DOI: 10.15201/hungeobull.70.3.4 Hungarian Geographical Bulletin **70** 2021 (3) 249–266.

Uneven geographies in the various language editions of Wikipedia: the case of Ukrainian cities

Oleksiy GNATIUK[1] and Victoria GLYBOVETS[1]

Abstract

The paper tackles the issue of uneven geographical representations on Wikipedia, the most visible and powerful user-generated encyclopaedia. In particular, it addresses language imbalances on Wikipedia with regard to geographical information and uneven spatial patterns of territory coverage on the different language versions in an attempt to verify expectations about the cultural factors that influence these imbalances and uneven spatial patterns. Ukraine is a promising case for testing the formulated expectations, as it has a large number of neighbouring countries, and most of them had political and cultural influence on its territory in the past. The volumes (word counts) of articles about the Ukrainian cities were analysed for seven language versions of Wikipedia, including the Ukrainian version and the versions of all bordering countries. The results show that historical geography is the strongest and central factor, and most of the key relic borders (former boundaries) can be traced. Ethnic composition appears to be another important factor, although weaker than the previous one. The role of the border factor is often unclear, but in some cases it definitely makes an impact and therefore cannot be completely ignored. Thus, the geographies of Wikipedia are not indifferent to the issues of ethnicity and geopolitics. The research calls into question the ability of modern Wikipedia to be a reliable and balanced source of geographical knowledge, as the described imbalances may create lopsided and biased geographical representations in people from different countries and nations.

Keywords: Wikipedia, geographical representations, uneven geographies, language inequalities, word count, cultural factors, Ukraine

Received April 2021, accepted July 2021

Introduction

Wikipedia is an international online project which attempts to create a free encyclopaedia in multiple languages, collaboratively and successfully edited by plenty of volunteers (Voss, J. 2005; Mamadouh, V. 2019a, b). Simultaneously, Wikipedia is a convenient and accessible source of information for millions of people around the globe. Geographical information, referring to particular places (countries, regions, cities and villages, etc.), is not an exception here. Today, when people search for information about a certain geographical location, they often use an Internet search engine, and one of the first search results will commonly be the article on Wikipedia (Lewandowski, D. and Spree, U. 2011).

However, unlike traditional encyclopaedias, the content of Wikipedia is created not by professionals but by ordinary users. Not surprisingly, the accuracy, comprehensiveness and balance of information on Wikipedia are often questioned. This applies especially to such culturally and geopolitically sensitive issues as history and geography. The worlds represented on Wikipedia are affected by those who write these representations of local places in specific languages and for specific audiences (Mamadouh, V. 2019a, b; Osborne, C. *et al.*

[1] Taras Shevchenko National University of Kyiv, Department of Economic and Social Geography, 01601, Kyiv, Ukraine. E-mails: alexgnat22@ukr.net, victoriasatiya@gmail.com

2021). Although factors defining the uneven geographies on Wikipedia have been already addressed in the literature in the global dimension (Graham, M. *et al.* 2014; Dittus, M. and Graham, M. 2019; Mamadouh, V. 2019b), there is a need to tackle the issue in greater depth in relation to individual factors and regions of the globe. In this article, we leave aside the accuracy and reliability of the content of geographical representations on Wikipedia and focus instead on their spatial and language imbalances. In particular, we deal with (1) imbalances between the language versions with regard to geographical information, (2) uneven spatial patterns of territory coverage in the language versions, and (3) factors of predominantly cultural origin that influence these imbalances and patterns.

Theoretical background

Wikipedia, positioned by its creators as a free encyclopaedia, is one of the world's most visible, most used, and most powerful repositories of user-generated content (Graham, M. *et al.* 2014). At the same time, Wikipedia is one of the predominant ways in which internet users obtain knowledge about the world (Dittus, M. and Graham, M. 2019). In particular, it contains "a massive cloud of geographic information about millions of events and places around the globe put together by millions of hours of human labor" (Graham, M. *et al.* 2014). Thus, it can be asserted that Wikipedia plays an important role in the construction of geographical imaginations of various types of places in the minds of Internet users (Graham, M. *et al.* 2014; Gribok, M.V. and Tikunov, V.S. 2019).

Wikipedia is particularly interesting to cybermetric research, not least because of the richness of phenomena and full accessibility of the data (Voss, J. 2005). In particular, a large corpus of literature addresses the issues of quality and reliability of information presented on Wikipedia, giving particular attention to the fact that it can be edited by anyone who wishes to do so (Stvilia, B. *et al.* 2005; Ortega Soto, J.F. 2009; Javanmardi, S. and Lopes, C. 2010; Mamadouh, V. 2019b). Researchers have opposing opinions on this issue: from strong questioning, suspicion and disrespect by academic circles (López Marcos, P. and Sanz-Valero, J. 2013; Jemielniak, D. and Aibar, E. 2016) to conclusions about the high credibility of information in different fields of knowledge, which is ensured by multiple mutual controls of many users (Giles, J. 2005; Rosenzweig, R. 2006; Kittur, A. and Kraut, R. 2008; Mesgari, M. *et al.* 2015; James, D. 2016; Michelucci, P. and Dickinson, J.L. 2016; London, D.A. *et al.* 2019). Being by most measures the most widely read knowledge repository on Earth, Wikipedia is often treated as unworthy of academic attention (Jemielniak, D. 2019), although some opposite experiences have been already presented (e.g. Selwyn, N. and Gorard, S. 2016; Konieczny, P. 2017; Di Lauro, F. and Johinke, R. 2017).

While traditional encyclopaedias are broadly unbiased thanks to the involvement of reputable specialists, credible sources, and appropriate editorial gate keeping, it should be acknowledged that the community of Wikipedia editors (or simply Wikipedians) have also elaborated their own quality control procedures. In particular, Wikipedia is built on collective efforts and consensus seeking; the diverging viewpoints should be taken care of through deliberation and argumentation; the full history of editions and discussions is open, public, and archived. To prevent various kinds of vandalism, editors have an important role in controlling new edits and policing articles under their supervision; bots (web robots) are also involved in registering suspected activities and preventing damage to the articles. The quality assurance mechanisms also include the special statuses for distinct articles like "good articles" and "features articles" (Mamadouh, V. 2019b). However, the editorial mechanisms and standards of quality vary widely among Wikipedia projects (Jemielniak, D. and Wilamowski, M. 2017). The basic reasons are the diverging number of editors involved and

different cultural traditions regarding hierarchy and autonomy (Hara, N. *et al.* 2010; Mamadouh, V. 2019b). While the most developed versions draw on a very large number of volunteers and, consequently, a lot of reads and corrections, less developed versions usually depend on a few (but very active) volunteers. Some versions (e.g. Volapük and Cebuano) have expanded dramatically using machine translation through the work of bots generating articles by translating them automatically from the other Wikipedias, although the value of such articles is questioned by some Wikipedia editors who prefer quality to quantity (Mamadouh, V. 2019b).

Geographic representations on Wikipedia are not an exception; they greatly depend on their creators and audiences, and therefore they are asymmetric and biased both spatially and in terms of content. Graham, M. *et al.* (2015), addressing this issue, distinguish uneven geographies of access, participation, and production. The most typical phenomenon is a self-focus bias (Hecht, B. and Gergle, D. 2009, 2010a, b), when "articles about places, people, and events where the language of the edition was spoken were more prominent than those in other regions" (Hale, S. 2014, 99). This is one of the reasons why different language versions of Wikipedia have different quality of coverage with regard to specific regions of the globe. Consequently, the most prominent articles about local places and events are often (but not always) written in local languages (Sen, S.W. *et. al.* 2015; Kim, S. *et al.* 2016).

At the same time, in many parts of the world, socioeconomic realities and digital divides constrain participation in Wikipedia editing (Dittus, M. and Graham, M. 2019), which causes numerous exceptions to the rule. A significant number of people are being excluded from the collective process of knowledge production due to technical, social, economic, political, regulatory, and infrastructural barriers that arise often solely on the basis of their native language (van Dijk, Z. 2009; Friedman, U. 2016; Osborne, C. *et al.* 2021). That is why the geography of articles related to the geographical places is highly uneven and clustered in developed countries, and simultaneously, large areas of the developing world remain invisible (Graham, M. *et al.* 2014). For example, there are more geotagged articles in the Netherlands than in Africa as a whole (Graham, M. *et al.* 2014). In the global context of today's digital knowledge economies, these digital absences are likely to have very material effects and consequences (Graham, M. *et al.* 2014). Similarly, the analysis of timelines of national histories across Wikipedia language versions showed that narratives about national histories are distributed unevenly across the continents with a significant focus on the history of European countries. Also, national historical timelines vary across language editions, although average inter-lingual consensus is rather high. In this sense, Wikipedia's historical reference articles are not free from gaps and biases (Samoilenko, A. *et al.* 2017).

Furthermore, the uneven involvement of people from different countries and regions in the editing of Wikipedia contributes to the language imbalances: some languages are overrepresented, while some other are represented more than modestly (Dittus, M. and Graham, M. 2019). Especially this refers to the dominant position of English, currently being the most powerful global language and de-facto standard language of the Internet (Danet, B. and Herring, S.C. 2007). It has been shown that for many countries in the Global South, which includes Africa, Asia, and South America, there are more articles written in English than in the respective native languages (Graham, M. *et al.* 2014; Dittus, M. and Graham, M. 2019).

Research expectations

Relying upon the literature (Graham, M. *et al.* 2014; Kim, S. *et al.* 2016; Dittus, M. and Graham, M. 2019; etc.), we assumed that articles on Wikipedia about geographical places are written mainly by three groups of authors: (1) locals who have knowledge

about the place, as well as strong physical and/or mental attachment to it, and want to convey this knowledge to a wide range of users, (2) people that are not locals but who are interested in the particular place due to the cultural ties shaped by the national, cultural, professional identity, etc., and (3) specialists in one topic (for example geography, history, etc. of cities or regions) that are not particularly interested in a specific city. Also, to contribute to a specific language version of Wikipedia, the author must be proficient in the respective language, and articles in such a language will be targeted primarily at the speakers of respective language, and therefore will be devoted to those places and aspects that are of interest to these speakers. On the other hand, the audience is important: people living in a city/region/country usually need more detailed information about that place than people living far away; the Wikipedia community does not promote the translation of articles without localization in the societal context associated with the language to serve the intended audience (Mamadouh, V. 2019b). Thus, it is assumed that each of the Wikipedias is focused primarily on geographical places that are related to the geography, history and culture of the respective nations and countries.

Taking into account the abovementioned remarks, we formulated three research expectations. The first expectation is that there should be a correlation between the ethnic/linguistic composition of the population of a given place and the size of article in the respective language version of Wikipedia about this place. The second expectation is about the positive correlation between the distance from a given place to the border of the country and the size of articles in the respective language version of Wikipedia about this place. The third expectation implies that places that sometime in the past where under the political and cultural influence of a particular state or ethnic group should be more widely represented in respective Wikipedia than the places having no common political and cultural background.

Data and methods

Ukraine is a promising case for checking the outlined expectations. It borders a large number of neighbouring countries, and in the past its territory has been under their political and cultural influence. Among the countries having land borders with Ukraine, only Belarus and Moldova have never politically controlled a part of Ukrainian state territory (in this context, we refer not to modern states but to their predecessors). Also, Ukraine is not a mono-ethnic state: sizeable national minorities live on its territory, including titular ethnic groups of neighbouring countries.

We analysed six versions of Wikipedia in the official languages of countries that have a land border with Ukraine, in particular the Russian, Polish, Romanian, Belarusian, Hungarian and Slovak versions. The Ukrainian version was covered by the study as well. It is important to keep in mind that these are language versions and not national versions, and therefore they are serving not only people from the respective countries but the whole language audiences. In this way, the Polish version serves a Polish audience, concentrated predominantly in Poland, of which the ethnic Poles in Ukraine are a tiny minority; the same applies to the Hungarian and Slovakian versions. On the other hand, the Russian version serves a transnational audience of Russian speakers across the world, especially from the former Soviet Union countries (not only Russians in Russia or Russian speakers in Ukraine), while the Romanian version of Wikipedia now serves principally the Romanian-language audience in both Romania and Moldova. The separate Moldavian Wikipedia in Cyrillic alphabet was closed because the Moldovan language was found to be a version of Romanian (even according to the 1989 Language Law of Moldova), and there is a software to navigate the two scripts (Mamadouh, V. 2019a). Regarding the Belarusian version, it may be supposed that it is principally serving nationally minded Belarusians all over the world,

while the majority of people in Belarus prefer the Russian version as it is better understood and more developed. Finally, the Ukrainian version serves not only the audience in Ukraine, but the vast Ukrainian diaspora.

Among all the geotagged articles related to geographical places in Ukraine, we focused on the articles about the cities. In this manner we clearly defined and shortened the list of scrutinized articles. At the same time, today's human activities are mostly tied to cities, and public representations about countries and regions are often constructed under the lens of urban geographies. In total, articles about 457 cities were analysed. As a rule, the content of the articles includes information blocks on the city's site and situation, physical geography (relief, climate, soils, flora, fauna, etc.), history, contemporary demographics and economic development, culture, transport, social sector, landmarks and prominent personalities, etc.

The key analysed parameter was the volume of an article, defined as a word count of the main text, including the captions of the illustrations and the lists of notes and references, but without the side inserts. When the article about a particular city is absent, the volume of the article is equated to 0 (zero word articles). Here we supposed that the volume of the article correlates with the amount of information contained in this article, thus, being indicative of the potential usefulness of the article for readers, i.e. the difference in word counts translates into differences in quality. The objection here could be the fact that the volume of the article depends on the city's size: the bigger a city, the larger the expected volume of the article. However, this rule is neither strong nor linear (cf. Gribok, M.V. and Tikunov, V.S. 2019), and the dependence function varies between different language versions of Wikipedia. That is why we decided to avoid the use of relative indices, such as ratio of the article volume to the city population, but to supplement the main parameter with two additional ones. First, the mean volume of the article for each language version was calculated for administrative regions of Ukraine (regions and main cities are shown in *Figure 1*; Crimea was 'de facto' annexed by the Russian Federation in 2014 but is claimed by Ukraine and recognized as Ukrainian by the United Nations, affirming the territorial integrity of Ukraine within its internationally recognised borders, and by most other countries). In this way the fluctuations in article volumes for cities of different sizes were smoothed out within the regions, and general trends could be seen more easily. Second, the rank of the articles by volume among the seven analysed language versions was defined for each language version. This means that for each specific city, the language version with the largest volume of the article received the 1st rank, the next – the 2nd rank, and so on until the language version with the smallest article that received the 7th rank (zero word articles were subjects for ranking as well, being assigned the 7th rank). Here the absolute size of the article is substituted with the ratio of the volume of different language versions of the same article, and in this manner articles about cities of different size may be compared. Also, the ranking approach makes visible the relationship between language versions for particular cities or regions, often revealing subtle but important trends and differences. Thus, the ranking was used (1) to show disproportions between the different language versions in the national dimension, and (2) to reveal the uneven relationship between language versions in the regional dimension.

The obtained patterns were compared with the factors that may influence the situation according to the initial expectations *(Figure 2).*

First, the factor of ethnicity: the share of the respective ethnic groups in each administrative region is shown in the form of cartograms. The data are taken from the 2001 census; for the Romanian language, cumulative share of Romanians and Moldovans is shown. It is expected that the higher share of a particular ethnic group in a city/region should correspond to more extensive articles on the respective Wikipedia, because of a larger number of local Wikipedians.

Fig. 1. Administrative territorial division of Ukraine and main cities. Notes: Administrative centres of the regions are written in capital letters. The names of the regions correspond to the names of their centres except for Volyn region (centre in Lutsk), Zakarpattia region (centre in Uzhhorod) and Crimea (centre in Simferopol). The Crimea was annexed by Russian Federation in 2014, but is claimed by Ukraine and recognized as Ukrainian by the United Nations. Kyiv and Sevastopol are the cities of the special status (equated to the regions), and the city of Slavutych is an exclave of the Kyiv region. For this research they were counted as belonging to Kyiv region, the Crimea, and Chernihiv region respectively.

Second, the factor of historical geography: hatching denotes areas controlled by the respective states in the past. For the Slovak language, the area controlled by Czechoslovakia in 1920–1939 is shown. For countries such as Russia and Poland, different types of hatching show gradations of impact. In particular, for Poland these are the territories controlled by the Second Polish Republic (1921–1939) and by the Polish-Lithuanian Commonwealth (from the 15th century to 1792), and for Russia the lowest level of influence was determined for the regions of Western Ukraine annexed to the USSR only after 1939, the high level for the left bank Ukraine (obtained by Russia under the Truce of Andrusovo in 1667), the Black Sea region (densely settled during the Russia-led colonization in 18–19th centuries), and the highest level for Crimea (transferred to the Soviet Ukraine only in 1954 and annexed by the Russian Federation in 2014). It is expected that cities/regions with such historical ties to other countries should be of greater interest to the Wikipedians from these countries. This factor has no sense for Ukrainian Wikipedia or, indeed, for the Belarusian one as the Belarusian state has never owned any part of the contemporary Ukrainian territory.

Third, the factor of a border: the maps show the borders with respective countries. For the Romanian language, the borders of Romania and Moldova are shown; for the Russian language, the borders of the Russian

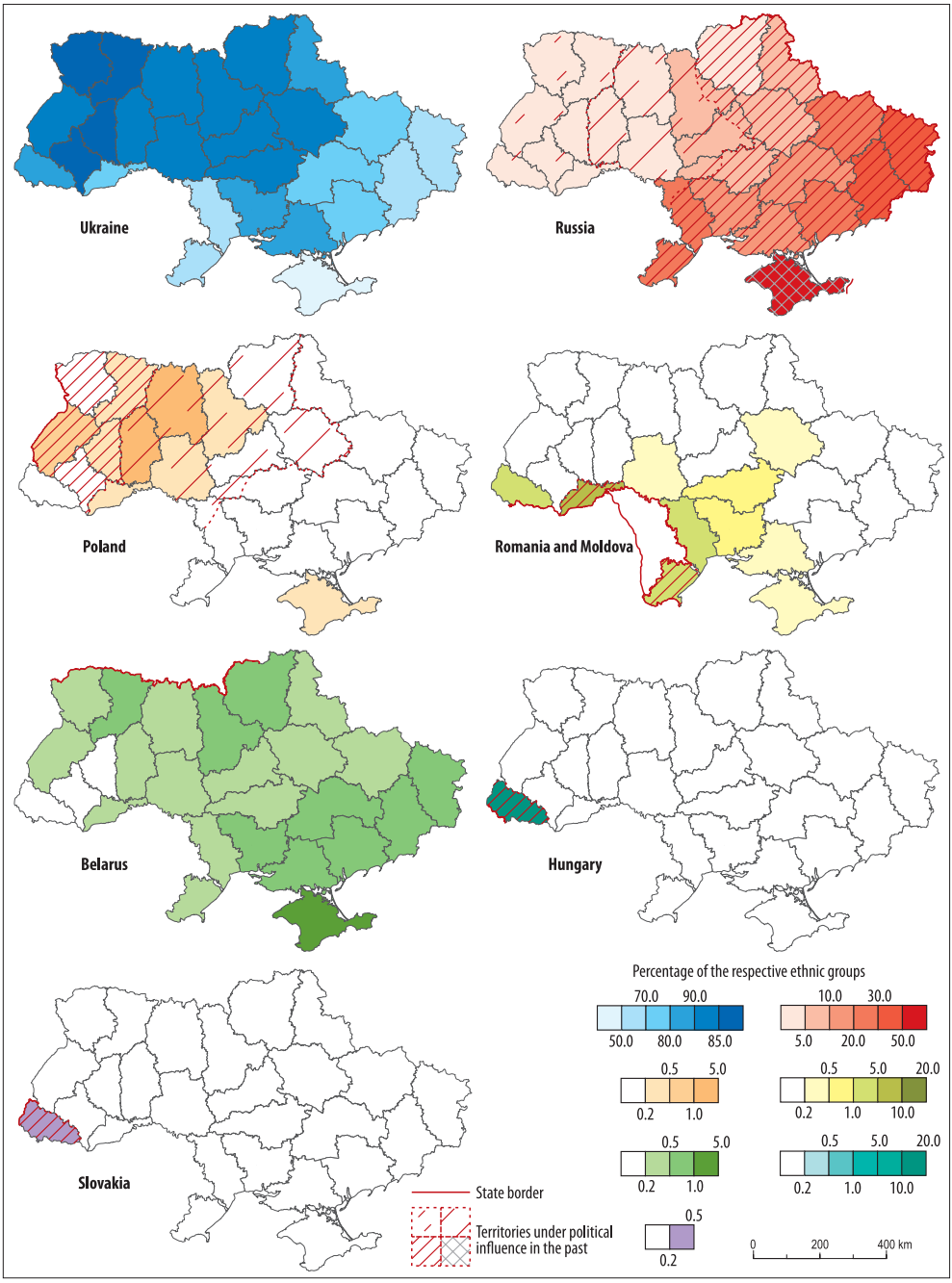

Fig. 2. Factors that potentially influence geographical representations on Wikipedia

Federation and Transnistria (Pridnestrovian Moldavian Republic) are shown. The expectation is a higher level of interest among Wikipedians living in each bordering country in nearby cities/regions of Ukraine (compared to the more distant ones).

It is worth remembering that these factors mostly are not fully independent: the ethnic structure often is shaped by the historical geography, and the border areas are often settled by the respective ethnic group or were controlled in the past by the neighbouring state. Also, some other factors may influence the representations on certain language versions, such as the size (and number of editors) of a particular version of Wikipedia, the use of bots-generated or translated content, and huge ethnic diasporas.

Results and discussion

The further analysis is divided into two subsections. In the first subsection, we discuss the revealed imbalances of representation between the language versions of Wikipedia in the national dimension, leaving aside the differences between the regions. In the second subsection we consistently consider each language version, focusing on the interregional differences in representation, as well as on the interregional variations in the relationship between language versions (using the ranking). At the end of this subsection, the results are compared with the initial research expectations. The results in terms of the research parameters are shown in *Figures 3, 4,* and *5.*

General imbalance between the language versions

The first important observation is a substantial imbalance in the coverage of Ukrainian cities by the studied language versions of Wikipedia (see *Figures 3, 4* and *5*). The basic statistical parameters are given in the *Table 1*: mean, median, maximal and minimal values of the volume of article (in words). Also, the coefficient of variation (CV) has been calculated for the volume of articles for each language version to assess the uniformity of the representation of cities: the lower the CV, the higher the observed uniformity, and vice versa. The last two columns show the percentage of cities with articles of less than 100 words (which can be considered uninformative) and the percentage of the cities with no article at all (zero word articles).

It is seen from the table that the longest articles are typical of the Ukrainian and Russian versions. Also, these language versions show the lowest coefficients of variation, which means that all cities across the country are more or less evenly reflected; in particular, there are no articles shorter than 100 words and there are no cities without an article. The leading position of the Ukrainian and Russian versions is explained by the leading role of the respective languages: Ukrainian is the official and the most widespread language; Russian takes the second place by the number of speakers, and it is still widely used as the lingua franca in the post-Soviet space. They are followed by the Polish and Belarusian versions with medium volume of articles, greater variation of values and a certain percentage of very short articles (less than 100 words). Although articles in Romanian are available for almost all cities, a high proportion of articles contain less than 100 words. This means that the vast majority of these articles are uninformative. Interestingly, most of such uninformative articles have been created by bots using the standardized template. The less elaborated are the Hungarian and Slovak versions: if all cities are taken into account, they have the lowest average volume of the articles, and only circa 20 percent of cities are reflected in these language versions. However, if we narrow the view to the actually existing articles (excluding zero volume articles), their average volume will be comparable to the Romanian and Belarusian versions. This means that while the Romanian Wikipedia provides limited information but on merely every city, the Hungarian and Slovak versions contain sufficiently expanded

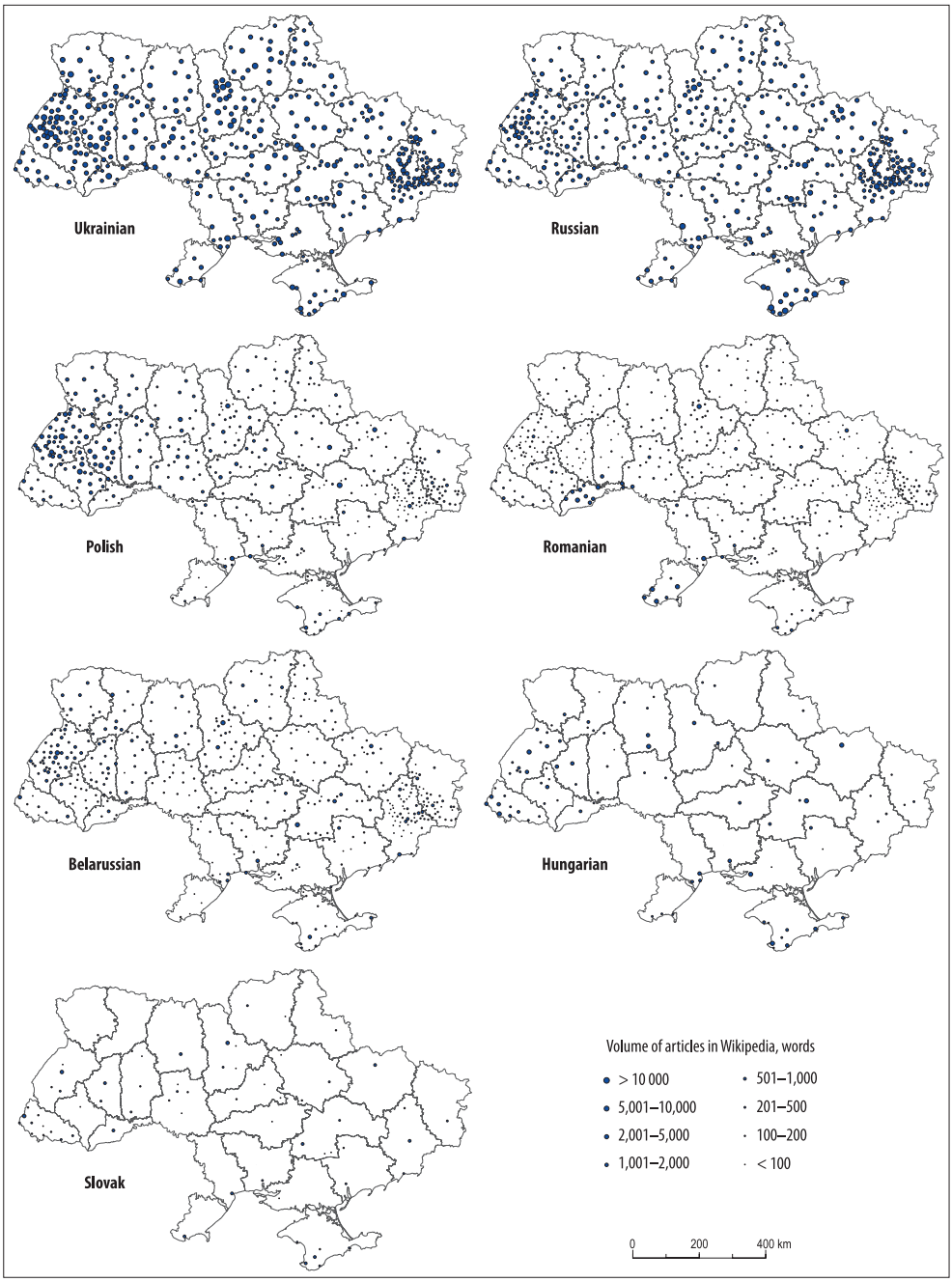

Fig. 3. Volume of the articles about cities on Wikipedia

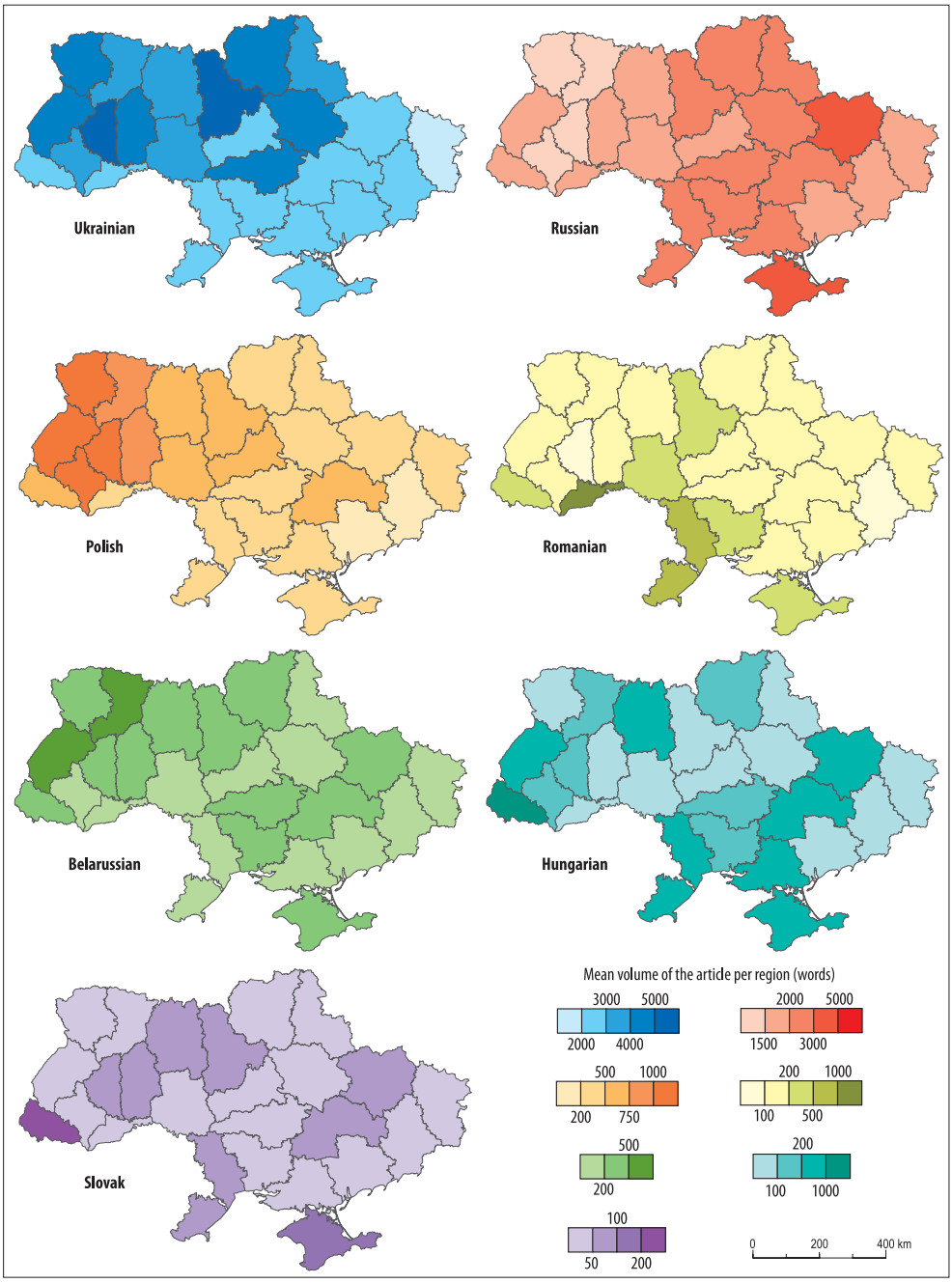

Fig. 4. Mean volume of the articles about cities on Wikipedia

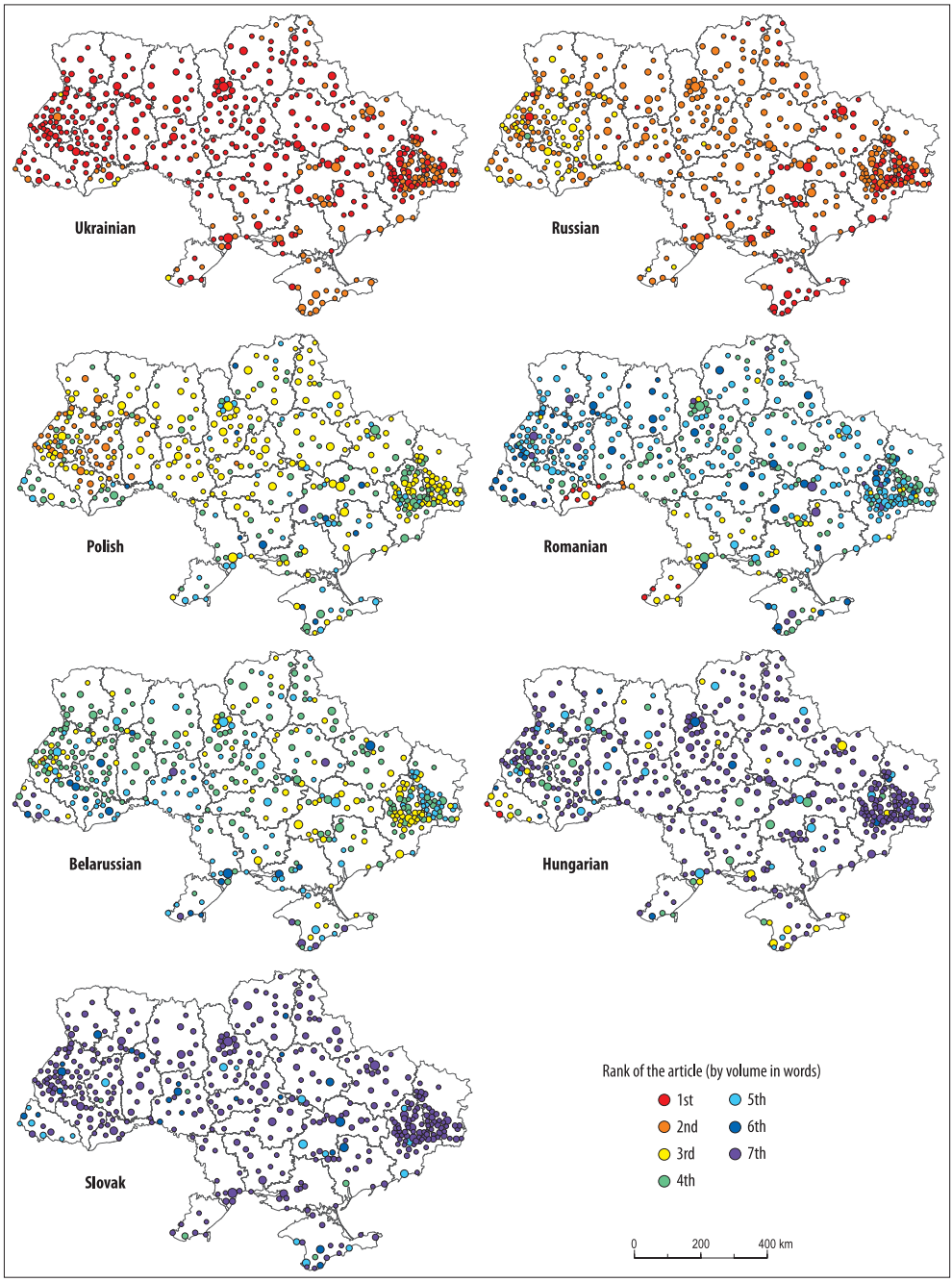

Fig. 5. Rank of the articles about cities on Wikipedia

Table 1. Article statistics for the language versions of Wikipedia

Language	Mean	Median	Max	Min	CV	V < 100, %	V = 0, %
Ukrainian	3,424	2,339	21,161	254	0.89	0.0	0.0
Russian	2,072	1,351	16,195	179	1.11	0.0	0.0
Polish	591	257	12,798	9	1.68	15.8	0.0
Belarusian	269	157	2,436	0	1.13	18.4	0.0
Romanian	206	113	3,869	8	2.08	34.6	1.8
Hungarian	157	0	4,791	0	3.31	81.0	79.6
Slovak	43	0	1,265	0	3.46	89.1	85.3

information (contributed by the human editors rather than bots) but for a smaller group of cities. The distribution of ranks leads to similar conclusions.

The overall size of the different Wikipedia versions varies greatly, so it is to be expected that the larger ones feature more information (also about Ukrainian cities) than the smaller ones. In reality, the described differences of representation do show clear a correlation with the size of these versions. In particular, the Russian Wikipedia has 1,700,000+ articles, Polish – 1,400,000+ articles, Ukrainian – 1,000,000+ articles, Hungarian and Romanian – 400,000+ articles, Slovak and Belarusian – 200,000+ articles. The most obvious differences are the better positions of Ukrainian (because it is the official and native language for the majority of Ukrainian citizens), Belarusian (probably due to the common post-Soviet context and the ease of translation from Ukrainian or Russian), and Romanian (due to the extensive use of bots for the creation of articles) versions. The positions of the Russian and Polish versions may have been further strengthened by the sizeable Ukrainian diasporas in these countries.

The findings suggest that people from different nations who choose Wikipedia in their native language as a source of geographical knowledge will have different opportunities to access knowledge about a specific area. Not only will users of some language versions receive less information about the same geographical location, but also most of the places will simply not exist for them (cf. Graham, M. 2009 on virtual "terra incognita"). This calls into question the usefulness of present-day Wikipedia as a source of geographical knowledge – at least for certain language versions and for certain territories.

Uneven geographies of representation on the language versions

Besides the general imbalances between the studied Wikipedias, each of them has specific territorial imbalances of coverage within the country, and most of these spatial patterns fit one or more initial research expectations.

Although the correlation between the share of ethnic Ukrainians and the mean volumes and ranks of the articles in the Ukrainian Wikipedia is not strong, both maps *(Figures 1* and *4)* clearly show the same fault line between the west and centre of the country, where the proportion of ethnic Ukrainians is higher than 80 percent, and the rest of the country, where this proportion is less than 80 percent (the only exception is the Cherkasy region with a high share of ethnic Ukrainians but shorter articles in the Ukrainian Wikipedia). Also, Ukrainian-language articles rank almost exclusively 1st to the West and North of this fault line, while they often rank 2nd or even 3rd to the East and South of it. The extreme case in this regard is Crimea, where the article in Ukrainian ranks 1st for only one city. This fault line is well known to researchers addressing issues of Ukrainian geopolitics, in particular electoral patterns (Osipian, A.L. and Osipian, A.L. 2012; Diesen, G. and Keane, C. 2017). Although the current differences in the share of ethnic Ukrainians are themselves determined by ancient geopolitical and natural boundaries, Wikipedia's content is

directly influenced by the modern ethnic composition of the local editors, and therefore we consider the described pattern as evidence of the ethnic factor's influence.

In the Russian Wikipedia, the highest volume of articles is observed for Crimea, the region with the most powerful political and cultural ties with Russia (and where Russians are a dominant ethnic group). It is followed by the Black Sea region (where both ethnicity and historical geography are major factors) and the northern part of the left-bank Ukraine (where the factor of historical geography is the most important). The lowest volume of articles is observed in the regions of Western Ukraine that were annexed by the Soviet Union after 1939 *(Figures 3* and *4).* Russian-language articles rank 1st in Crimea (except for Yevpatoria), interchangeably 1st or 2nd in the other regions of the south-east and in the extreme north-east, predominantly 2nd in central Ukraine and mainly interchangeably 2nd and 3rd in that part of Western Ukraine annexed by the Soviet Union after 1939. An especially low rank of the Russian-language articles is observed in the Galician regions (Ternopil, Ivano-Frankivsk, Lviv), which together with the Zakarpattia and Chernivtsi regions were not controlled by the Russian Empire *(Figure 5).* The influence of ethnicity and borderline factors can also be traced, but to a much lesser extent.

The mean volume of articles in the Polish Wikipedia decreases with increasing distance from the Polish border – from the west to the south-east of Ukraine. A more detailed look reveals the influence of historical geography. In particular, the eastern border of the Second Polish Republic (which included contemporary Lviv, Ivano-Frankivsk, Ternopil, Volyn and Rivne regions) is still visible on the maps *(Figures 3*, *4* and *5)*: the mean volume of articles here generally exceeds 1,000 words, and many articles have the 2nd rank, overtaking the Russian version. This is especially true for the three Galician regions, where Polish articles rank 2nd for more than half of the cities. The influence of the Kingdom of Poland and the Polish-Lithuanian Commonwealth can also be observed, although less obvious at first glance. In particular, a significant decrease in the volume of articles takes place with the transition of the Dnieper, i.e. from the right-bank to the left-bank Ukraine *(Figures 3* and *4)*, that is west of the former eastern border of the Commonwealth *(Figure 2).* However, it is important to remember that the map *(Figure 2)* shows the most eastern position of the border, but after the Truce of Andrusovo (1667) it generally passed along the Dnieper, and the right-bank Ukraine has been under Polish political influence for a longer time than the left-bank Ukraine. On the contrary, those areas of Western Ukraine that have not been under Polish rule (the Zakarpattia and Chernivtsi regions) are less covered by the Polish Wikipedia both in terms of volume and rank of the articles. Although the area that is best covered by the Polish Wikipedia generally overlaps with the area with the highest proportion of ethnic Poles, the factor of ethnicity is clearly not the crucial one here. If this were the case, we should have expected the best coverage in the Zhytomyr and Khmelnytskyi regions, which are home to the largest number of Ukrainian Poles.

In the Romanian Wikipedia, two territories are clearly distinguished against the general background: the Chernivtsi region and the southern part of the Odessa region within the historical area of Bessarabia. Here, the mean volume of articles is approximately 10 times higher than the average values across Ukraine *(Figures 3* and *4).* The average rank of the articles in Romanian is also significantly higher than in the rest of the country. In particular, the Romanian Wikipedia ranks first for the majority of cities in the Chernivtsi region and for three among 19 cities in the Odessa region; the minimum rank of Romanian Wikipedia in these areas does not fall below the 3rd *(Figure 5).* Both of these territories were controlled by the Principality of Moldavia from the 14th to 18th centuries, and by Romania in 1918–1940 and 1941–1944, and have the highest proportions of the Romanian/Moldovan population in Ukraine.

At the same time, the rather high share of Romanians in Zakarpattia is not accompanied by such a significant increase in the volume of articles, although Transcarpathia also boasts in terms of the Romanian Wikipedia more articles than do the other regions. The other Ukrainian territories with a relatively high share of Romanians/Moldovans, including the Mykolaiv and Vinnytsia regions, Crimea and so forth also typically show higher volumes and ranks of articles in the Romanian Wikipedia; however, this correlation is not very strong. The influence of Romanian/Moldovan border is debatable: on the one hand, the majority of regions bordering Romania/Moldova have better representations of the cities in Romanian Wikipedia compared to the rest of the country, but on the other, the Ivano-Frankivsk region (which has no significant Romanian/Moldavian minority and has never been controlled by the Romanian or Moldovan state) breaches the rule. Thus, for the Romanian Wikipedia, the historical geography factor is most important, followed by the factor of ethnicity, and the influence of the border factor is only in third place and is of questionable significance.

Belarus has never politically controlled any part of Ukraine. Thus, in the case of the Belarusian Wikipedia, we can ignore the factor of historical geography and take a closer look at the two other factors: border and ethnicity. Three groups of regions have a relatively better representation in the Belarusian Wikipedia: (1) Western Ukraine, (2) the regions near the Belarusian border (in the northern part of Ukraine), and (3) the regions of the southeast of Ukraine *(Figure 4).* The same three groups of regions are distinguished in the terms of ranks *(Figure 5).* The second group may be explained by both the higher share of Belarusians in the population and the direct proximity to the Belarusian border *(Figure 2).* The third group is distant from the Belarusian border, but the share of ethnic Belarusians is the highest in these very regions, therefore the factor of ethnicity may be an explanation here. However, the good representation of Western Ukraine in the Belarusian Wikipedia cannot be explained by the considered factors; probably, Western Ukraine, especially the Lviv region, is interesting for a Belarusian audience as a vibrant touristic area. Another possible explanation for the interest of Belarusian Wikipedians in Western Ukrainian cities could be an inspiration to strengthen their national language against Russian.

As for the Hungarian and Slovak Wikipedias, the interregional differences in the volume of actually existing articles are not so impressive *(Figure 3).* However, Zakarpattia (Transcarpathia) is the only region where absolutely all cities have their articles in these language versions. If we consider that most cities in Ukraine have no Hungarian and Slovak articles at all (i.e. their volume is equated to zero), Zakarpattia will stand out sharply against other regions in terms of the average volume of articles *(Figure 4)*. The same applies to ranks *(Figure 5).* While in the other regions the Hungarian Wikipedia most often ranks 7th (sometimes 5–6th, very seldomly 4th or 3rd), in Zakarpattia two cities have the 1st rank, and six cities – the 2nd rank (out of a total of 11 cities). Similarly, while in the other regions the Slovak Wikipedia is often ranked 7th (sometimes 5–6th, very rarely 4th), in Zakarpattia, six out of 11 cities have 5th rank. This result is perfectly consistent with the historical geography of Zakarpattia, as it was for many centuries a part of the Hungarian state (from the 11th century until the Treaty of Trianon in 1920, and also in the World War II period of 1939–1944) and was a part of Czechoslovakia in the interwar period (1920–1939). It should also be noted that the political influence of Hungary and Czechoslovakia has never spread to any other region of Ukraine. Zakarpattia is also the only region bordering modern Hungary and Slovakia and having a significant representation of relevant national minorities. However, the favourable positions of the Hungarian and Slovak Wikipedia are observed for the entire Zakarpattia region, not only for those parts that lay closer to the respective state borders; also, there are no signs

of better elaborated Slovak articles in the southwest of the Lviv region, which is very close to the Slovak border. Another observation is the absence of a strong correlation between the number of ethnic Hungarians and Slovaks and the mean volumes and ranks of the Wikipedia articles across the other Ukrainian regions. That is why, for both language versions, we may consider the historical geography factor to be the central one, the factor of ethnicity to be also influential, although less important, and the border factor as less significant and rather unclear.

The estimated influence of all three studied factors is summarized in *Table 2*.

The table shows that the factor of historical geography is the strongest and the central one, as its influence is clearly traced in all five cases when this factor is relevant. The factor of ethnicity appears to be also important, although weaker than the previous one. Finally, the role of the border factor is often unclear; in two cases it is estimated as weak, and only in one case (the Belarusian Wikipedia) as strong. Interestingly, this is the exact case when the historical geography factor is eliminated. Therefore, although the border factor cannot be completely ignored, we can definitely assert its relative weakness compared with the other two factors.

Nevertheless, these factors to a greater or lesser extent may contribute to the uneven geographical representations in the linguistic versions of Wikipedia. That means that people from different nations, using Wikipedia in their native language as a source of geographical knowledge, are receiving uneven spatial representations of the real world. For example, Poles, being well informed about Western Ukraine, receive limited information about the south-eastern part of the country, and for Slovaks or Hungarians the vast majority of the country, with the exception of a few islands, will be "terra incognita". Given the nature of the factors considered, this applies in particular to neighbouring countries/nations/cultures having a complicated history of mutual relationships, including territorial exchanges in the past. The geographies of Wikipedia are not indifferent to nationality and geopolitics; they are mirroring ethnic identities and exhibit phantom boundaries no worse than the election results.

Conclusions

The research shows the uneven geographical representation of Ukrainian cities on Wikipedias written in the official languages of countries bordering Ukraine, as well as in the Ukrainian Wikipedia. The revealed patterns are well explained by the two factors: historical geography (the strongest one) and ethnicity (less strong). The third presumed albeit ambiguous factor is the distance to the border of the respective country. Also, the study documented significant disproportions in the amount of information between the language versions caused, first of all, by the differences in their size (and, respectively, the number of active editors). However, a shared recent history (e.g. the common experiences of Ukraine, Russia and Belarus in

Table 2. Influence of factors on the language versions of Wikipedia

Language	Ethnicity	Historical geography	Border
Ukrainian	strong	not relevant	not relevant
Russian	weak	strong	weak
Polish			
Belarusian	strong	not relevant	strong
Romanian	weak	strong	unclear
Hungarian	strong		
Slovak			

the post-Soviet space) and contemporary social and cultural ties (e.g. the presence of large Ukrainian diasporas in Poland and Russia) contribute to the better representation of Ukrainian urban geography on the respective Wikipedias. The editorial policies and mechanisms of different Wikipedias are important as well, as shows the example of Romanian Wikipedia ballooned via the use of bots-generated geotagged articles.

Jemelniak, D. (2019) expressed a hope that "in 2019 Wikipedia turned 18, so maybe academics should start treating it as an adult". However, nowadays language versions of Wikipedia often behave like disengaged, discordant and obsessed teenagers. Our research confirmed the risk that Wikipedia "might not just be reflecting the world, but also reproducing new, uneven, geographies of information" (Graham, M. *et al.* 2014). The different language versions of Wikipedia, taken separately, constitute neither objective nor impartial sources of information. Even being based on purely quantitative research methods and leaving aside the content-related issues, our research calls into question the ability of Wikipedia to be a reliable and balanced source of geographical knowledge. The imbalances and uneven spatial patterns create lopsided and biased geographical representations in people from different countries and nations, which in the conditions of modern information society may have negative economic and social effects. Further research is required in this field before the next step can be taken with a switch to the biases in content, reflecting the subjective view of the Wikipedia editors and audiences – the bearers of a certain cultural traditions, geopolitical ideas and representations about the 'true' versions of history and, consequently, geography of the own country and the surrounding world. The edit wars on Wikipedia, reflecting controversies with regard to the selection and rendering of historical periods and current affairs, are another promising topic for further research, particularly in the geopolitically divided country that Ukraine currently represents. The first shoots of such academic investigations (see e.g. Rogers, R.A. and Sendijarevic, E. 2012; Jemielniak, D. 2014; Kumar, S. 2017; Kopf, S.E. 2018) need further development.

Billions of people do not have access to free knowledge, and expanding the corpus of knowledge on Wikipedia is an effective way to feel this gap (Jemielniak, D. 2019). Thus, Wikipedia editors, including representatives of academia, must try hard to overcome the imbalances and to substantially improve Wikipedia's quality with regard to geographical representations.

REFERENCES

Danet, B. and Herring, S.C. 2007. *The Multilingual Internet: Language, Culture, and Communication online*. Oxford, Oxford University Press.

Diesen, G. and Keane, C. 2017. The two-tiered division of Ukraine: historical narratives in nation-building and region-building. *Journal of Balkan and Near Eastern Studies* 19. (3): 313–329. Doi: 10.1080/19448953.2017.1277087.

Di Lauro, F. and Johinke, R. 2017. Employing Wikipedia for good not evil: Innovative approaches to collaborative writing assessment. *Assessment & Evaluation in Higher Education* 42. (3): 478–491. Doi: 10.1080/02602938.2015.1127322.

Dittus, M. and Graham, M. 2019. Mapping Wikipedia's geolinguistic contours. *Digital Culture & Society* 5. (1): 147–164. Doi: 10.14361/dcs-2019-0109.

Friedman, U. 2016. The lopsided geography of Wikipedia. Founder Jimmy Wales discusses the barriers to the encyclopedia's expansion. In *Atlantic,* June 21, 2016. Available at https://www.theatlantic.com/international/archive/2016/06/geography-wikipedia-jimmy-wales/487388/

Giles, J. 2005. Internet encyclopaedias go head to head. *Nature* 438. (7070): 900–901. Doi: 10.1038/438900a.

Graham, M. 2009. Wikipedia's known unknown. In *The Guardian,* December 2, 2009. Available at https://www.theguardian.com/technology/2009/dec/02/wikipedia-known-unknowns-geotagging-knowledge

Graham, M., Hogan, B., Straumann, R.K. and Medhat, A. 2014. Uneven geographies of user-generated information: patterns of increasing informational poverty. *Annals of the Association of American Geographers* 104. (4): 746–764. Doi: 10.1080/00045608.2014.910087.

Graham, M., De Sabbata, S. and Zook, M. 2015. Towards a study of information geographies: (im) mutable augmentations and a mapping of the

geographies of information. *Geo: Geography and Environment* 2. (1): 88–105. Doi: 10.1002/geo2.8.

Gribok, M.V. and Tikunov, V.S. 2019. Wikipedia as a data source for studies of collective mental representations of geographical objects (examplified by the cities of the Russian Arctic zone). *Izvestiya Russkogo geograficheskogo obshestva* 151. (4): 50–60. (In Russian). Doi: 10.31857/S0869-6071151450-60.

Hale, S. 2014. Multilinguals and Wikipedia editing. In *WebSci 2014: Proceedings of the 2014 ACM Web Science Conference*, 99–108.

Hara, N., Shachaf, P. and Hew, K.F. 2010. Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology* 61. (10): 2097–2108.

Hecht, B. and Gergle, D. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the Fourth International Conference on Communities and Technologies, C&T '09*. New York, 11–20. Doi: 10.1145/1556460.1556463.

Hecht, B. and Gergle, D. 2010a. On the "localness" of user-generated content. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 229–232.

Hecht, B. and Gergle, D. 2010b. The Tower of Babel meets Web 2.0: User-generated content and its applications in a multilingual context. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI '10*. New York, 291–300.

James, R. 2016. WikiProject medicine: Creating credibility in consumer health. *Journal of Hospital Librarianship* 16. 344–351. Doi: 10.1080/15323269.2016.1221284.

Javanmardi, S. and Lopes, C. 2010. Statistical measure of quality in Wikipedia. In *Proceedings of the First Workshop on Social Media Analytics*. July 25–28, 2010. Washington D. C., District of Columbia, 132–138. Doi: 10.1145/1964858.1964876.

Jemielniak, D. 2014. *Common Knowledge: An Ethnography of Wikipedia*. Stanford, Stanford University Press.

Jemielniak, D. 2019. Wikipedia: Why is the common knowledge resource still neglected by academics? *GigaScience* 8: 1–2. Doi: 10.1093/gigascience/giz139.

Jemielniak, D. and Aibar, E. 2016. Bridging the gap between Wikipedia and academia. *Journal of the Association for Information Science and Technology* 67. 1773–1776. Doi: 10.1002/asi.23691.

Jemielniak, D. and Wilamowski, M. 2017. Cultural diversity of quality of information on Wikipedias. *Journal of the Association for Information Science and Technology* 68. 2460–2470. Doi: 10.1002/asi.23901.

Kim, S., Park, S., Hale, S., Kim, S., Byun, J. and Oh, A.H. 2016. Understanding editing behaviors in multilingual Wikipedia. *PLoS ONE* 11. (5): e0155305. Doi: 10.1371/journal.pone.0155305.

Kittur, A. and Kraut, R. 2008. Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, November 08–12, 2008. San Diego, CA, USA, 37–46. Doi: 10.1145/1460563.1460572.

Konieczny, P. 2017. Joining the global village. Teaching globalization with Wikipedia. *Teaching Sociology* 45. (4): 368–378. Doi: 10.1177/0092055X17714030.

Kopf, S.E. 2018. *Debating the European Union Transnationally – Wikipedians' Construction of the EU on a Wikipedia Talk Page (2001–2015)*. Lancaster, Lancaster University.

Kumar, S. 2017. A river by any other name: Ganga/Ganges and the postcolonial politics of knowledge on Wikipedia. *Information, Communication & Society* 20. 809–824. Doi: 10.1080/1369118X.2017.1293709.

Lewandowski, D. and Spree, U. 2011. Ranking of Wikipedia articles in search engines revisited: Fair ranking for reasonable quality? *Journal of the American Society for Information Science and Technology* 62. (1): 117–132. Doi: 10.1002/asi.21423.

London, D.A., Andelman, S.M., Christiano, A.V., Kim, J.H., Hausman, M.R. and Kim, J.M. 2019. Is Wikipedia a complete and accurate source for musculoskeletal anatomy? *Surgical and Radiologic Anatomy* 41. (10): 1187–1192. Doi: 10.1007/s00276-019-02280-1.

López Marcos, P. and Sanz-Valero, J. 2013. Presencia y adecuación de los principios activos farmacológicos en la edición española de la Wikipedia. *Atención Primaria* 45. (2): 101–106.

Mamadouh V. 2019a. Wikipedia: mirror, microcosm, and motor of global linguistic diversity. In *Handbook of the Changing World Language Map*. Eds.: Brunn, S. and Kehrein, R., Cham, Springer, 3730–3756. Doi: 10.1007/978-3-030-02438-3_200.

Mamadouh, V. 2019b. Writing the world in 301 languages: A political geography of the online encyclopedia Wikipedia. In *Handbook of the Changing World Language Map*. Eds.: Brunn, S. and Kehrein, R., Cham, Springer, 3757–3780. Doi: 10.1007/978-3-319-73400-2_199-1.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F.Å. and Lanamäki, A. 2015. "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology* 66. 219–245. Doi: 10.1002/asi.23172.

Michelucci, P. and Dickinson, J.L. 2016. The power of crowds. *Science* 351. (6268): 32–33. Doi: 10.1126/science.aad6499.

Ortega Soto, J.F. 2009. *Wikipedia: a quantitative analysis*. Doctoral Thesis. Madrid, Universidad Rey Juan Carlos.

Osborne, C., Graham, M. and Dittus, M. 2021. Edit wars in a contested digital city: mapping Wikipedia's uneven augmentations of Berlin. *The Professional Geographer* 73. (1): 85–95. Doi: 10.1080/00330124.2020.1800493.

Osipian, A.L. and Osipian, A.L. 2012. Regional diversity and divided memories in Ukraine: Contested past as electoral resource, 2004–2010.

East European Politics and Societies 26. (3): 616–642. Doi: 10.1177/0888325412447642.

Rogers, R.A. and Sendijarevic, E. 2012. *Neutral or national point of view? A comparison of Srebrenica articles across Wikipedia's language versions.* Berlin, Wikipedia Academy: Research and Free Knowledge.

Rosenzweig, R. 2006. Can history be open source? Wikipedia and the future of the past. *The Journal of American History* 93. (1): 117–146. Doi: 10.2307/4486062.

Samoilenko, A., Lemmerich, F., Weller, K., Zens, M. and Strohmaier, M. 2017. Analysing timelines of national histories across Wikipedia editions: a comparative computational approach In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media, ICWSM 2017*, 210–219.

Selwyn, N. and Gorard, S. 2016. Students' use of Wikipedia as an academic resource – Patterns of use and perceptions of usefulness. *Internet and Higher Education* 28. (1): 28–34. Doi: 10.1016/j.iheduc.2015.08.004.

Sen, S.W., Ford, H., Musicant, D.R., Graham, M., Keyes, O.S. and Hecht, B. 2015. Barriers to the localness of volunteered geographic information. In *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* April 2015, 197–206. Doi: 10.1145/2702123.2702170.

Stvilia, B., Twidale, M.B., Smith, L.C. and Gasser, L. 2005. Assessing information quality of a community-based encyclopedia. In *Proceedings of the International Conference on Information Quality, ICIQ 2005*, 442–454.

van Dijk, Z. 2009. Wikipedia and lesser-resourced languages. *Language Problems & Language Planning* 33. 234–250.

Voss, J. 2005. Measuring Wikipedia. In *Proceedings of ISSI 2005, 10th International Conference of the International Society for Scientometrics and Informetrics.* Eds.: Ingwersen, P. and Larsen, B., Stockholm, Karolinska University Press, 24–28.