

**Olga GORITSKAYA**  
(Minsk State Linguistic University, Belarus)  
goritskaya@gmail.com

## **Belarusian flavour in Russian: how to measure gradual differences between the varieties of pluricentric languages?**

### **Abstract**

The article focuses on lexical features of Belarusian Russian. It shows that the distinctness of Belarusian Russian results from the presence of elements unique for the variety (qualitative differences between the varieties), and the functional specificity of lexemes shared by different national varieties of Russian (quantitative differences). The paper highlights that differences between the national varieties of the Russian language (in particular, between the non-dominant Belarusian and the dominant Russian varieties) are of a gradual nature, which is proved by statistical methods applied to the corpus material, as they demonstrate a significant increase or decrease in the observed frequency of lexical items compared to their expected frequency.

### **1. Introduction**

The specific structural (phonetic, lexical, grammatical, etc.) features of the varieties of pluricentric languages are sometimes quite subtle. This is explained not only by the fact that there are some structural limits of variation within the varieties of *one* language. Equally important are sociolinguistic factors that regulate the functioning of language varieties – in particular, the influence of a codified standard that is often found only in the dominant varieties of pluricentric languages. See, for example, an observation of Ronald Wardhaugh (2010:34), relevant not only for World Englishes:

“Today, written Standard English is codified to the extent that the grammar and vocabulary of written varieties of English are much the same everywhere in the world: variation among local standards is really quite minor, being differences of ‘flavor’ rather than of ‘substance,’ so that the Singapore, South African, and Irish varieties are really very little different from one another so far as grammar and vocabulary are concerned”.

As Michael Clyne (1992:459) notes, the dominant nations “have difficulty in understanding the ‘flavor rather than substance’ notion”. However, it is clear that the non-dominant nations also come across this problem: speakers are often not

aware of the specific features of their own speech or ignore them as not prestigious enough (people say “there are no such elements in the language”, meaning that these elements are not characteristic of the codified variety). These attitudes hinder the development of the non-dominant varieties. That is why the detection and explanation of structural specifics of the varieties (primarily, the non-dominant ones) is one of the most important tasks for pluricentric studies and, in particular, for corpus planning.

This article deals with specific lexical features of Belarusian Russian. This is a highly-variable language variety at the stage of its formation and development (Del Gaudio 2013, Woolhiser 2012, etc.). Belarusian Russian exists within a Belarusian-Russian language continuum where different social, regional, genre, and other varieties of Russian and Belarussian, as well as Belarusian-Russian mixed speech (so-called *Trasjanka*) are found (Hentschel 2017; Goritskaya 2019). In this language continuum, linguistic items interact and migrate, which complicates drawing the boundaries between the varieties, see, for example, the following observation of a Belarusian:

*A jaščè belarusy ŭnikal’nyja tym, što majuc’ upèŭnenasc’ – jany razmaŭljajuc’ na čys’cjutkim ruskim jazyku. Bez anijakaha akcèntu. Ne toe što maskvičy, z ix žudasnym maskoŭskim akcèntam. Bo jany, belarusy – ruskija sa znakam jakasci. Pry hètym ruskija verac’, što ruski jazyk belarusaŭ – hèta takiaja vos’ dziŭnavataja, ale zrazumelaja im, ruskim, belaruskaja mova belarusaŭ<sup>1</sup>.*

[One more unique feature of Belarusians is that they are completely sure they speak the purest Russian. With no accent. Unlike Moscow citizens, with their horrible Moscow accent. As they, Belarusians, are Russians with a quality mark. At the same time, Russians believe that Belarusian Russian is quite strange, but understandable for them the Belarusian language of Belarusians.]

## 2. Types of Belarusianisms

Belarusian flavor in Russian manifests itself in different ways. As a rule, linguists take into account words and other linguistic items of Belarusian origin, but the distinctness of the variety goes further than that: a lot of specific features originate from Russian itself.

(1) Contact-induced lexical Belarusianisms: They are not restricted to loanwords from Belarusian for notions important for the Belarusian culture, as *mova* ‘language, as a rule, Belarusian’ or *šufljedka* ‘drawer’ (Russian Russian *vydvižnoj jaščik*) (a symbol of Belarusian Russian). They also include items that do not fill any language lacuna and have one-word equivalents in Russian, for instance, *po-*

<sup>1</sup> URL: [https://www.facebook.com/permalink.php?story\\_fbid=10207028336587419&id=1850149328](https://www.facebook.com/permalink.php?story_fbid=10207028336587419&id=1850149328)

*tixu* (characteristic of Belarusian Russian) – *potixon'ku*, *tixon'ko* (used in different countries) ‘on the sly or slowly’. Within the continuum, contact-induced Belarusianisms take a place between code-switching and loanwords (Goritskaya 2019). Potentially, any lexical items of the Belarusian language can be used in Belarusian Russian, but in fact Belarusianisms differ in their frequency: usage of some items is limited, while others are quite widespread (Goritskaya, Suprunchuk 2018).

(2) Intra-systemic Belarusianisms: This type appeared in Belarusian Russian without a direct influence of Belarusian and realizes the potential inherent in the Russian language system, see also (Gorickaja 2018). This – a much subtler – class includes lexemes derived from Russian morphemes, new meanings of Russian words, multi-word expressions, including the periphrastic ones.

Locally specific lexical items, emergence of which is not due to language contacts, denote both Belarusian realia (*palatka* ‘the Lower House of the Belarusian Parliament’), and the phenomena characteristic of other countries as well (for example, *morožko* – diminutive of *moroženoe* ‘ice-cream’). The emergence of such universal items is caused by the constant need for renewal of the vocabulary that may take place independently in different countries. Some items refer to the texts well-known to Belarusians. For example, *žestočajšij* (the superlative from *žestokij* ‘cruel’), *čarka* ‘glass (of alcoholic drink)’ and *škvarka* ‘cracklings’ are associated with the speeches of Alexander Lukashenko, the Belarusian president.

However, because of a close relatedness of Belarussian and Russian, it is not always easy to draw a clear boundary between intra-systemic and contact-induced phenomena. For example, words seen as Belarusianisms by the inhabitants of Belarus (e.g., *burak* ‘beetroot’ or *cibulja* ‘onion’) are fixed in different dictionaries of Russian, sometimes marked as *regional*, *colloquial*, *dated*, *spoken*, etc.

### 3. Aim of this study

So, the following question arises: which words can be regarded characteristic of Belarusian Russian? The aim of this article is to present and provide the methodology that allows to define lexical items characteristic of the variety under investigation. The paper stresses that the specificity of varieties of pluricentric languages is not solely about some unique items. The functional peculiarities of lexical items with wide localization also need to be taken into account. In general, in modern pluricentric languages, there are few words found in the speech of inhabitants of only one city or one country. More often, such lexical items are more or less typical for one or several countries. Dirk Geeraerts (2010:824) noticed:

“The examples we have quoted so far may suggest that the differences in

lexical choices between languages varieties, like American English and British English, are a matter of black and white choices: *subway* is the American term and *underground* is the British term. In actual practice, however, the choices are seldom of a categorical nature. Rather, the selection from among the various lexical alternatives that occur within a language is a gradual matter”.

Words and other language items with different frequency in several countries where one language is used lead the researcher beyond the dichotomy of dominant and non-dominant varieties of pluricentric languages and highlight the natural graduality of structural differences between the varieties. Such language items embody ‘flavor’ rather than ‘substance’ notion and act as “both unifiers and dividers of peoples” (Clyne 1992a:1), and therefore need to be analyzed in more detail.

#### 4. Corpus study of lexical variation in Belarusian Russian

##### 4.1. The choice of corpus

Texts in standard Russian and Belarusian can be separated quite clearly, however in real communication languages interact and mix, which makes it difficult to draw the boundaries within the Belarusian-Russian language continuum. The solution to this problem is an important methodological question in the studies of the communicative space of Belarus. The need to address this issue predetermined the choice of data for the empirical study of the variety. Belarusian-Russian mixed speech (“Trasjanka”) is a linguistic phenomenon that exists primarily in an oral form, so I analyzed only written texts to exclude the examples of mixed speech from the sample and focus on the Belarusian variety of Russian.

The data for this study was extracted from General Internet-Corpus of Russian (GICR, [webcorpora.ru](http://webcorpora.ru)) (Belikov et al. 2013a, 2013b). For this article, I used the “Livejournal” blog sub-corpus that contains 8.72 billion words. The sub-corpus covers different countries where speakers of Russian live.

In GICR, the information on the speaker’s location is rather diverse: in some cases, only the locality (e.g. city) is mentioned, others include the name of a country, but in many cases, the location is not marked. The total number of localizations in the sample is over 2500 (according to “Corpus statistics” section on GICR web site), therefore, a further processing of the data is required. With the help of a specially created software that allows to group the data by countries, I have created profiles (for more about profiles in linguistic studies see, e.g. (Janda 2016)) reflecting the distribution of words in different countries.

Further, while creating the profiles that demonstrate the distribution of lexical items in different national segments of GICR, I have not taken into account the contexts with unmarked localization. There are three main samples in my data: Belarus, Russia, and Ukraine. There is not enough data in the corpus to draw any sound conclusions on the distribution of variants in other post-Soviet countries, therefore contexts from Latvia, Lithuania, Kazakhstan, etc. were grouped together with contexts from countries where migrants from the ex-USSR live (USA, Israel, Germany, etc.). It is a control sample, where the differences between the varieties of Russian are blurred.

#### 4.2. National profiles of variants

With the help of corpora, it is possible to find out where lexical items are used and create national profiles of variants that reflect the specificity of their distribution in different national segments of GICR.

First, the data demonstrates frequency correlation of the competing variants in Belarus, Russia, Ukraine, and other countries. For example, consider the diagram 1 demonstrating different preferences for the name of All Souls' Day celebrated in a week after Easter. There are several names for it: *Radunica*, *Radonica*, *grobki*, *Provody* (*Provoda*), etc. The figure 1 shows that speakers from Belarus, Russia and Ukraine prefer different nominations.

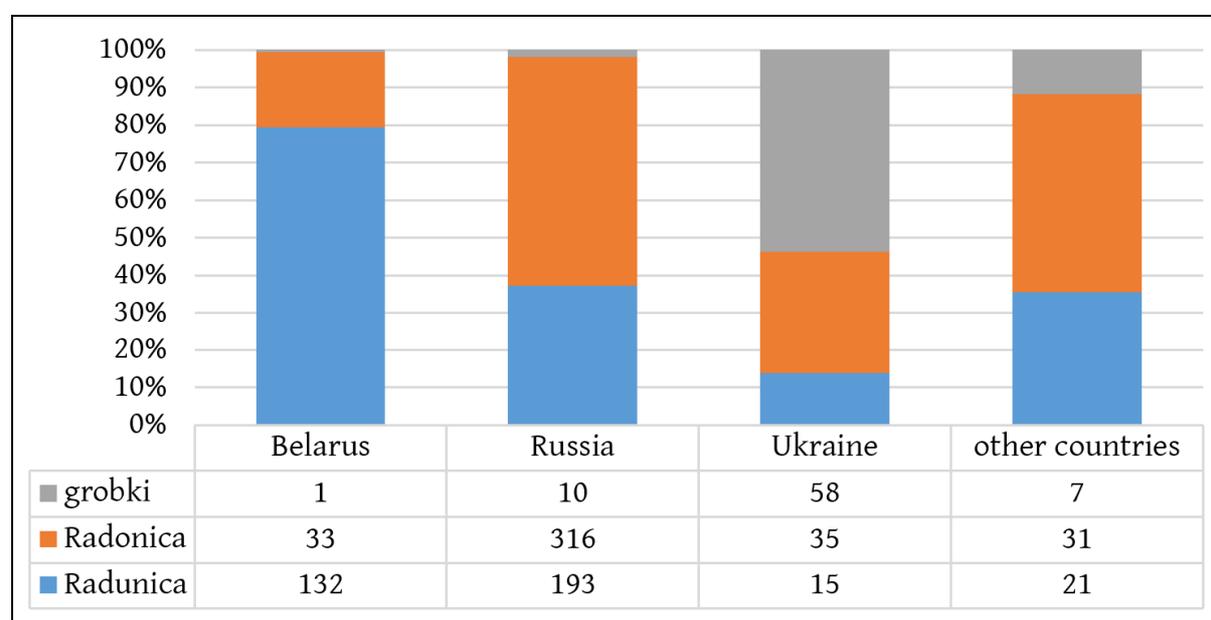


Fig. 1: Ratio of variants *Radunica*, *Radonica* and *grobki* in national segments of GICR

Second, the national profiles help notice the shift of data to one or several countries. In the figure 2, the profile of the “universal” variants *svekla* / *svĕkla* for ‘beetroot’ reflects the distribution of national segments in GICR. The Belarusian

usage of *burak* is more prominent than of *svekla / svėkla*. At the same time, Ukraine contributes significantly more to *burjak*.

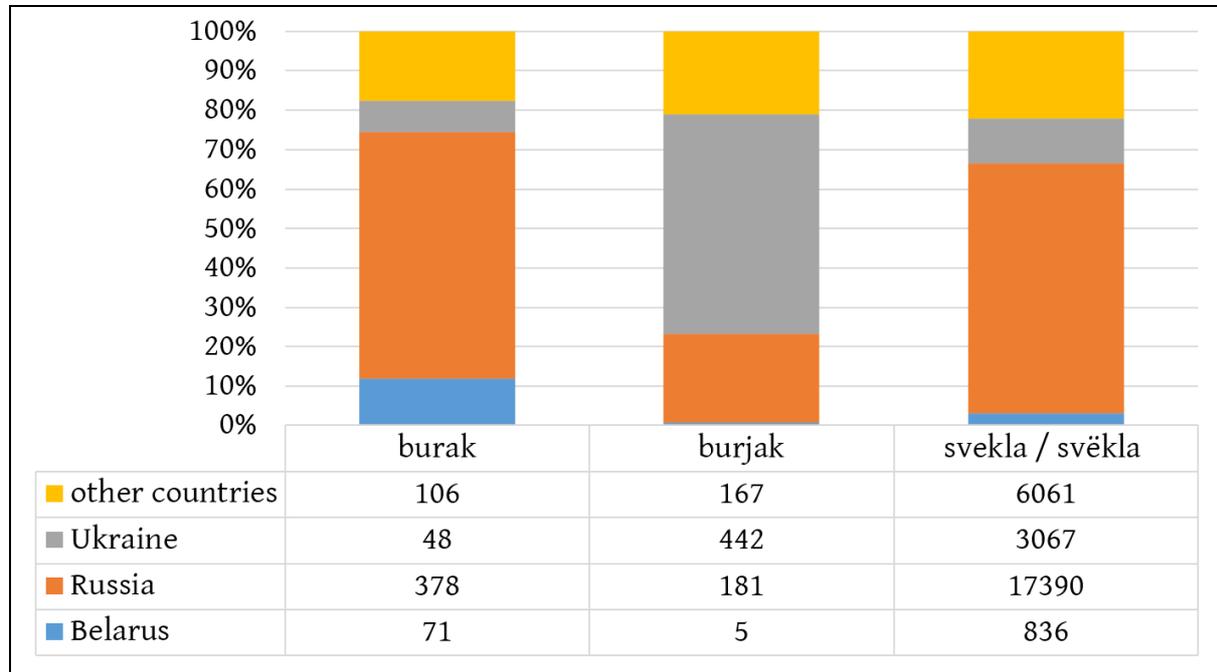


Fig. 2: The proportion of a national segment in the data (based on the distribution of *burak*, *burjak* and *svekla / svėkla* variants)

When can we say that a language item is characteristic of a particular national variety? One can analyze only those cases when the majority of contexts are found in a particular country. It is obvious, for example, that *burjak* is characteristic of Ukrainian Russian. Alternatively, we can also take into account words like *burak*, cf. the Belarusian segments in the profiles of *burak* and *svekla* (*burak* is a Belarusian equivalent for *beetroot*, but is also found in some regional varieties of Russian in Russia; *svekla* is a “Standard” Russian word). Anyway, there is a question: how can we justify our decision for it not to be seen as analytical arbitrariness?

#### 4.3. Statistical analysis

To determine whether the differences in frequencies of lexical items in Belarusian and Russian segments of GICR are significant, the log-likelihood function was used (Dunning 1993; Rayson 2002; Rayson et al. 2004). This measure of statistical significance allows us to determine the difference between the observed and expected frequency of a language item. In my paper, I used the log-likelihood cal-

culator in Excel made by Paul Rayson (the tables and formulas below as well as the calculator itself are from his website<sup>2</sup>).

	<i>Corpus 1</i>	<i>Corpus 2</i>	<i>Total</i>
<i>Frequency of word</i>	<i>a</i>	<i>b</i>	<i>a+b</i>
<i>Frequency of other words</i>	<i>c-a</i>	<i>d-b</i>	<i>c+d-a-b</i>
<i>Total</i>	<i>c</i>	<i>d</i>	<i>c+d</i>

Tab.2: Contingency table to calculate log-likelihood

To calculate the expected frequency, the following formula is used:	Log-likelihood (G2) is calculated according to the formula as follows:
$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$	$-2 \ln \lambda = 2 \sum_i O_i \ln \left( \frac{O_i}{E_i} \right)$

To guarantee the accuracy of the results, the critical value  $G^2$  for log-likelihood should be not less than 15.13 (Rayson 2002: 155),  $p < 0.0001$ . If  $G^2$  is more than 15.13, the differences in frequency between the corpora are considered statistically significant. The log-likelihood function allows us to determine the cases when language items are used more and less frequently in one subcorpus compared to the other. In my study, I have focused on the cases when the observed frequency of a lexical unit is more than expected. In the following example, a Russian speaker from Belarus shares his observations on the differences in the Belarusian and Russian varieties of Russian:

*... provereno na neskol'kix žiteljax Moskvy (voznost ot 21 do 81 goda): slova **ssobojka** i **restik** neponjatny ljudjam... T.e. ssobojku ešče koe-kak ponjali po kontekstu, a vot s restikom naprjaženka. Vse dumali, čto èto rostiks s opečatkoj. Gygy. (Dlja nemoskvičej - Rostiks - set' mestnyx kurinyx fastfudov tipa KFC ili Mak-donal'dsa). Ešče mne govorili, čto slovo **šufljedka** takže neponjatno tut... Gm... (GICR, Livejournal: lelik\_afrika)*

[... it was proved by several Moscow citizens (from 21 to 81 years): *ssobojka* 'food carried to work, school, etc.' and *restik* 'restaurant' (diminutive) are confusing to people ... I mean, people at least understood *ssobojka* relying on the context, but with *restik* it was much worse. Everyone thought that this is *rostiks* with a typo. Lol. (For those outside Moscow, *Rostiks* is a local chain of chicken fast food, like KFC or McDonalds.) Also, I was told that they don't understand the word *šufljedka* 'drawer' here... Hm...]

To define if all of these words are characteristic of Belarusian Russian, I have analyzed their frequency in Russian and Belarusian segments of GICR and

<sup>2</sup> URL: <http://ucrel.lancs.ac.uk/llwizard.html>

calculated their log-likelihood, see the following table (2). The cases when the frequency of a lexical item is lower than expected are also of interest. Often, a lower frequency of a language item demonstrates the competition of variants. If the frequency of one of the competing items increases in one sub-corpus, the frequency of other ones may decrease. For instance, in Belarus, the frequency of the word *sadit* 'to plant' or 'to seat, to put', including 'to put somebody to prison', is higher than expected ( $G^2=349.24$ ), while the frequency of the rivaling *sažat* is, vice versa, lower than expected ( $G^2=378.06$ ).

Lexeme	Observed frequency		Expected frequency		Over (+) / under (-) use <sup>3</sup>	Log-likelihood (G <sup>2</sup> )
	Belarus	Russia	Belarus	Russia		
<i>ssobojka</i>	141	40	6.92	174.08	+	732.54
<i>šufljadka</i>	138	33	6.53	164.47	+	735.85
<i>restik</i>	26	756	29.88	752.12	-	0.55

Tab.2: Frequency of *ssobojka*, *šufljadka* and *restik* in Belarusian and Russian segments of GICR

This requirement is not necessarily followed, especially in those cases when the competing variants differ in their frequency and/or  $G^2$  of the variant with a higher frequency is close to the critical value. For instance, in Belarus, the frequency of *razukraška* 'coloring book' is higher than expected ( $G^2=83.90$ ), but this does not lead to a statistically significant decrease of a much more frequent *rask-raska* ( $G^2=0.25$ , that is lower than the critical value), which has several meanings. However, things are different with the verbs these words are derived from: the frequency of *razukrašivat*' is higher than expected ( $G^2=26.10$ ), while the frequency of *raskrašivat*' is lower ( $G^2=16.37$  that is a little higher than the critical value). All of this demonstrates the diversity of language variation.

Using this method, I have found around 170 words (to date) that, on the one hand, are widely used in Russia and are present in the dictionaries of the Russian language, and on the other hand, show higher frequency in the Belarusian segment of corpus than expected. These lexical variants, along with loanwords from the Belarusian language (not mentioned in the dictionaries of "Standard" Russian), are considered as lexical Belarusianisms. In addition, I have found around 30 words, which are under-used in the Belarusian segment of GICR in comparison with Russian. Increase and decrease in frequency of these words is caused by language contact, political, cultural, and other factors.

<sup>3</sup> In Belarus compared to Russia.

In the studies of specific structural (including lexical) features of non-dominant varieties of pluricentric languages the primary task is to determine the specificity compared to the dominant variety. As only the Russian variety of Russian is codified, it is reasonable to use the Russian sub-corpus as a referent one. At the same time, it is clear that in every country, especially such a large one as Russia, an internal or second-level pluricentricity is also observed (Clyne 1992a: 3; Muhr 2016: 20). However, the comparison of Belarus with different regions of Russia did not seem appropriate, as it would blur the boundaries of the object of the study and would not take into account the specificity of a national variety of the language developing in a sovereign state and put Belarusian Russian in line with other Russian regional varieties (regiolects).

#### 4.4. Functional specificity of lexical variants in the Belarusian and Ukrainian varieties of Russian

The data of the study allows us to determine some functional specifics of language items in non-dominant varieties of Russian – Belarusian and Ukrainian. Due to the structural closeness of Belarusian and Ukrainian and similarities of their sociolinguistic history, Belarusian Russian and Ukrainian Russian share several items that distinguish them from Russian Russian, and the limitation of the sample to the words unique for Belarus only does not take into account the graduality of differences between the varieties of Russian. The following metalinguistic comment illustrates this fact:

*Ja živu v Minske s sentjabrja i rešil pereexat' v Belarus' nadolgo. V celom Minsk poxož na Ufu, otkuda ja priexal. On nebol'šoj i spokojnyj, mne zdes' komfortno. Kogda ja žil v Rossii, v moej sem'e ispol'zovali nekotorye ukrainskie slova – v nas est' ukrainskaja krov'. Kogda èti slova vyletali u menja v reči, na menja stranno kosilis' i prosili perevesti. Zdes', kogda tak proisxodit, menja vse ponimajut – èto tak kruto. Kažetsja, ja priexal kuda nužno<sup>4</sup>.*

[I have been living in Minsk since September and I have decided to move to Belarus for a long time. In general, Minsk is similar to Ufa, where I came from. It's not large, I like it here. When I lived in Russia, my family used some Ukrainian words – there is some Ukrainian blood in our veins. When I used some of these words, people looked at me sideways and asked to translate them. Here, when it happens, everyone understands me – this is so cool. It seems I am where I need to be.]

There are certain patterns in the assessment of language items by Russian speakers from different countries. For instance, structural closeness of Belarusian

<sup>4</sup> URL: <https://www.facebook.com/humansminsk/photos/a.245920882552009/742456769565082>

and Ukrainian (and in some cases non-linguistic factors, mainly historical and political) lead to the similarities in specific linguistic features of Belarusian and Ukrainian Russian. As a result, there are lexical items perceived as Belarusianisms by Belarusians and as Ukrainianisms by Ukrainians, e.g. *mová* ‘language’ (often about Belarusian or Ukrainian), *nesgrabnyj / nezgrabny(j) / nesgrabnij* ‘clumsy, awkward’, *ščiryj / ščyry / ščirij* ‘true, frank, sincere’ (often about patriots, ironic).

It is interesting that Russians associate some of such words more with Ukraine than with Belarus. For instance, *nezaležnyj* and *nezaležnost* are included into the new Russian academic spelling dictionary<sup>5</sup>. According to corpus data, their frequency in Belarus is not higher than expected, as there is a shift of the sample to Ukraine, and, as a rule, talks about Ukrainian issues in the contexts from Russia. Moreover, the corpus analysis has shown that the frequency of some formal (graphic and phonetic) variants varies by countries, which is caused by language contact. While in Belarus one finds *zababony*, *šufljedka*, *svjadomy(j)* more often, in Ukraine *zabobony*, *šuxljedka*, *svidomy(j)* are more widespread.

## 5. Sociocultural aspect of lexical variation

The usage of the analytical instruments described in this paper leads to the enlarged circle of lexical items characteristic of Belarus. For example, the sample includes the items that have higher frequency in Belarusian Russian than in the dominant variety only due to the extralinguistic factors. As in Belarus the police forces were not renamed (the Soviet name is still used), the frequency of *milicija* is higher than expected ( $G^2=221.69$ ), *policija* – lower than expected ( $G^2=1933.04$ ). Another example: *okroshka* is a dish more popular in Russia than in Belarus (there is another recipe for a traditional Belarusian cold soup), and this is reflected in the frequency of the lexical items: *okroška* in Belarus is used less often than expected ( $G^2=83.31$ ), *xolodnik* (Belarusian dish) – more often ( $G^2=532.53$ ). In addition, according to GICR, Belarusians more often use such words as, for example, *kartoška* ‘potato’, *kolhoz* ‘collective farm’, *partizan* ‘partisan’, which demonstrates discursive differences between the varieties of Russian resulting from cultural, historical, political and other factors, see also (Šajkevič / Savčuk 2014), cultural factors of innovation diffusion in Latvian Russian are described in (Berdicevskis 2014:239).

Earlier, I have discussed the correlation of *Radunica*, *Radonica* and *grobki* in Belarusian, Russian and Ukrainian varieties of Russian. The quantitative analysis has shown that in the Belarusian segment of GICR one can see a statistically significant increase (compared to Russia) in frequency both of a more frequent

<sup>5</sup> URL: <http://orfo.ruslang.ru>.

*Radunica* ( $G^2=437.87$ ), and of a less widespread *Radonica* ( $G^2=21.65$ ). Indeed, this traditional celebration is of a great significance for Belarusians: this is an official holiday in the country, and the subject itself is important for the culture. Due to the same reasons, in the texts created in Belarus one can more often find *Kupal'e* as a word for the holiday celebrated on the day of the summer solstice ( $G^2=402.70$ ).

Detection of lingua-cultural and discursive differences between the varieties of Russian can be an area for further research. In one of his studies, Biber (1987) demonstrates that there are functional differences in grammar of American and British English. The systematization of the discovered distinctive features points to the fact that American written texts are more interactive and abstract compared to British ones. The development of corpora that reflect functioning of Russian in different countries will make it possible to conduct similar studies of scientific, media, literary, and other texts. Apparently, people living in different countries may discuss different topics with different frequency and, as a result, use some lexical items more or less frequently (Geeraerts 2010:823). Recently, scholars note that to understand the specificity of a variety it is important to analyze not only the items that are culturally, ideologically, etc. loaded, but also “ordinary” words (including collocations), see (Baker 2010:66, Koteyko 2014:66; Stubbs 1995:387).

## 6. Conclusion

In general, our study demonstrates that differences between the varieties of Russian are more of a quantitative than qualitative nature. In other words, the specificity of Russian in Belarus, Russia and other countries is determined primarily not by some unique lexical items, but by differences in preferences: given the choice, people from different locations can choose some preferable options, and this is one of the reasons why speakers are not aware of all the differences between the varieties. Therefore, words can be considered Belarusianisms due to their functional characteristics (increase of frequency in Belarus), not only to their origin (borrowings from the Belarusian language). The emergence of lexical Belarusianisms is a result of intra-systemic factors as well as language contacts that need to be analyzed in socio-political and cultural context. The gradual character of differences between the varieties of Russian results from several factors. One of the most obvious extralinguistic factors is human migration, which is reflected in the corpus data: some Belarusianisms are found outside Belarus resulting from the mobility of its speakers. The deeper reasons lie in the nature of linguistic dynamics. For instance, changes in Belarusian Russian arise from both the

internal factors of language development (which means they can be seen throughout the territory where Russian is spoken), and from language contact<sup>6</sup>. At the same time, the boundaries between the internal and contact-induced phenomena are blurred because of the Belarusian-Russian language continuum. Therefore, the gradual character of the differences between the varieties of pluricentric languages results from a non-discrete and fuzzy nature of language.

Finally, the description of differences between the varieties as gradual ones is caused by the research methodology. According to Leech (2015:155), corpus data itself makes the scholars pay attention to the gradual character of linguistic phenomena and search for the ways of their description. Particularly, this becomes possible due to the frequency analysis of language items. In introspective studies, the line between the linguistic phenomena is much clearer, which can be explained by the human's striving for the more precise boundaries. However, reality has a lot of grey areas, and with the help of corpora, linguists can go beyond their own experience and see some new nuances, previously hidden. It must be noted that the gradual character of differences between the varieties of pluricentric languages is not absolute. There is a tendency to some stabilization in the functioning of variants in a country or in some parts of the countries. This is caused by a separate communicative space, within which particular linguistic habits are formed and innovations are spread. In these communicative conditions, the characteristic features of non-dominant varieties of pluricentric languages, in particular Belarusian and Russian, are being established.

## References

- Aikhenvald, Alexandra Y. (2008): *Grammars in Contact: a Cross-Linguistic Perspective*. In: Alexandra Y. Aikhenvald, R. M. W. Dixon (eds.) (2008): *Grammars in Contact: a Cross-Linguistic Typology*. Oxford, Oxford University Press. p. 1-66.
- Backus, Ad (2004): *Convergence as a Mechanism of Language Change*. Vol. 7, Issue 2 of *Bilingualism: Language and Cognition*. p. 179-181.
- Baker, Paul (2010): *Sociolinguistics and Corpus Linguistics*. Edinburgh, Edinburgh University Press.
- Belikov, Vladimir et al. (2013a): *Corpus as Language: from Scalability to Register Variation*. *Dialogue, Russian International Conference on Computational Linguistics, Bekasovo, Vypusk 12 (19)*. Moskva, RGGU. p. 83-95.

---

<sup>6</sup> In general, language contact influences the productivity and frequency of language items (Aikhenvald 2008:22), and changes in frequency are a diagnostic sign of language change (Backus 2004). Under the influence of language contact, language items can change their frequency (if there are no changes in other spheres), for instance, language items become more or less neutral, etc. (Johanson 2002:292).

- Belikov, Vladimir et al. (2013b): Big and Diverse is Beautiful: A Large Corpus of Russian to Study Linguistic Variation. In: Web as Corpus Workshop (WAC-8). URL: <http://corpus.leeds.ac.uk/serge/publications/2013-wac8.pdf/>
- Berdicevskis, Aleksandrs (2014): Predictors of Pluricentricity: Lexical Divergences between Latvian Russian and Russian Russian. In: L. Ryazanova-Clarke (ed.) (2014): The Russian Language Outside the Nation: Speakers and Identities (Russian Language and Society). Edinburgh, Edinburgh University Press. p. 225-245.
- Biber, Douglas (1987): A Textual Comparison of British and American Writing. Vol. 62(2)/1987 of American Speech. p. 99-119.
- Clyne, Michael (1992a): Pluricentric Languages – Introduction. In: Clyne, Michael (ed.) (1992): Pluricentric Languages. Different Norms in Different Countries. Berlin/New York, Mouton/de Gruyter. p. 1-9.
- Clyne, Michael (1992b): Epilogue. In: Clyne, Michael (ed.) (1992): Pluricentric Languages. Different Norms in Different Countries. Berlin/New York, Mouton/de Gruyter. p. 455-465.
- Del Gaudio, Salvatore (2013): Russian as a Non-Dominant Variety in Post-Soviet States: a Comparison. In: Rudolf Muhr et al. (eds.). Exploring Linguistic Standards in Non-Dominant Varieties of Pluricentric Languages. Frankfurt am Main et al., Peter Lang Verlag. p. 343-363.
- Dunning, Ted (1993): Accurate methods for the Statistics of Surprise and Coincidence. No. 19(1)/1993 of Computational Linguistics. p. 61-74.
- Geeraerts, Dirk (2010): Lexical Variation in Space. In: Peter Auer, Jürgen Erich Schmidt (eds.) (2010): Language in Space: An international handbook of linguistic variation. Vol. 1: Theories and methods. Berlin/New York, Mouton de Gruyter. p. 821-837.
- Gorickaja, O. S. (2018) (Goritskaya, Olga): Leksičeskie osobennosti russkogo jazyka v Belarusi, ne svjazannye s vlijaniem belorusskogo jazyka. No. 34 (39) / 2018 of *Respectus Philologicus*. s. 48-60.
- Goritskaya, Olga (2019): Belarusian Russian in a Language Continuum: Contacts and Conflicts. In: Rudolf Muhr / Josep Àngel Mas Castells / Jack Rueter (eds.) (2019): European Pluricentric Languages in Contact and Conflict. Wien et al., Peter Lang Verlag. p. 259-272.
- Goritskaya, Olga / Suprunchuk, Mikita (2018). Frequency Dictionary of Belarusian Borrowings in the Belarusian Variety of the Russian Language. In: SlaviCorp 2018, 24-26 September 2018, Charles University, Prague: book of Abstracts. p. 50-52. URL: [slavicorp.ff.cuni.cz/wp-content/uploads/sites/144/2018/09/SlaviCorp2018\\_Book\\_of\\_Abstracts.pdf](http://slavicorp.ff.cuni.cz/wp-content/uploads/sites/144/2018/09/SlaviCorp2018_Book_of_Abstracts.pdf).
- Hentschel, Gerd (2017): Eleven Questions and Answers about Belarusian-Russian Mixed Speech ('Trasjanka'). Volume 41, Issue 1/2017 of *Russian Linguistics*. p. 17-

42. URL: <https://link.springer.com/article/10.1007/s11185-016-9175-8>.
- Janda, Laura (2016): *Linguistic Profiles: A Quantitative Approach to Theoretical Questions*. Vol. 3/2016 of *Jazyk i metod*. p. 127-146.
- Johanson, Lars (2002): *Contact-Induced Change in a Code-Copying Framework*. In: Mari C. Jones / Edith Esch (eds.) (2002): *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*. Berlin, Mouton de Gruyter. p. 285-313.
- Koteyko, Nelya (2014): *Language and Politics in Post-Soviet Russia: A Corpus Assisted Approach*. New York, Palgrave Macmillan.
- Leech, Geoffrey (2015): *Descriptive Grammar*. In: Douglas Biber / Randi Reppen (eds.) (2015): *The Cambridge handbook of English corpus linguistics*. Cambridge, Cambridge University Press. p. 146-160.
- Muhr, Rudolf (2016): *The State of the Art of Research on Pluricentric Languages: Where We Were and Where We Are Now?* In: Rudolf Muhr, et.al. (eds.) (2016): *Pluricentric Languages and Non-Dominant Varieties Worldwide. Volume 1: Pluricentric Languages Across Continents – Features and Usage*. Wien et. al., Peter Lang Verlag. p. 13-37.
- Rayson, Paul (2002): *Matrix: A Statistical Method and Software Tool for Linguistic Analysis Through Corpus Comparison*. PhD thesis, Lancaster University. URL: <http://ucrel.lancs.ac.uk/people/paul/publications/phd2003.pdf>
- Rayson, Paul / Berridge, Damon / Francis, Brian J. (2004): *Extending the Cochran Rule for the Comparison of Word Frequencies Between Corpora*. In G. Purnelle, C. Fairon, A. Dister (eds.) (2004): *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004)*, Louvain-la-Neuve, Belgium, March 10–12, 2004, vol. II. Louvain-la-Neuve, Presses universitaires de Louvain. p. 926–936.
- Šajkevič, A. Ja. / Savčuk, S. O (2014): *Analiz leksiko-semantičeskix osobennostej regional'noj pressy (na primere gazet grodnenskogo regiona Belarusi)*. In: V. Selegej (ed.) (2014): *Komp'juternaja lingvistika i intellektual'nye texnologii: po materialam ežegodnoj Meždunarodnoj konferencii "Dialog"* (Bekasovo, 4–8 ijunja 2014 g.). Moskva, RGGU. Vyp. 13 (20). s. 585–598.
- Stubbs, Michael (1995). *Collocations and Cultural Connotations of Common Words*. Vol. 7, No. 4 / 1995 of *Linguistics and Education*. p. 379-390.
- Wardhaugh, Ronald (2010): *An Introduction to Sociolinguistics*, 6th ed. Chichester / Malden, Wiley-Blackwell.
- Woolhiser, Curt (2012): *"Belarusian Russian": Sociolinguistic Status and Discursive Representations*. In: Rudolf Muhr (ed.) (2012): *Non-dominating Varieties of Pluricentric Languages. Getting the Picture. In Memory of Prof. Michael Clyne*. Wien et. al., Peter Lang Verlag. p. 227-262.