stock with new units. The acceleration of neophraseologisation processes in the Slavic languages over the past decades has been governed primarily by the need to linguistically define and assess newly emerged or re-emerging phenomena from various spheres of public life. According to N. Alefirenko and N. Semenenko, phraseological neonomination can be interpreted in a number of ways: it is the creative process of a linguistic consciousness which is reorienting; it represents an adaptation to new value-based and meaning-based priorities; it can be viewed as a peculiar nominative response to rapid changes in the sociocultural sphere, its major drive being the formation of a new concept through the condensation of the semantic content of the symbol (Alefirenko & Semenenko, 2009, p. 227).

Since the end of the 20th century the new realities in Bulgarian socio-political and economic life have produced phraseological word combinations such as *бунт на мравките*, *син кинжал* 'a Union of Democratic Forces hardliner', *червени бабички*, *добре облечен бизнесмен* 'organised crime group leader', *преяждам с власт*, *път към Европа*, *заиграва червената* (*синята*, *жълтата*) *метла*, *спускам информационна завеса*, *слагам на тезгяха* 'privatise', *данъчна секира*. Some of these word combinations have been falling into disuse due to the loss of currency of the phenomena and concepts they denote (Georgieva & Velichkova, 2008, p. 286).

Polish language examples of newly coined phraseological units are the set word combinations *biała szkoła* 'a winter school trip for recreational and educational purposes, which is usually several days long', *z niższej/wyższej półki* 'something of high/low quality', *rozbierana randka* 'a meeting of two people whose main purpose is to engage in sexual intercourse'. Some archaic phraseological units are also gaining currency. One such example is the well-known phraseological unit *jechać/jeździć na saksy* ('travel abroad to earn money'). The expression dates back to the 19th century: in Partition-era Poland, after the abolition of obligatory unpaid peasant labour service, Poles from Galicia and Congress Poland, as well as in the territories held by Prussia, started travelling abroad in large numbers to take up seasonal employment in Germany, mainly in Saxony. Hence, the familiar expression *wyjazd na saksy*. Nowadays it has regained popularity to reflect a new reality: with Poland's accession to the European Union the economic migration destination points have increased to include not only Germany but also Ireland, the UK and other EU countries.

In the Ukrainian language various historical and socio-political factors have paved the way for phraseological units reflecting concepts, phenomena and events related to international relations, world politics, and the socio-political situation in Ukraine and other countries. Examples of such phraseological neologisms are the set word combinations *небесна сотня* 'the patriots who died in the 2013–2014 protest rallies in Ukraine', *зоряна війна* 'interplanetary armed conflict', *диванні партії* 'political parties which exist on paper only', *зелений коридор* 'an evacuation route opened to allow civilians to leave a danger zone or a combat zone'. The aforementioned examples of

phraseological units in the three languages can be defined as 'socio-political' as they originate and function mainly in media discourse.

Besides socio-political discourse, advertising discourse can also produce phraseological innovations (Bondzolova, 2015; Lusińska, 2010). Popular television commercials have given rise to the following new phraseological expressions *връхът на сладоледа*, *вдигам самолета* in Bulgarian, *a świstak siedzi i zawija w te sreberka*, *podaruj sobie odrobinę luksusu*, *z pewną taką nieśmiałością*, *prawie robi wielką różnicę* in Polish.

In recent years lexical borrowing and loan translations (calques) from Western European languages have played a major role in the growth of the phraseological stock of the three languages, with English being the principal source (Styshov, 2015). These influences should be attributed to globalisation tendencies and the expansion of international relations, which result in the enrichment of the international phraseological stock. Illustrative examples of new phraseological calques are the following set word combinations in the three languages: *пране на мръсни пари*, *pranie brudnych pieniędzy*, *відмивання брудних грошей* 'legalisation of illegally obtained funds' (originating from the English *money laundering*), *сапунена опера*, *opera mydlana*, *мильна опера* (originating from the English *soap opera*), *мокра поръчка* 'a contract murder' (originating from the English *wet order*), *холивудска усмивка*, *голлівудська посмішка* 'a wide, dazzlingly white-toothed smile' (originating from the English *Hollywood smile*), *п'ята влада* 'organised crime in a country' (originating from the English *fifth power*), also *to zrobiło mój dzień* (originating from the English *it made my day*), *черен петък*, *чорна п'ятниця* 'a huge sales day' (originating from the English *Black Friday*), etc. The new phraseological calques usually have a similar lexical structure and the same figurative core meaning in the three languages. However, in some cases there are variations cf. *мозъчна атака* in Bulgarian and *burza mózgów* in Polish (origin from English *brainstorm*). The borrowing of phraseologisms from other languages is rare compared to loan translations and is typical of Polish (*patchwork family*, *biznes jest biznes*, *American dream*) and Ukrainian (*олд скул* 'classical style, a classic' example of origin from English *old school*).

In the period following the political and social changes in Bulgaria, Poland and Ukraine there has been an expansion of colloquial language (Kita, 1991) which some researchers have defined as *a stylistic revolution* (Videnov, 1997). A representative example of the tendency towards colloquialisation is the increase in the number of colloquial and slang-based phraseologisms. The material in *The dictionary of active Polish and Ukrainian phraseology* [*Leksykon aktywnej frazeologii polskiej i ukraińskiej*] (Tymoshuk, Sosnowski, Jaskot, & Ganoshenko, 2018)[1] reveals the same tendency. A contrastive analysis of the material shows that phraseological units from the lower lexical strata have entered mainstream Polish and Ukrainian: *woda sodowa uderzyła do głowy/sodówka uderzyła*, *albo rybki, albo akwarium/albo rybka, albo pipka*; *до лампочки/до лампи*, *не мати клепки (в голові)/клепки повилітали*, *п'яний в дрова/в зюзю*, *Богом забуте місце/Богом забута діра*. Such types of phraseologisms operate on the border of standard use and are usually considered to be substandard and rude. In Bulgarian there has been an expansion in the use of a number of colloquial phraseological neologisms such as *гушвам букетчето (босилека)*, *духам супата*, *избивам (изтрепвам) рибата*.

Phraseological units from various sociolects have been actively entering the phraseological stock of the three languages. Youth slang has been the source of phraseological expressions such as *кърша мивки*, *цепя мрака* 'create a strong impression with one's qualities or behaviour' in Bulgarian, *być jazzy*, *być trendy*, *być na gigancie*, *dawać sobie w żyłę* in Polish. Over the last decades the following Ukrainian slang-based phraseologisms have been growing in popularity: *бути в темі*

---

[1] *The dictionary of active Polish and Ukrainian phraseology* presents the semantic counterparts of phraseologisms in Polish and Ukrainian. It contains more than 1000 phraseological units which are actively used in both languages in the contemporary period. The explication of phraseological semantics is achieved via established universal definition schemes, which allows the inclusion of an unlimited number of other languages, including non-Slavic languages. A strong point of the work is the selection of phraseological neologisms listed along with general phraseological units and phraseological culturemes which are popular in contemporary language, but which have no equivalents in other languages, as they reflect the culture of the given people and its characteristic worldview (*see* Sosnowski & Tymoshuk, 2017; Tymoshuk & Sosnovs'kyĭ, 2017).

'be very familiar with something', *дах поїхав* 'angry, mad for some reason', *бути на одній хвилі* 'have similar views on something', *ловити кайф* 'be happy with something'.

Another typical example of the colloquialisation tendency is the creation of new structural variations of familiar phraseological word combinations through the exchange of stylistically neutral components with colloquial components which are expressively marked, cf. the Polish *umierają jak muchy* and *padają/zdychają jak muchy*, *mieć nierówno pod sufitem* and *mieć nierówno pod kopułą*[2]; *śmierdzi jak cholera, jak sarna* (Bańko, 2004) and *śmierdzi jak z murzyńskiej chaty*; *gładki jak pupa niemowlęcia* (Bańko, 2004) and *gładki jak pupcia niemowlęcia*. Colloquialisation finds expression in the addition of different types of extensions to familiar phraseologisms, which increases their expressive power, cf. in Polish *głupi jak but* (*z lewej nogi*), *brzydki jak noc* (*listopadowa*), *ciemno, jak w dupie* (*u Murzyna*) + (*po czarnej kawie*) + (*w piwnicy*), *od groma* (*i ciut, ciut/i trochę*), *od Sasa do lasa* (*i z powrotem*).

Closely linked to the colloquialisation tendency are examples of vulgarisation which are common in contemporary phraseology. Vulgarisation can affect both the whole phraseological unit and its separate components, cf. *potrzebne jak kurwie majtki*, *wchodzić bez mydła* (*włazić/wchodzić do dupy bez mydła*) in Polish; *руки з дупи/не звідти ростуть*, *бігати/літати як сраний віник/як сраний кіт*, *влізти/залізти* (*в дупу*) *без мила* in Ukrainian.

The opposite tendency of euphemisation can also be observed in the creation of new phraseologisms, although to a lesser degree, cf. in Bulgarian *ляво братство* 'homosexuals', *минавам на левия тротоар* (*на левия бряг*) 'become homosexual'; in Polish *kochający inaczej* 'homosexual', *sprawny inaczej* 'a disabled person, invalid'. In the Polish examples the euphemisation is accompanied by the attribution of jocular style markers.

The mechanism governing the creation of a large number of neophraseologisms in the three languages essentially works by reshaping a prototypical free word combination through a metaphorical or (on rare occasions) metonymic transfer, cf. *сменям чипа* 'change your way of thinking, your mindset', *дръпвам шалтера* 'put an end to an activity', *връзвам кранчето* 'stop financing an activity', *клатя стола на някого*, *дебели вратове* in Bulgarian; *cisnąć do dechy, kręcić lody, urwał się film komuś, zasuwać jak mały samochodzik/jak mały parowozik* in Polish; *іти в тінь* 'be engaged in illegal economic activity', *бути в одному човні* 'be in the same situation and be faced with the same problems' in Ukrainian. A parallel tendency is phraseologisation via the determinologisation of new or old terminological combinations, cf. *бета версия* 'a copy of an original', *висш пилотаж* 'absolute professional mastery in a given field', *летящ старт* 'begin an activity, which provides the opportunity for fast progress because of some initial advantages', *шокова терапия* 'reforms geared to the quick overcoming of a crisis, but creating difficulties for those affected by them' in Bulgarian; *masa krytyczna* 'a condition the breaking of whose boundaries results in a dramatic change', *pas transmisyjny* 'somebody or something, that sets certain values and directions of development', etc. in Polish; *важка артилерія* 'a reliable means to be used as a last resort', *збився приціл* 'lack the necessary skill and accuracy' in Ukrainian.

Some neophraseologisms are created following the models of phraseological word combinations which already exist. Examples of these in Bulgarian are the new phraseologisms *удрям бингото*, *уцелвам джакпота* (following the familiar phraseological expressions *удрям* (*уцелвам*) *шестицата*), in Polish — *czarny weekend* (related to the large number of accidents during the so-called long weekends) following the model of *czarny dzień*.

Various types of jokes and puns are an important source of neophraseologisation material. *We need to add jokes as expressive means to the well-known and extensively described reasons for the creation of neologisms. A speaker shares a joke with their interlocutors, who find it so amusing that they start spreading it. Undoubtedly, such a phenomenon is related to the need to keep a distance from the surrounding world.* Such are the observations of J. Satoła-Staśkowiak, who is a researcher in contemporary Polish and Bulgarian lexis (Satoła-Staśkowiak, 2016, pp. 189–

---

[2]The variant *pod kopułą* has been listed only in the Bulgarian-Polish dictionary of colloquial lexis (Sotirov, Mostowska, & Mokrzycka 2011–2013).

190). Her thesis has been supported by the present observations of contemporary phraseology. For the purposes of this research a survey was conducted. Several hundred respondents, who are native speakers of Polish, were requested to read two Polish set comparison models: *wystroić się jak...* and *znać się na czymś jak...*, and to identify variants of the comparatum. Along with well-established variants of the listed set comparison models included in the Dictionary of Comparisons (Bańko, 2004) such as *wystroił się jak lalka, jak na bal, jak do ślubu, jak na wesele, jak stróż w Boże Ciało*,[3] the respondents came up with a large number of other variants: *jak szczur na otwarcie kanału, jak choinka, jak biedronka na święto lasu*. For the set comparison model *zna się ktoś na czymś jak...*, along with the dictionary variant *jak kura na pieprzu*[4] the respondents included the following variants: *jak wilk/pies na gwiazdach, jak (krowa) na balecie, jak kura na jaju, jak Żyd na świni*. All the examples in the survey are clearly stylistically marked as jocular. It must be noted, however, that they have a high frequency of use in contemporary communication.

Examples of neophraseologisation achieved via wordplay mechanisms are several phraseologisms based on homophony (cf. in Polish *a nuż, widelec* playing upon the homophones *nuż* and *nóż*), or on rhyming (cf. in Polish *celny jak dziad kościelny, mądra jak flądra, cyce jak donice* (Bańko, 2004, p. 213)), as well as *łeb jak sklep, ni pies, ni wydra coś na kształt świdra*.

A unique aspect of neophraseologisation is the structural variation within existing phraseological units. Structural transformations can affect either the lexical composition of the phraseme or its grammatical structure. Examples of a variety of lexical exchange or extension have been discussed in the preceding paragraphs. A reduction of lexical components is also possible, cf. in Polish *woda sodowa uderza komuś do głowy* and *sodówka uderza/uderzyła*. In the context of maximum structural and semantic condensation, the reduction of components produces the lexicalisation of the phraseologism, cf. in Polish *woda sodowa uderza komuś do głowy* and *sodówka, spadać na drzewo/na bambus* and *spadówa/spadówka*.

3 Opportunities for electronic corpora application in the research on new Bulgarian, Polish and Ukrainian phraseology

An indispensable part of contemporary linguistic research is the use of electronic linguistic corpora, as well as of linguistic infrastructures which can process extremely large collections of information in the natural languages in real time. A linguistic corpus is generally held to be a large, standardised, structured body of natural language texts which has been linguistically annotated and presented in a computer readable form. The corpus management system is based on more or less universal software tools for the extraction and processing of a variety of linguistic information (Shyrokov, Buhakov, Hriaznukhina, et al., 2005, pp. 11–17). The advantages which the electronic corpus offers to researchers pertain to the opportunity to work with a vast collection of linguistic material, and to achieve a high degree of breadth and efficiency while processing the information in the context of direct access to a great variety of linguistic facts.

In contemporary Slavic Studies national linguistic corpora play an important role.[5]

The Bulgarian National Corpus (BulNC)[6] was created by the Institute for Bulgarian Language at the Bulgarian Academy of Sciences. *According to contemporary classifications the BulNC is a large, unbalanced and dynamically changing one* (Koeva, 2014, p. 47). Currently, it contains over 240,000 documents of approximately 1.2 billion words. The main contemporary principles of its creation are: *a standardised approach to the collection, classification and processing of texts in different languages; (mainly) automatic identification and collection of suitable online texts regardless*

---

[3]Our research points to this being the most widely-used comparison model (several hundred respondents' replies).
[4]Our research points to this being the most widely-used comparison model (several hundred respondents' replies).
[5]An overview and a description of the existing electronic corpora for Slavic languages have been provided by Hebal-Jezierska, 2014.
[6]http://dcl.bas.bg/bulnc/

of the particular task; a taxonomically organised metadata classification model for text description, which allows the inclusion of new categories and easy reorganisation; an annotation model based on the principle of accumulation of linguistic data (Koeva, 2014, p. 47; Koeva, Stoyanova, Leseva, Dimitrova, Dekova, & Tarpomanova, 2012). The corpus has a semantically-annonated section.

For the Polish language, the largest linguistic corpus is The National Corpus of Polish (NCP).[7] It is the result of a joint initiative between the Institute of Theoretical Foundations of Computer Science, the Institute of Polish Language at PAN, the publishing house Wydawnictwo Naukowe PWN, as well as the Department of Computational and Corpus Linguistics at the University of Lodz (Przepiórkowski, Bańko, Górski, & Lewandowska-Tomaszczyk, 2012). It is a representative linguistic corpus of over 1.5 billion words. There is a balanced part of 300 million words as well as a manually annotated section. The list of sources for the corpus contains not only classical literature, but also daily newspapers, specialist periodicals and journals, transcripts of conversations, and a variety of Internet texts.

At the Ukrainian Lingua-Information Fund of the NAS of Ukraine, as part of the development of the National dictionary database, there exists The Ukrainian National Linguistic Corpus (UNLC).[8] The corpus contains approximately 180 million words from sources which date from the beginning of the 19th century to the present day: both translated and original literature from different periods, popular science and research papers, newspaper and magazine texts, etc. The corpus is being used to compile the new 20-volume dictionary of the Ukrainian language.

The national electronic corpora are a productive environment for the extraction and statistical processing of linguistic data. Additionally, they are a vital resource for the study of the linguistic dynamics of phraseological units. Data from the Polish and the Ukrainian national linguistic corpora has been successfully incorporated into the latest research work into Polish and Ukrainian phraseology — *The dictionary of active Polish and Ukrainian phraseology* [*Leksykon aktywnej frazeologii polskiej i ukraińskiej*] (Tymoshuk et al., 2018).

The size and composition of the corpus are equally important for the observation of phraseological innovations. Many research papers have highlighted the importance of using the largest possible corpus for the successful completion of various linguistic tasks (Kilgarriff & Grefenstette, 2003; Meyer, 2004 et al.). Such an approach is especially vital for the study of linguistic phenomena of low-frequency use such as neologisms (phraseological ones included), which fall into this category by definition. As Svetla Koeva rightly observes, *the larger corpora imply a more reliable illustration of a wider range of linguistic phenomena of high frequency and a broad distribution across a number of thematic fields, styles and genres*. Corpora size is also a precondition for the availability of *a sufficient number of hits, even for rarely used words, collocations and compound lexical units* (Koeva, 2014, p. 39).

For research into phraseological innovations it is essential for the corpus to include an adequate number of texts with an implied wider use of phraseological units. Examples of such texts are literature, sociopolitical writing, spoken colloquial language, dialogues from contemporary films, etc.

Appropriately compiled electronic corpora are a valuable source for the extraction of neophraseologisms and the observation of their functioning in real contexts, cf. Figures 1, 2 and 3.

Another useful characteristic of a corpus is greater chronological depth, which facilitates the comparison of language status over time. For example, the BulNC includes texts dating from 1945 to the present day; the classification of texts according to their year of creation allows for the compilation of subcorpora following a chronological principle (Koeva, 2014). Such flexibility provides the opportunity to establish, with greater or lesser accuracy, the period of occurrence of a certain neophraseologism, as well as to trace the dynamics of its functioning. Corpus data on the phraseologism *сини мравки*, for example, points to its first occurrence in the periodical press and sociopolitical writing in 1993. There was a peak in the frequency of use of the said phraseologism

---

[7] http://nkjp.pl/
[8] http://unlc.icybcluster.org.ua/virt_unlc/

Figure 1: Concordances with the new phraseologism *пране на мръсни пари* in BulNC.

between 2002–2003 and a respective decline in the frequency of use in the period that followed. The loss of currency of the phenomenon it described has resulted in its sporadic use over the past several years.

In the case of phraseologism variation, a corpus-based approach offers the opportunity to establish the extent of the use of the individual variants, and to determine which is the principal variant. For instance, it has been found that among the variants of the Polish phraseologism *gra* (*mecz, pojedynek*) *do jednej bramki* the most frequent in written texts is the variant containing the component *mecz* (30 hits in NCP, *see* Figure 4), followed by the variants containing the component *gra* (18 hits) and *pojedynek* (9 hits).

In Bulgarian, among the variant phraseologisms *разпъвам/разпъна (опъвам/опъна, разтварям/разтворя, спускам/спусна, слагам/сложа, разпервам/разперя, поставям/поставя, държа) политически чадър над някого или нещо* the variant with the highest frequency of use is that with the central verbal component *разпъвам/разпъна*. This detail is the rationale for the lexicographic decision to list the phraseologism *разпъвам/разпъна политически чадър* as the representative variant in the Bulgarian dictionaries of neologisms and to provide the alternative versions via cross-references to it (Blagoeva & Kolkovska, 2017).

Parallel electronic corpora, whether bilingual or multilingual, are also practical tools for the extraction of information about innovations in the phraseological systems of the three languages. The process of searching for the equivalents of phraseological units provides a good illustration of how these types of corpora can be used in translation, dictionary development and language teaching. The working group of the Institute of Slavic Studies at the Polish Academy of Sciences has been developing a parallel Polish-Bulgarian-Russian-Ukrainian corpus, which is to be incorporated into the CLARIN framework.[9]

---

[9]Common Language Resources and Technology Infrastructure is a project granted the status of ERIC (European

Figure 2: Concordances with the new phraseologism *pranie brudnych pieniędzy* in NCP.

A new development in the compilation of multilingual resources is the inclusion of texts whose characteristics come close to spoken colloquial language. The inclusion of such texts in the multilingual parallel corpora at this stage is achieved through the processing and adding of collections of dialogues from contemporary television series and films in the respective languages. This type of text constitutes a substantial part of the material in the developed corpora. The parallel Polish-Bulgarian-Russian-Ukrainian corpus allows users to search for new phraseological units.

It must be taken into consideration, however, that phrasemes can exist:

In two languages

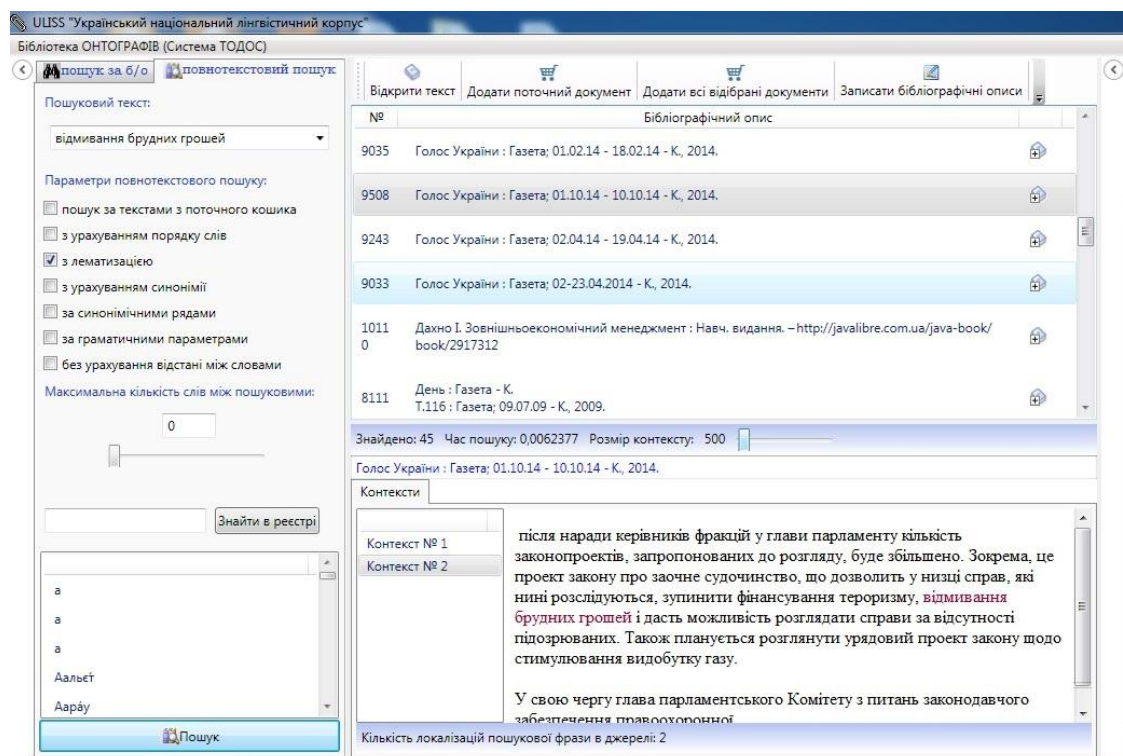|	
— Serio?	— Невже?
— W sumie nie wiem.	— Я маю на увазі, я не знаю.
— Na mnie krzyczeć nie będzie.	— Слухай, вона не може кричати на мене, правильно?
W końcu jedziemy na tym samym wózku.	Не після... тобто ми ж в одному човні.

"Fargo"

Figure 3: Concordances with the new phraseologism *відмивання брудних грошей* in UNLC.

In only one language

Це останній писк.	Najnowszy model.
Трохи дорого.	Trochę drogi.
Якщо взяти кредит…	Tutaj jest plan spłat…
Корпус з металу.	Metalowy.
Таке мене не цікавить.	Nie jestem zainteresowana.

"Frozen Land"

… Znowu z Hooli.	Знову дзвонять з Hooli.
Pewnie ci wyluzowani koderzy.	Мабуть, це друзі-програмісти.
Chcą mi powiedzieć, że strona jest do dupy.	Напевно, не терпиться пожартувати з приводу мого сайту.

"Silicon Valley"

As part of the international project CLARIN, parallel bilingual corpora are being compiled, with Polish being one of the target languages, including the Polish-Bulgarian Parallel Corpus. The overall volume of these resources will exceed 20 million wordforms. The results of the work completed to date are available on the CLARIN-PL project web page which offers access to the KonText software,[10] designed for the search of language resources. Figure 5 illustrates a search

---

[10] https://kontext.clarin-pl.eu/run.cgi/first?shuffle=1&reload=1&corpname=polish_bulgarian_corpus_PL&queryselector=iqueryrow&iquery=&phrase=&word=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&sca_turn.speakers=&sca_seg.soundfile=

Przeszukiwany zbiór zawiera 240,192,461 słów. Znaleziono **30** akapitów pasujących do zapytania w **0.079s.** Bieżąca strona zawiera 30 przykładów z 30 różnych tekstów.

#			
1.	Petro Jerzego Kasalika, który stwierdził, że był to mecz	**do jednej bramki**	. Kasalik, absolwent wrocławskiej AWF, powinien mieć na
2.	na boisku rywala, jedenastkę Orlika Gogolinek 2:1. Mecz	**do jednej bramki**	, w ogromnym upale mógł skończyć się wyższym zwycięstwe
3.	- To był mecz praktycznie	**do jednej bramki**	. Wprawdzie przy stanie 4:1, trener naszego zespołu Leszek
4.	zespołem, uzyskaliśmy dużą przewagę i mecz toczył się	**do jednej bramki**	. Była szansa na remis, ale po akcji Daniela Koryciaka,
5.	prowadzić 2:1 po trzech setach. Tie break toczył się już	**do jednej bramki**	. Z meczu na mecz Gosia grała coraz lepiej
6.	— To był w zasadzie mecz	**do jednej bramki**	. Goście oddali tylko dwa strzały na naszą bramkę. Szybko
7.	To był mecz praktycznie	**do jednej bramki**	. Widać było, że spotkały się zespoły z przeciwnych
8.	opadają z sił. Trzecia tercja to był już typowy mecz	**do jednej bramki**	. Goście często nie byli nawet w stanie wybić krążka z
9.	podwójną karę Morzyka. Trzykrotnie zmniejszyli stratę	**do jednej bramki**	, po raz ostatni w 56 minucie. Nie wykorzystał jednak
10.		Mecz	**do jednej bramki**	
11.	Jojczyk, Tomasiak, Sroka – Sieniawa, K. Pawlik. Mecz	**do jednej bramki**	. W 5 min po rzucie rożnym Mrzygłód z 5 metrów trafił do
12.	MECZ	**DO JEDNEJ BRAMKI**	
13.	(14:20 w 41 min.), Witaminki znowu zmniejszyły stratę	**do jednej bramki**	. Na 13 minut przed końcem po kolejnej dobrej akcji Marty
14.	W Nowym Dworze Gdańskim znów odbył się mecz "	**do jednej bramki**	", bez historii. Trochę więcej emocji było w dwóch
15.	Mecz	**do jednej bramki**	. Trampkarze Gorc oddali bowiem jedynie trzy strzały na
16.	- Mecz	**do jednej bramki**	i znów nie wykorzystujemy okazji. Straciliśmy gola po
17.	by przytrzymywać piłkę, nie szarpać się. To był mecz	**do jednej bramki**	, według zasady: kto strzeli pierwszy, ten wygrywa. Cały
18.	Klasyczny mecz	**do jednej bramki**	, zakończony wysokim i gładkim zwycięstwem Korei.
19.	obraz gry nie uległ zmianie. Mecz w zasadzie toczył się	**do jednej bramki**	. W 63 min po zagraniu Kaczówki Maik znajdując się w
20.	zastanawiające, dlaczego rozgrywa się tutaj mecz	**do jednej bramki**	. Poza głosem pana senatora Kruka, a wcześniej
21.	Mecz nabierał kolorów, napięcie rosło, bo różnica zmalała	**do jednej bramki**	. Po sfaulowaniu na polu karnym jednego z księży, sędzia
22.	Finał pomiędzy Gryzem a Bochenkiem to praktycznie gra	**do jednej bramki**	. Doskonale przygotowany Gryz co rusz trafia potrójne
23.	ale, niestety, były środki . Pierwszy półfinał to mecz	**do jednej bramki**	. Kapitalny popis swoich możliwości dał Marcin Miller,
24.	sezonie. Sam mecz z Niegowonicami był typowym spotkaniem	**do jednej bramki**	. Na odnotowanie zasługuje fakt, że na nietypowej dla
25.	w odpowiedniej chwili. Pierwszy półfinał to mecz	**do jednej bramki**	. Kapitalny popis swoich możliwości dał Marcin Miller,
26.	ale, niestety, były środki . Pierwszy półfinał to mecz	**do jednej bramki**	. Kapitalny popis swoich możliwości dał Marcin Miller,
27.	w odpowiedniej chwili. Pierwszy półfinał to mecz	**do jednej bramki**	. Kapitalny popis swoich możliwości dał Marcin Miller,
28.	wielkich emocji. – Wydawało się, że będzie to spotkanie	**do jednej bramki**	, spokojne. W pierwszej połowie zdobyliśmy dwa gole
29.	czuli się na nim doskonale. Był to wwięc w zasadzie mecz	**do jednej bramki**	. Tym bardziej, że niemal całe spotkanie graliśmy w
30.	2:3 Pawłowi Władyce, ale później mecz toczył się już	**do jednej bramki**	. Kolping pewnie usadowił się na fotelu lidera. Za bardzo

Figure 4: Concordances with the new phraseologism *mecz do jednej bramki* in NCP.

for the Bulgarian correspondences of the Polish phraseologism *dać za wygraną* into the Polish-Bulgarian Parallel Corpus.

4   Conclusions

The comparative study of phraseological innovations in Bulgarian, Polish and Ukrainian shows that over recent decades there have been thoroughgoing neophraseological processes which have led to a substantial enrichment of the phraseological stock of each of the three languages. The factors responsible for these changes are primarily related to the increased communication needs for new nominative tools through which newly emerged phenomena, or phenomena which have regained prominence, can be identified and categorised in terms of their expressive power.

In terms of the sources and mechanisms regulating the formation of new phraseological units in Bulgarian, Polish and Ukrainian, a considerable inter-lingual similarity has been established. Nevertheless, regardless of the availability of a certain number of formally and semantically corresponding new phraseologisms in these languages, the neophraseologisation results are language specific.

The neophraseologisation processes in Bulgarian, Polish and Ukrainian reveal a common set of tendencies, the most prominent of which is the distinct tendency of linguistic democratisation shared by all contemporary Slavic languages. A number of phenomena affecting the formation of new phraseological units in the three languages are related to this tendency, namely colloqualisation, slangisation, and vulgarisation. They intensify the expressive power of the nominative tools. The opposite tendency of intellectualisation is observed in a limited number of cases, such as the phraseologisation of individual terminological word combinations.

In the field of new Bulgarian, Polish and Ukrainian phraseology there is a distinct tendency towards internationalisation, which stimulates the further expansion of the international phraseological stock, which is the result of loan translation — and on rare occasions of lexical borrowing — of phraseological units from English. A significant number of the new phraseologisms shared by the three languages are calques from English. Nonetheless, the increased importance of phraseological loan translation testifies to the opposite tendency of nationalisation, as the reproduction

Figure 5: Concordances with the phraseologism *dać za wygraną* in the Polish-Bulgarian Parallel Corpus (D. Roszko, R. Roszko, & Sosnowski, in press; R. Roszko, D. Roszko, Sosnowski, & Satoła-Staśkowiak, 2018).

of the foreign phraseological prototypes via domestic linguistic material is preferred to passive borrowing.

The processes of neophraseologisation are also characterised by other general linguistic tendencies and phenomena, such as the linguistic economy principle whose effect is obvious in the reduction of lexical components in the structure of some of existing phraseologisms.

New phraseology is a distinctly dynamic segment in the lexical systems of languages. It is very often the case that after a period of extensive use, neophraseologisms lose currency and are replaced by new ones. A sure sign of the fluidity of this segment is the particularly prominent structural variation. Which of the phraseological neologisms are here to stay is for the language to decide, following its natural path of development.

It is crucially important for research into the developmental processes in phraseology to be based on corpora data. The large and constantly updated electronic corpora constitute an objective basis for the tracing of *the lifecycle* of phraseologisms, their context and frequency of use, as well as for observations of their variants.

References

Alefirenko, N. F., & Semenenko, N. N. (2009). *Frazeologiia i paremiologiia*. Moskva: Flinta, Nauka.

Bańko, M. (2004). *Słownik porównań*. Warszawa: PWN.

Blagoeva, D., & Kolkovska, S. (2017). Sŭvremenni podkhodi v bŭlgarskata neografiia. *Spisanie na BAN*, *4*, 16–20.

Bondzholova, V. (2015). *Osobenosti na reklamniia ezik i stil: Leksikalen aspekt*. Veliko Tŭrnovo: Faber.

Georgieva, S., & Velichkova, S. (2008). Frazeologichnite neologizmi v bŭlgarskiia ezik i tekhnite sŭotvetstviia v drugi slavianski ezitsi. In *Slavianska filologiia* (Vol. 24, pp. 286–299). Sofiia: AI "Prof. M. Drinov", 286–299.

Hebal-Jezierska, M. (Ed.). (2014). *Praktyczny przewodnik po korpusach języków słowiańskich*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the Special Issue on Web as Corpus. *Computational Linguistics*, *29*(3), 333–347. https://doi.org/10.1162/089120103322711569

Kita, M. (1991). Ekspansja potoczności. In A. Kowalska, & A. Wilkoń (Eds.), *Studia polonistyczne* (pp. 81–99). Katowice: Wydawnictwo Uniwersytetu Śląskiego. (Prace Naukowe Uniwersytetu Śląskiego w Katowicach. Prace Językoznawcze, 19).

Koeva, S. (2014). Bŭlgarskiiat natsionalen korpus v konteksta na svetovnata teoriia i praktika. In *Ezikovi resursi i tekhnologii za bŭlgarski ezik* (pp. 29–52). Sofiia: AI "Prof. M. Drinov".

Koeva, S., Stoyanova, I., Leseva, S., Dimitrova, T., Dekova, R. & Tarpomanova, E. (2012). The Bulgarian National Corpus: Theory and practice in corpus design. *Journal of Language Modeling*, *2012*(1), 65–110. https://doi.org/10.15398/jlm.v0i1.33

Lusińska, A. (2010). *Reklama a frazeologia: Teksty reklamowe jako źródło nowych frazeologizmów*. Toruń: Wydawnictwo Adam Marszałek.

Meyer, C. F. (2004). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Mokienko, V. M. (2003). *Novaia russkaia frazeologiia*. Opole: Uniwersytet Opolski.

Nedkova, E. (2017). Osnovni kontseptosferi na frazeologichnite inovatsii v bŭlgarskiia ezik. In *Arnaudov sbornik* (Vol. 9, pp. 429–432). Ruse: Leni-An.

Pajdzińska, A. (2013). Frazeologia — szkic do naszej zbiorowej autobiografii. In J. Mazur, A. Dunin-Dudkowska, & A. Małyska (Eds.), *70 lat współczesnej polszczyzny* (pp. 145–164). Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Przepiórkowski, A., Bańko, M., Górski, R. L., & Lewandowska-Tomaszczyk, B. (Eds.). (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.

Roszko, D., Roszko, R, & Sosnowski, W. (in press). Polish-Bulgarian Corpora ISS PAS (IS PAN) and CLARIN-PL. *Slavica Lodziensia*, *2018*(2).

Roszko, R., Roszko, D., Sosnowski, W., & Satoła-Staśkowiak, J. (2018). Polish-Bulgarian Parallel Corpus, CLARIN-PL digital repository, http://hdl.handle.net/11321/536

Roszko, R., Tymoshuk, R., Duszkin, M., & Sosnowski, W. (2018). Polish-Ukrainian Parallel Corpus, CLARIN-PL digital repository, http://hdl.handle.net/11321/535

Satoła-Staśkowiak, J. (2016). Żart w procesie neologizacji — na podstawie języka młodego pokolenia. *Językoznawstwo*, *1*(10), 185–192. https://doi.org/10.25312/2391-5137.10/2016_185-192

Shyrokov, V. A., Buhakov, O. V., Hriaznukhina, T. O., et al. (2005). *Korpusna linhvistyka*. Kyïv: Dovira.

Sosnowski, W., & Bonnard, P. (2015). The current evolution of Slavic languages in Central and Eastern Europe in the context of the EU multilingualism policy. *Cognitive Studies/Études cognitives*, *15*, 397–411. https://doi.org/10.11649/cs.2015.028

Sosnowski, W. & Tymoshuk, R. (2017). On the Dictionary of Active Polish and Ukrainian phraseology [Leksykon aktywnej polskiej i ukraińskiej frazeologii]: Contrastive linguistics and culture. *Cognitive Studies/Études cognitives*, *2017*(17). https://doi.org/10.11649/cs.1317

Sotirov, P., Mostowska, M., & Mokrzycka, A. (2011–2013). *Bułgarsko-polski słownik leksyki potocznej* (Vols. 1–2). Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Styshov, O. A. (2005). *Ukraïns'ka leksyka kintsia XX stolittia (na materiali movy zasobiv masovoï informatsiï)*. Kyïv: Vydavnychyĭ tsentr KNLU.

Styshov, O. A. (2015). Osnovni dzherela popovnennia frazeolohichnoho skladu ukraïns'koï movy kintsia XX – pochatku XXI stolit'. *Movoznavstvo*, *2015*(1), 33–46.

Tymoshuk, R. P., Sosnovs'kyĭ, V. (2017). Novi pidkhody do stvorennia suchasnykh frazeolohichnykh slovnykiv (na materiali "Leksykona pol's'koï ta ukraïns'koï aktyvnoï frazeolohiï"). *Movoznavstvo*, *2017*(2), 69–77.

Tymoshuk, R., Sosnowski, W., Jaskot, M., & Ganoshenko, Y. (2018). *Leksykon aktywnej frazeologii polskiej i ukraińskiej*. Warszawa: KJV Digital.

Videnov, M. (1997). *Ezikŭt i obshtestvenoto mnenie*. Sofiia: AI "Prof. M. Drinov".