COGNITIVE STUDIES | ÉTUDES COGNITIVES, 22
Warsaw 2022

Article No.: 2770
DOI: 10.11649/cs.2770

Citation: Hentschel, G., & Palinska, O. (2022). Restructuring in a mesolect: A case study on the basis of the formal variation of the infinitive in Ukrainian–Russian Surzhyk. *Cognitive Studies | Études cognitives*, *2022*(22), Article 2770. https://doi.org/10.11649/cs.2770

Gerd Hentschel, **Olesya Palinska**

Carl von Ossietzky University, Oldenburg, Germany
{gerd.hentschel,olesya.palinska}@uni-oldenburg.de

Restructuring in a Mesolect: A Case Study on the Basis of the Formal Variation of the Infinitive in Ukrainian–Russian Surzhyk

Abstract

In addition to Ukrainian and Russian, Ukraine is linguistically characterized by a Ukrainian–Russian mixed speech called Surzhyk. Given the background of Ukrainian–Russian relations and the emancipation of Ukrainian from the previously dominant Russian, Surzhyk has become the subject of an emotional discussion in independent Ukraine. The majority of Ukrainian scholars working with pre-Labovian and implicit theoretical sociolinguistic (and contact linguistic) models view the distribution of Ukrainian and Russian elements in Surzhyk as spontaneous and chaotic. Furthermore, Surzhyk – together with many who use it – has been widely stigmatized, even by linguists, as a post-colonial legacy from the times of Russian and Soviet dominance.

Taking as an example the forms of verb infinitives, a corpus-based quantitative analysis of about 10,000 instances evidences that Surzhyk shows a considerable degree of stabilization in the use of competing morphological forms. This stabilization can be interpreted best as an instance of structure building in a mesolect between Ukrainian dialects on the one hand and, on the other hand, Russian and (to a certain degree) Ukrainian standard languages in competing roles during the recent history of Ukraine.

Keywords: bilingualism; dialect levelling; code-mixing; fused lects; colonial hybridization; Surzhyk

Funding

This paper is based on material from field work in three projects: (i) *Variability and stability in a mixed substandard in extensive and time-stable language contact: The Ukrainian–Russian Surzhyk in the Ukraine*, funded by the Fritz-Thyssen-Foundation (no. 10.14.1.066); (ii) Inflectional irregularity in contemporary contact varieties of North-Slavic languages, funded by the Deutsche Forschungsgemeinschaft (DFG, no. 155014374), headed by Gerd Hentschel; (iii) *Hybridization from two sides? Ukrainian–Russian and Russian–Ukrainian code mixing in the context of the (socio-) linguistic situation in Southern Ukraine along the Black Sea coast* (DFG no. 419468937, FWF no. I 4189-G30, an ongoing project, headed by Gerd Hentschel and Tilmann Reuther, Klagenfurt, Austria).

1 Introduction

Even today, large parts of the linguistic landscape in Ukraine can be characterized as bilingual and diglossic[1] (Ukrainian and Russian), although in western oblasts Russian has to a considerable degree decreased in prominence during the decades of the independent Ukrainian state. Of course, there are other languages spoken in Ukraine, such as Polish in western oblasts or Bulgarian and Moldovan in the south on the Black Sea coast. For these groups of speakers in Ukraine one may speak of trilingualism or triglossia.[2] This paper, however, is solely concerned with Ukrainian–Russian language contact. In this respect two points should be underlined:

Firstly, on the level of standard or literary languages the young Ukrainian language and the more established Russian language have coexisted in an asymmetric constellation of bilingualism since the second half of the 19th century. In the last decades of the Russian Empire and in the Soviet era (with the exception of the 1920s), there was generally a very strong asymmetry favouring Russian (cf. Taranenko, 2007). Political endeavours towards the linguistic Russification of Ukraine started even in the middle of the 17th century, before the emergence of the modern standard languages (cf. Danylenko & Naienko, 2019). Language policies in post-Soviet independent Ukraine, however, have at least legally and institutionally strengthened the position of Ukrainian. The last step in this direction was the adoption of a new language law by the Ukrainian parliament (Verkhovna Rada) in April 2019 and its ratification by the outgoing president Poroshenko in May. It came into force in July 2019. Nevertheless, both languages are strongly present and – metaphorically speaking – to some degree competing in Ukrainian society, at least in large parts of the country (cf. Hentschel & Taranenko, 2021). The Ukrainian language conflict is one of the modern stereotypes about independent Ukraine, yet it seems to be a conflict occurring at the level of the political and cultural elites in Ukraine and Russia (cf. Hentschel & Brüggemann, 2015). The alleged conflict has been instrumentalized by Kremlin propaganda against Ukraine as a state and by Ukrainian elites with a national, if not nationalistic, orientation against Russian as the language of the former colonial oppressor and contemporary aggressor. Doubtlessly, endeavours to emancipate the Ukrainian Standard language and to strengthen its role in various communicative contexts in independent Ukraine are more than understandable. However, the vast majority of the Ukrainian population, Ukrainian and Russian speaking, takes a rather relaxed position on these questions (cf. Hentschel & Zeller, 2016 for central Ukraine[3]).[4]

Today there are clear differences in the distribution or preference constellations of the two languages across the vast Ukrainian territory. There are several minority languages in the country, which will not be included in the following discussion. Ukrainian is traditionally seen as dominating in the west, with Russian dominant in the east and south (cf. *Vseukraïns′kyĭ perepys naselennia*, 2001). A recent study by Hentschel and Taranenko (2021) suggests that things are much less clear cut than usually presented in studies by Ukrainian social science institutes like KIIS (Kyïvs′kyĭ mizhnarodnyĭ instytut sotsiolohiï, "Ispol′zuemyĭ iazyk", 2003, as cited in "IAzyki Ukrainy", n.d.;

[1] The traditional differentiations of diglossia and bilingualism (analogously, triglossia and trilingualism) definitely show several shortcomings for a sound characterization of complex constellations of coexistence of two or more languages (codes) in one society (cf. Jaspers (2017) in general and Hentschel (2017, pp. 32–35) for the Belarusian situation, which is in many respects similar to the Ukrainian one, but in many others different – see below, fn. 17). Nevertheless, the old dichotomy of diglossia and bilingualism is still useful here, at least for an initial approximation.

[2] Cf. for example the study on Ukrainian–Russian–Polish trilingualism with Ukrainians of non-Polish descent by Levchuk (2020).

[3] There are several, broader and narrower, conceptions of the "centre" of Ukraine. The two authors, as generally in the Oldenburg investigation on Ukrainian Surzhyk, delimit the centre in their investigations in a broader way (see below).

[4] Please note that this paper was written before the Russian invasion of Ukraine in February 2022. Attitudes towards Russian and the language question in Ukraine have most probably changed in the meantime.

Kyïvs′kyĭ mizhnarodnyĭ instytut sotsiolohiï [KIIS], 2019), and are obviously in flux. However, the well-known historically-based tendencies of the increasing strength of Russian and simultaneously decreasing strength of Ukrainian from the west to the east and towards the south (the Black Sea) remain. The same holds for regions near the Russian border.

Secondly, the diglossic aspect of the linguistic landscape of Ukraine can be seen on the one hand in the existence of traditional Ukrainian dialects under the "roof" of at least one of the two standard languages. These areal dialects are still mainly spoken in the countryside. On the other hand, a colloquial mixed Ukrainian–Russian subvariety exists, which has received a name of its own: Surzhyk (cf. Taranenko, 2004, 2014). This variety is the subject of this paper. It is spoken by millions of people, at least in informal conversational settings. This variety or social dialect (as it is also traditionally assumed to be) can mainly be found in small and medium-sized towns in central Ukraine (cf. Taranenko, 2007, p. 125). More precisely, Hentschel and Taranenko (2021) give evidence that Surzhyk is especially strong in eastern parts of central Ukraine and in western parts of the south coast. Surzhyk is often spoken by the same individuals alongside Ukrainian and/or Russian, which are preferred in more public and formal settings (cf. Hentschel & Zeller, 2016, 2017). Surzhyk has been heavily stigmatized as a reflection of Russian colonialism, again by representatives of elites with a clearly national disposition (cf. Stavyc′ka, 2014).

The donor codes for Surzhyk are, of course, Ukrainian and Russian. As for the former, one must differentiate between the literary or standard language and the various areal dialects. Dialects are of special relevance because it is widely acknowledged (cf. Taranenko, 2014, p. 270) that Surzhyk (as well as its Belarusian-Russian equivalent Trasjanka, cf. Zaprudski, 2007, pp. 105–107) developed when the rural population moved into towns and cities in various waves of industrialization and steadily growing urbanization (under Imperial Russian rule until 1917 or later under Soviet rule), where they had to adapt linguistically to Russian. Today it is beyond doubt that many Ukrainians, even if they mainly spoke rural Ukrainian dialects during childhood, have a good, if not excellent, command of Russian. This follows the lines of the regionally varying strength of Russian mentioned above. In the Ukrainian variant of Russian there may be, of course, several phonetic features or specific words of Ukrainian origin (cf. Kamusella, 2018). Such local colour is normal in spoken variants of European Standard languages. A large number of Ukrainians, however, have preserved Surzhyk as an informal vernacular, at least for the private sphere. Regarding Russian, the literary language (be it in a colloquial variant) is of central importance as a donor code for Surzhyk, as it dominated almost all official or public spheres of communication and therefore had an enormous impact on education. Some social subvarieties of Russian (e.g. the Russian *prostorechie*) may also have played a certain role, especially in industrial centres which saw an influx of workers from other parts of Russia or the Soviet Union. However, these varieties obviously did not function as a point of linguistic orientation for the new Ukrainian city dwellers.[5]

The delimitation of what is Ukrainian and what is Russian in Surzhyk is, of course, already complicated by the fact that both languages are genetically and structurally closely related. Of special interest for the discussion of mixed or fused lects are deeper layers of linguistic structuring than phonetics and phonology (cf. Hentschel, 2008) but even then, many common, interlingual Ukrainian–Russian "diamorphs" remain. A further complication is related to the triangular constellation of donor codes: Standard Ukrainian – Standard Russian – Ukrainian dialects. Many structural phenomena in the latter, or at least a subset of them, are congruent with the Russian Standard and not with the Ukrainian Standard. This, for example, is the case with the forms of the infinitive (see details below). The variation of infinitive forms in Surzhyk, with its double donor background (Ukrainian dialect or Standard Russian), will serve as the phenomenon to illustrate the general points we intend to make. The infinitive is a widespread and salient phenomenon in speech and as such it is an ideal object for a case study.

[5] For example, the inflectional morphology in Belarusian Trasjanka, although clearly exhibiting Russian influence, does not include typical inflectional traits of Russian *prostorechie*, unless they already have analogies in literary or dialectal Belarusian (cf. Brandes, 2015).

2 General Aims

The first general aim is to interpret the position of Surzhyk in the linguistic "architecture" of Ukraine. Of course, one can qualify it as a sociolect or social dialect (cf. Taranenko, 2014) in a broader sense: not as a dialect used by people of a certain social status, but as a colloquial variety for informal conversational settings. (As a matter of fact, Surzhyk is present in a broad spectrum of social groupings, of course with some differences – cf. Hentschel & Taranenko, 2015; Hentschel & Zeller, 2017). This view, however, ignores two aspects. Firstly, Surzhyk has to a large degree replaced traditional areal dialects (cf. Taranenko, 2013, pp. 48–49). This is definitely the case with many people who migrated from villages to towns and their children, if not their grandchildren. The latter two groups at least partially lost the command of old rural dialects. To what extent Surzhyk has replaced traditional areal dialects in the Ukrainian countryside is still an open question. Secondly, if (as is widely and correctly assumed – see above) Surzhyk arose in the course of migration from rural to urban settings, then the dialectal background should be reflected in the mixed speech, i.e. in the fusional process with the standard languages, not least with Russian. In other words, it is more than doubtful that there is a (Ukrainian-based – see below) Surzhyk that is free of Ukrainian dialectal traits. Surzhyk thus should to some degree reflect differences in the dialectal background and thus exhibit areal differences itself. If this can be shown, then Surzhyk could be seen as a further instance of fusions of autochthone dialectal substrata and standard varieties, i.e. of the development of regionally restricted mixed sub-varieties if not "regiolects" (or regional vernaculars),[6] which are to be found in many areas of Europe as an epiphenomenon of the retreat and loss of old rural dialects – for example in southern Germany (cf. Schmidt & Herrgen, 2011). The specific situation in Ukraine would consist in the fact that for at least a century there have been two competing standard languages (Ukrainian and Russian), and not, as for example in Germany, just the one.

Such a view has been proposed by Hentschel (2013, 2014) on empirical grounds for the Belarusian equivalent to Surzhyk, Trasjanka, and hypothetically extended to Surzhyk by Hentschel and Brüggemann (2015). This means that, although characterized by a high degree of spontaneous variation, Surzhyk (like Trasjanka) should exhibit a considerable regularization (stabilization, reduction of variation), though possibly in Ukraine with regional differences. This is plausible, given the fact that Surzhyk is the linguistic code that millions of people have grown up with.[7] For the second and following generations of speakers of Surzhyk, it is not the case that they first learn Ukrainian and mix it when trying to learn and speak Russian. Rather, they first acquire Surzhyk and later, starting at the latest in school, they more or less successfully develop the competence or filters to supress elements from one of the donor codes when they are expected to speak the other one (cf. Hentschel, 2017).

[6] Of course, there are further regional varieties of Standard or Literary Ukrainian that show traits of autochthone areal dialects of Ukrainian.

[7] As Hentschel and Zeller (2017, pp. 53–54) and Hentschel and Palinska (2022, ch. 4) have shown for central Ukraine and the three Black Coast oblasts respectively, throughout these areas at least 20 percent of more than 1,000 randomly selected respondents in each of the two areas named the mixed code of Surzhyk as their code of first linguistic socialization (before school). A slightly higher proportion named Ukrainian or Russian additionally. It is thus easy to interpolate that for millions of people Surzhyk was without any doubt the central code of first linguistic socialization, although the two other codes mentioned were certainly present. This holds the more so, because due to stigmatization of Surzhyk and its speakers, not least by several Ukrainian linguists, it is safe to assume that some respondents felt ashamed to clearly name Surzhyk as the central code in their family during early childhood and preferred to name Ukrainian or Russian instead.

The second general aim of this paper is to try to provide evidence for the mesolectal[8] status of Surzhyk, which shows a considerable structural stabilization with regional differences and socially conditioned variations: a sort of socially conditioned mesolect between autochthone areal dialects and the standard languages, resulting from both dialect levelling and the influence of the two standard languages. In this respect, the southern oblasts along the Black Sea coast are of special interest. Here there are (apart from rather small, peripheral areas) no traditional autochthonous Ukrainian dialects, compared with the centre or the west of Ukraine. The Slavic settlement of this part of Ukraine, mainly by people from Ukrainian and Russian dialectal areas, started only after its conquest by the Russian Empire at the end of the 18th century.

3 The Corpus

The corpus of Surzhyk our analysis is based on can be divided into four subcorpora. The dividing features are (a) region: centre vs. south, (b) the settings in which linguistic material was collected in the field: family conversations vs. interviews. Where necessary, we thus differentiate between the CentreFam, SouthFam, CentreInt and SouthInt subcorpora.[9] For the analyses differentiating between centre and south, it should be noted that there is not only a regional difference but a temporal one as well. In the south, field work was conducted in 2020 (family) and 2021 (interviews). In the centre, however, field work was conducted in 2010 (family) and 2014 (interviews).[10] All four corpora contain between 170,000 and 200,000 word forms, stemming from between 55 and 70 speakers.[11] Thus, altogether there is a corpus of about 750,000 word forms from more than 200 speakers.

The family corpora (CentreFam, SouthFam) contain conversations in which family members, friends, colleagues and neighbours took part. The corpora of interviews (CentreInt, SouthInt) contain transcribed fragments of so-called open (semi-structured, "deep") interviews, in which respondents outlined their views on the language question in Ukraine, their own linguistic biography, attitudes and preferences for the choice between languages and codes, and the role of languages for Ukrainian culture, religion, education and statehood. The respondents considered for the open interviews form a subset of 1,400 participants in the centre and 1,290 in the south who were recruited after completing a closed fully-structured interview (opinion poll[12]) because they had stated that they use Surzhyk regularly. It should be noted that the respondents in the interview corpora were neither in contact with each other nor with respondents taking part in the recorded family conversations.

The material collected from each individual respondent varies in extent, especially in the family corpus. This is due to the fact that in family conversations there are usually a few protagonists

[8] The term *mesolect*, though originally coined for an intermediate variety between *basilect* and *acrolect* in a post-creole continuum, has been adopted for mixed speech with Russian as the acrolect in the last decade (cf. Woolhiser, 2011, pp. 26–28). The latter however, in contrast to the former, are highly stigmatized in the corresponding societies.

[9] The subcorpora from the centre were established on the basis of the first two grants mentioned above, the ones from the south from the third grant mentioned.

[10] A full documentation for both corpora and the corpora themselves will be published electronically in 2023 (e.g. on the website of the authors' department). The conceptual design of the corpora for Surzhyk largely follows that developed for Belarusian-Russian Trasjanka; cf. Hentschel et al. (2014).

[11] As a matter of fact, in the family conversations there were far more participants than the numbers mentioned: about 30 in both the centre and in the south. Their contribution to the material in both regions was minimal, see below. These "peripheral participants" contributed less than 500 word forms each, quite often less than 100, because they took part in the family conversations only occasionally. We have not included their socio-biographic information (apart from their places of residence) in the quantitative analysis to follow, as the material compiled from these individuals makes up only 3 percent of the total.

[12] Analyses of these opinion polls have been presented by e. g. Hentschel and Taranenko (2015, 2021), Hentschel and Zeller (2016, 2017), Zeller et al. (2019).

who dominate the conversation. The share of contributions by the other family members tends to be lower and that of occasional participants (neighbours, friends) sometimes much lower. The variation in respondents' size of contribution in the interviews depends on the selection criteria adopted: Firstly, for both corpora only those fragments of discourse in which the utterances showed a considerable degree of mixing of Ukrainian and Russian expressions were selected for narrow transcription. Secondly, within the transcripts for linguistic analysis only mixed utterances (mostly sentences or sometimes syntactically incomplete, shorter utterances) were considered, i.e. utterances in which at least one specific morph from each of the two donor languages or a hybrid morph occurred. Utterances with specific morphs from only one of the two donor languages and / or common morphs were not considered. There are many common "diamorphs" in the mixed speech from the two closely related languages. The determination of morphs (and composed units such as word forms, phrases etc.) as specific, common or hybrid was based on a "deeper" layer of morphophonemic representation. Phonetic and (surface) phonological phenomena were neglected (cf. Hentschel, 2008; Menzel & Hentschel, 2017, pp. 156–164).[13] The interviews, in contrast to the family conversations, are certainly not instances of completely spontaneous speech. Respondents tended to try to speak "pure" Ukrainian or, less often, "pure" Russian, especially at the start. If, however, phenomena in family conversations and in interviews overlap or do not contrast, the interview material can be relied on to the same degree as the conversational material for making generalizations.

The corpora in the large central[14] parts of Ukraine were compiled only from residents of smaller towns (including so-called "town-like settlements") and cities up to 1,000,000 inhabitants. This means that neither village dwellers, nor inhabitants of metropoles (i.e. Kyiv, Kharkiv, Dnipro[15] were included. Concentrating data selection on places of roughly medium size was based on the widely accepted assumption that these were the central 'melting pots' where Surzhyk developed first and is most stable. Metropoles in the area (Kyiv, Kharkiv, Dnipro) are considered broadly Russian-speaking (cf. for example Taranenko, 2007, p. 131), although this may be changing to a certain degree at present. Village dwellers are reported to still often practise traditional Ukrainian dialects, although this most probably varies in different areas of the country (cf. Hrytsenko, 2015).[16]

Most of the places where data were collected are located in the area of the south-eastern dialectal group (Ukr. *narichchja*), one of the three main Ukrainian dialectal groups. For our analysis, the regional dialectal distribution of the endings of the infinitive will be more important than the general dialectal partition of Ukrainian (see below).

It should be noted that the collection of data in the south of Ukraine had a higher density than in the centre, since it included only the three oblasts on the Black Sea coast. In the centre there were eleven oblasts. This was of course motivated by the differences in the general aims of the three grants. One important point was that due to the fact that the central region encompasses only areas with a traditional autochthonous Ukrainian dialectal base, a "traditional" Ukrainian-based

[13] Among other things, this means that, for example, an utterance that only consists of Russian and perhaps common morphs, but which has been pronounced with widely Ukrainian phonetic traits, would not be considered a mixed utterance.

[14] Note that the "central" parts of Ukraine where the data for this project were collected encompass a somewhat larger territory than is mostly understood under "central Ukraine", e.g. Khmelnytskyi in the west and Kharkiv in the east.

[15] The cities Dnipropetrovsk and Kirovohrad (but not the oblasts) have been renamed recently; in the text, correspondingly, we use Dnipropetrovsk and Kirovohrad as names of the oblasts, Dnipro and Kropyvnytskyi as names of the cities.

[16] Without any doubt, the linguistic situation in Ukrainian villages deserves a special investigation. Due to the expected great diversity among them this was not feasible within the framework of the two aforementioned grants focusing on the central regions. Villages were included in the south, since they were part of the special focus of interest of that grant, but this material will not be considered in the analyses in this paper, since it is not comparable.

Surzhyk should dominate there. This has been confirmed by Hentschel and Zeller (2016, 2017) and Hentschel and Taranenko (2015). On the other hand, the Black Sea coast oblasts are traditionally described as overwhelmingly Russian speaking (cf. KIIS, 2003, as cited in "IAzyki Ukrainy", n.d.; KIIS, 2019). Furthermore, in a strict sense, the dialectal base in the south is young and mixed. Since the southern regions on the Black Sea coast were incorporated into the Russian Empire after the end of the 18th century, all settlers in this region were Ukrainian or Russian immigrants. Therefore, one cannot speak of autochthonous Ukrainian dialects in the same sense that one can do so for central (or western) regions. Thus, if a "young" Russian-based Surzhyk had arisen during the last three decades of an independent Ukraine due to the political enforcement of the role of the Ukrainian language during this time (cf. Flier, 2008), this should have mainly happened in the south (as well as in the east, which is not at issue in this study). Nevertheless, due to the large number of inhabitants of the region who migrated from various other areas of Ukraine to the south in the 20th and even 21st century, the much older and therefore more stable Ukrainian-based Surzhyk should play a considerable role on the Black Sea coast as well. Furthermore, if the development of a supposed Russian-based Surzhyk had started in the 1990s, at the time when the Ukrainian language was being politically enforced in formerly Russian-speaking areas, then such a mixed speech could not have reached a reliable level of stabilization within 30 years, within one generation. We would not call such a mixed speech Surzhyk or a variety.

4 The Forms of the Infinitive in Ukrainian and Russian

The default ending of the infinitive in Standard Ukrainian today is a syllabic *ty*, in Standard Russian it is a non-syllabic *t'*. In Standard Russian a small number of verbs take different endings, e.g. *ti* (some of them are very frequent like *idti* 'go; directed') and there are some synchronically irregular infinitive forms, where a historical ending has been blended with a final consonant of the stem of the verb, e.g. Russian *moč'* 'can' (some again very frequent). In Ukrainian dialects, generally speaking, according to AUM (1984–2001, maps I/250, II/235, III-1/57-58, III-2/70-71), infinitive endings are sometimes the same as in Standard Ukrainian and sometimes the same as in Standard Russian. In the west and southwest (in the oblasts of Khmelnytskyi and Vinnytsja) dialects almost everywhere show *ty*. This ending, however, can be found in most other oblasts as well, as an alternative to the Russian *t'*. Only in the north (most clearly in the Chernihiv oblast) and in the Dnipropetrovsk oblast can a preponderance or exclusiveness of *t'* be observed. There are some other variants of infinitive endings in Ukrainian dialects, namely *čy* / *či*, *ti* and *t*. These can all be neglected because they are extremely rare in our corpus. There are several thousand instances each for *ty* and *t'*, but less than 100 instances of the others, so they do not play a significant role in Surzhyk and will not be considered in the analysis.

Dialectal influence seemingly plays a role in the occurrence of the *t'* ending in some oral forms of the Ukrainian Standard (cf. UkrPrvp, 2019, p. 153; ZHovtobriukh et al., 1980, p. 220). The latter source furthermore reports the sporadic occurrence of *t'* in literary Ukrainian. Summarising, it may be stated that in the linguistic landscape of Ukraine, *t'* cannot unequivocally be seen as a phenomenon mirroring Russian influence.

For Surzhyk, Del Gaudio (2010, pp. 97–99) reports that the normal ending of the infinitive is *t'*, with possible "deviations" in favour of *ty*. His observations are based on the northern areas of Polesia (northern regions of the oblasts of Zhytomyr, Kyiv, Chernihiv) and Kharkiv (cf. Del Gaudio, 2010, pp. 139–168), all along the Russian border. (His focus on Polesian varieties is somewhat strange because these varieties tend to be viewed as transitional ones towards Belarusian. Generalizations on this basis for Surzhyk are more than problematic.) Apart from Kharkiv, these are areas where the ending *t'* is strongly represented in the corresponding Ukrainian dialects and Del Gaudio is careful not to comment upon the origin of *t'* in his material. Previously, the occurrence of *t'* had been characterized by Flier (1998) as one of many symptoms of the Russian impact on Surzhyk. This again has been criticized by Moser (Mozer, 2016), hinting at the presence of *t'* in

Ukrainian dialects and rejecting its interpretation as a Russicism. A similar view is proposed by Dubichynskyi et al. (2016), presenting *t'* as a Ukrainian dialectism in Surzhyk. Opinions on the origin of *t'* in Surzhyk are thus diametrically opposed.

For Surzhyk, in which Russian influence is evident in various contexts of linguistic structures, the general question is to what extent the distribution and frequency of the two competing endings of the infinitive reflect Russian influence or a Ukrainian dialectal influence. In order to provide an answer to this question, further sub-questions need to be addressed: (i) Is the occurrence of *t'* in Surzhyk indeed all-encompassing? If the answer is positive, then the point of view that *t'* in Surzhyk is a Ukrainian dialectism will have to be rejected, at least for regions where Ukrainian dialects exclusively or overwhelmingly show *ty*. If the answer is negative, the following questions will arise: (ii) Are there any regional differences in the occurrence or frequency of use of one or the other infinitive ending? (iii) Are there factors other than the regional which determine the preference for one of the two endings?

5 Analysis

A total of 9,684 instances of infinitives are available for analysis. Table 1 illustrates the overall frequencies in the corpus showing the following details: (a) The share of material contributed by each of the four partial corpora is not equal: SouthFam contributed the least with 16 percent and CentrInt the most with approximately 36 percent. (b) In total, about one in four infinitives take the ending *ty*, thus three take *t'*. (c) This relation varies in the four partial corpora, but not dramatically: the smallest share of *ty* is observed in SouthFam, where only about 14% of infinitives take *ty*. The largest share of *ty* of almost 29% is found in CentrFam. The figures of the other two partial corpora are closer to the latter.

Table 1. Overall frequencies of infinitives in the corpus.

corpus: / ending:	***ty* in %**	***t'* in %**	**N**	**share of corpus in %**
Centr_Fam	28.6	71.4	2,460	25.4
Centr_Int	25.2	74.8	3,484	36.0
South_Fam	13.8	86.2	1,549	16.0
South_Int	22.5	77.5	2,191	22.6
Total	23.6	76.4	9,684	

Given the background of previous investigations cited above, the first insight is that *ty* is much more widespread in Surzhyk than is suggested by Del Gaudio (2010) on the basis of material from a much smaller area in north-central Ukraine. Thus, the first question posed above has to be answered negatively: *t'* is not the all-encompassing ending in Surzhyk. If, on average, one in four or five instances is *ty* in the centre and the south, it has to be asked whether there are any regularities in this distribution or whether it is chaotic, as many scholars from Ukraine would argue.

The regional differentiation between the centre and the south on the basis of the two corpora is of course too coarse. Hentschel and Taranenko (2021), refining their approach already presented in Hentschel and Taranenko (2015), proposed a model for the strength of the three basic codes in Ukraine, taking the division of the country (to be precise: of the area of investigation) into oblasts as a coordinate system. In other words, they calculated the "weight" of Ukrainian, Russian, and Surzhyk for each oblast. The "weight" should be understood as the frequency of usage stated by the respondents. Their findings are illustrated in Figure 1.

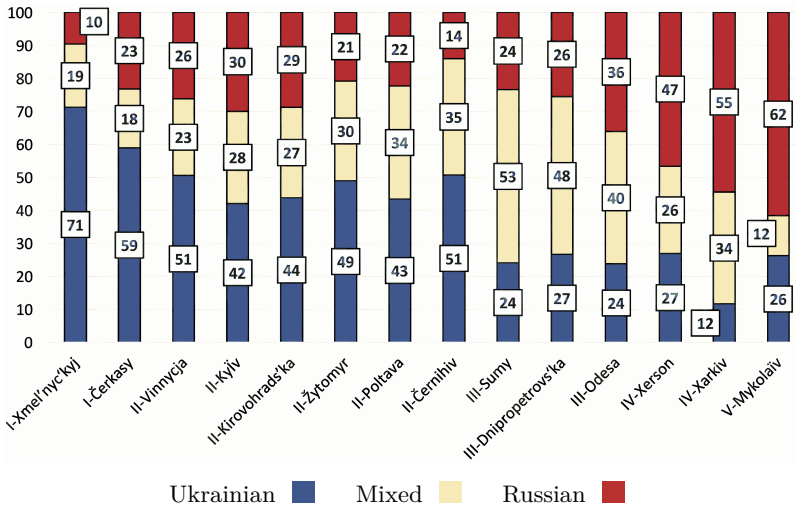


Figure 1. The "weight" of the three codes in clusters and oblasts on a scale from 0 to 100 points (cf. Hentschel & Taranenko, 2021, p. 295).

This model is of twofold interest for this study: Firstly, we propose a map below, illustrating the share of the typically Ukrainian ending *ty* among the infinitives in the material from each oblast. Again, the oblasts serve as a coordinate system. Secondly, if the usage of the competing ending *t'* is primarily grounded in Russian influence on Surzhyk, then this ending should be especially frequently used in oblasts where Russian is the dominant standard language, in spite of the fact that many Ukrainian dialects show *t'* and not *ty*.

The following Map 2 illustrates clear differences in the frequency of usage of the two endings in Ukrainian Surzhyk, at the same time mirroring the clusters of Hentschel and Taranenko (2021) (represented by Roman numerals in front of the name of the oblast) (see Fig. 2).

The first relevant observation is that there are obviously four extreme cases. These are, on the one hand, Khmelnytskyi and Vinnytsia, where *ty* has a share of 90 percent (and correspondingly *t'* a share of 10%), and on the other hand Chernihiv and Zhytomyr, where the share of *ty* is less than 10 percent (and correspondingly *t'* has a share of more than 90%). All four oblasts are located in the northwest of the area of investigation. In fact, Khmelnytskyi, Vinnytsia and Zhytomyr are the three westernmost oblasts. In all four oblasts the weight of Russian is low, between 10 and 26 on the normalized scale from 0 to 100 (cf. Fig. 1). It is not possible to explain the extreme differences in the distribution of the endings by contemporary differences in the frequency of Russian use in everyday life. Moreover, in all four oblasts Ukrainian is, according to Hentschel and Taranenko (2021), very clearly the most frequently used code.

The regional differences in the distribution of frequency of usage must therefore be conditioned by other factors, at least partially. The possible impact of dialects on the usage of the two endings has already been mentioned. The next variables to be controlled are regional differences in the distribution of the endings in traditional, autochthone Ukrainian dialects. Unfortunately, the dialectal distribution of the two endings, competing in Surzhyk, does not form one compact area for each of them (cf. AUM, 1984–2001, maps I/250, II/235, III 1/57-58, III-2/70-71). These dialectal maps only allow a classification of the locations where the material was gathered into five types: (a) all dialects around that location have *ty*: type "only ty", (b) there is a dominance of *ty* over *t'* in the area: type "more ty", (c) both endings are present to a more or less equal extent: type "ty = t'", (d) there is a dominance of *t'* over *ty*: type "more t'", (e) all dialects around that location have *t'*: type "only t'". The variable with these five values (types) will be called "infinitive dialect type", abbreviated as INF-DIAL-TYPE. The gradation behind this classification is of course very



Figure 2. Frequency of usage of the endings *ty* (darker red tone) and *t'* (lighter red tone) in Ukrainian Surzhyk.

rough, for several reasons. The data presented in the AUM are already rather fuzzy. Furthermore, the extent to which the dialectal base in the environment has influenced the mixed vernacular of Surzhyk in all the places where data were recorded remains unknown. Last but not least, we do not know the extent of the dialectal background's influence on the individual respondents' linguistic behaviour. Nevertheless, the hypothesis that the two endings should be frequently used when they are widespread in dialects of a microregion is sound. Table 2 suggests that this is correct, albeit only partially.

Table 2. Share of the ending in areas of different infinitive dialect types (note: the share of *t'* is the difference between the value given for *ty* and 100).

	Share of *ty*	
INF-DIAL-TYPE	**%**	**N**
only ty	83.9	1018
more ty	20.1	870
ty = t'	18.5	1782
more t'	20.4	2721
only t'	11.4	3293
Total	- - -	9684

The dialectal environment clearly has an impact on the frequency of the two endings in Surzhyk in the corresponding locations. However, this impact by no means supports the general interpretation of *t'* as a dialectism from Ukrainian in Surzhyk. The quantitative relations of the "extreme classes" (only ty / only t') seem to be clear. If one or the other ending is omnipresent in the dialects of the corresponding area, then that ending has a frequency of usage of close to 90 percent. Apart

from this, however, the gradation of the presence of the two endings in the dialectal environment is not mirrored by the "middle classes", where both endings are present in the dialectal environment. The frequency relations in the locations of these classes are much closer to the extreme class of only t', with a frequency of about 80 percent for *t'*. This is most puzzling in the area where *ty* dominates. This indicates that the dialectal background does have an influence, but that it is a restricted one.

We obtain a further approximation to answering the question of how far the influence of the dialectal relations extends when we differentiate the dialectal relations according to oblast. The results are shown in Table 3.

Table 3. Share of the material from the five Infinitive dialect types in the oblasts compared with the share of *ty*.

	Share (in %) of the material from the five Infinitive dialect types				**Share of ending (%) in Surzhyk**		**Total**
Oblast	**only ty**	**more ty**	**ty = t'**	**more t'**	**only t'**	**ty**	**N**
II-Chernihiv					100	4	597
II-Zhytomyr			100			9	342
III-Dnipropetrovs'ka			4		96	13	816
IV-Kherson			19	19	62	14	1168
II-Kyïv			2	98		15	514
IV-Kharkiv		80	11	9		17	832
II-Poltava	4	16	80			18	708
III-Sumy				100		18	458
V-Mykolaïv			29	30	41	20	1780
III-Odesa	18			24	58	24	792
II-Kirovohrads'ka		22		78		33	414
I-Cherkasy	44			56		52	734
I-Khmel'nyc'kyj	100					90	236
II-Vinnycja	100					92	293

The interpretation starts from the bottom:

(A) Khmelnytskyi, Vinnytsia: These were the locations with the absolutely highest share (9 out of 10) of the ending *ty* and where all places where data were collected belong to the only ty type. Therefore the assumption that it is this fact that determines the very high share of *ty* in Surzhyk in this area is more than sound.

(B) Cherkasy: This oblast ranks third according to the share of *ty* (about half of the instances). Interestingly, almost half of the set of infinitives analysed comes from only ty areas, with the other half coming from more t' ones. Considering only this oblast, the quantitative relations are once again very clear: In only ty areas, 9 out of 10 infinitives in mixed Surzhyk utterances show *ty*. Data from places in more t' areas on the other hand, show 8 out of 10 infinitives having *t'*. Incidentally, in this oblast there is a relatively clear dialectal isogloss dividing the oblast into the west with *ty* and the east with *t'* (cf. AUM, 1984–2001, map I/250).

(C) Chernihiv, Dnipropetrovsk – Kherson, Kyiv, Sumy – Mykolaiv, Odesa, Kirovohrad: The corpus material for the first two oblasts stems (almost) exclusively from only t' areas, which seems to be in accordance with the fact that in both oblasts more than, or nearly, 90 percent of the infinitive endings in mixed utterances show *t'*. Similarly, in the "middle" three oblasts

(Kherson, Kyiv, Sumy), where more t' areas or more t' areas together with only t' ones provide all or almost all infinitives, more than 80 percent of the infinitive endings are *t'*. The same holds for the three oblasts named last (Mykolaiv, Odesa, Kirovohrad), where the presence of *t'* in dialects in the environs is somewhat reduced and the share of *t'* in Surzhyk is a little less, but in both respects still clearly dominating. One is certainly inclined to formulate a correlation of the kind that if the dialectal base in the environs shows only or predominantly *t'*, then this ending clearly dominates in Surzhyk as well. However, this seems to be only part of the truth.

(D) Kharkiv – Zhytomyr, Poltava: The clearest argument against the general stochastic "rule" that has just been formulated is Kharkiv. It is *ty* that clearly dominates in dialects around the places where data were gathered. Nevertheless, it is *t'* that makes up more than 80 percent of the endings in mixed speech. What has to be underlined here is that in the Kharkiv oblast the Russian language has always been strong (cf. KIIS, 2003, as cited in "IAzyki Ukrainy", n.d.) (cf. Hentschel & Taranenko, 2021 and Fig. 1). The same preponderance of *t'* is to observed in Poltava and even more clearly in Zhytomyr, although dialectally the quantitative relation between the two endings is obviously balanced. The most important point to be made is that the high frequency of *t'* in these three oblasts can by no means be motivated by the influence of geographically close dialects.

The question then arises of whether there are other factors that impact the usage of the competing endings in Surzhyk.

There is one well-known effect on morphological borrowing which should be taken into consideration. Menzel and Hentschel (2017, p. 345) report for four spoken varieties that are heavily exposed to language contact (Surzhyk and Trasjanka being two of them) a clear tendency of "morphologic-etymological agreement" between stem and endings. This would mean in our case that Ukrainian verb stems should show a higher share of the Ukrainian infinitive ending *ty*, whereas Russian stems should show a clearly lower share of that ending and correspondingly a clearly higher share of *t'*. Stems common to Ukrainian and Russian, as well as hybrid stems consisting of at least one Ukrainian and one Russian morpheme, should be somewhere in between.[17] We thus postulate a variable STEM-TYPE. Table 4 illustrates that this is indeed the case.

Table 4. Share of the ending *ty* with stems of different morphological affinity in areas of different infinitive dialect types.

	INF-DIAL-TYPE									
	in percent					**N (absolute)**				
STEM-TYPE	**only ty**	**more ty**	**ty = t'**	**more t'**	**only t'**	**only ty**	**more ty**	**ty = t'**	**more t'**	**only t'**
Ukrainian	88.6	28.5	22.7	29.4	13.0	643	467	1,028	1,383	1,589
Russian	42.5	3.8	2.2	2.8	2.5	80	104	225	507	571
Common	87.8	14.9	18.4	18.0	16.2	237	235	435	674	903
Hybrid	72.4	4.7	12.8	8.3	4.0	58	64	94	157	226
						1,018	870	1,782	2,721	3,289

Table 4 illustrates the quantitative relations when differentiating the five infinitive dialect types on the one hand, and the different stem types on the other hand. In accordance with the

[17] The qualification of morphological elements as Ukrainian, Russian, common or hybrid is not based on judgements of appearances. The procedure for qualification is described in Menzel and Hentschel (2017, pp. 156–163).

hypothesis cited, Russian stems take the ending *ty* to a much lesser extent than Ukrainian ones. The differences in percentage points are shown in Table 5.

Table 5. Differences between Ukrainian and Russian stems in the share of *ty*.

only ty	**more ty**	**ty = t'**	**more t'**	**only t'**
46.1	24.6	20.4	26.6	10.6
percentage points				

Of course, the differences decrease alongside the general decrease in the share of *ty* from only ty areas to only t' ones, yet the impact of the morphologic-etymological affinity to Ukrainian and Russian respectively remains clear. By and large, common and hybrid stems show values of the share that lie between Ukrainian and Russian stems. There are nevertheless some interesting details that do not completely fit into this scheme. With hybrid and especially common stems in only ty areas, the same strong inclination to take *ty* can be observed as with Ukrainian nouns, i.e. all three stem types heavily contrast with Russian stems. In only t' areas (but also in more t' and even more ty ones), hybrid stems show very similar values to Russian stems and correspondingly contrast with Ukrainian stems. It would be a task for a perceptual dialectology analysis (cf. Preston, 1999) to determine whether different tendencies to perceive hybrid lexical stems as Russian or as Ukrainian exist in different regions of Ukraine, depending on aspects of the dialectal or social background. This cannot be done here.

The variables INFINITIVE DIALECT TYPE and STEM-TYPE interact in still another respect. It is well known that the impact of Russian on the mixed Ukrainian Surzhyk and on its Belarusian equivalent Trasjanka is highest in the lexicon.[18] Furthermore, it is well known that Russian, as the main means of linguistic communication, varies in strength in different parts of Ukraine. Therefore, it is necessary to ask whether the lexical impact of Russian varies in different areas of different infinitive dialect types. Of course, the analysis is restricted to the verbs tested for the infinitives. The answer is positive. In only ty areas, the share of Ukrainian verb stems is 63 percent, that of Russian stems 48 percent. The other three infinitive dialect types show figures in between. In only t' areas, the corresponding figures are 7 and 17 percent. The ratio for Ukrainian stems in the former areas to the latter is thus 9 to 1, for Russian stems 3 to 1. This means that the different shares of *ty* and correspondingly *t'* are to some degree dependent on differences in the origin (or morphologic-etymological affinity) of the lexical base. The quantitative data presented so far are simple cross tabulations of relative or absolute frequencies. They definitely offer a solid foundation for initial insights (and hypotheses). However, they consistently ignore the interdependencies between the variables.

For this reason, two multivariate analyses were conducted: an ANOVA analysis and a Generalized Linear Mixed Model (GLMM) analysis. As independent variables we included, of course, all the variables discussed so far, especially (i) INF-DIAL-TYPE and (ii) STEM-TYPE, and furthermore (iii) the weight of the codes in daily communication as calculated by Hentschel and Taranenko (2021 – cf. Figure 1). Furthermore, we include several variables that are usually considered relevant for linguistic behaviour, not only in the bilingual (or "three-code") constellation of Ukraine.

[18] Corresponding corpus-based analyses can be found in Hentschel (2018) for Surzhyk and Hentschel (2013) for Trasjanka. Hentschel (2018) offers evidence that Ukrainian Surzhyk is much less coined by Russian than Belarusian Trasjanka. This is of course due to the fact that Ukrainian has played a much larger role as a means of everyday communication over the last three decades (and even before) than Belarusian, as clearly shown by the studies by Hentschel and Kittel (2011) on the Belarusian situation and by Hentschel and Taranenko (2015, 2021), Hentschel and Zeller (2017) on central Ukraine. Even on the Black Sea Coast, often described as mainly Russian speaking, Ukrainian holds a stronger position than Belarusian in Belarus (cf. Hentschel & Palinska, 2022).

These are: (iv) age; (v) sex; (vi) mother tongue and (vii) first code (Ukrainian, Russian, Surzhyk, or combinations of them – both as stated by the respondents); (viii) education; (ix) size of place of residence (differentiating between "town-like villages", small towns, medium-sized towns, and large towns / cities); (x) where the respondent grew up, i.e. whether a respondent grew up in the countryside, a town, or in both. All these variables represent possible factors that may influence the choice of code and the gradual oscillation between codes in the sense of a certain style shifting (cf. Chambers, 2002; Meyerhoff, 2002) and thus the choice of single means of expression linked to one code or the other.

Both multivariate analyses yielded the same results: the only significant independent variables are indeed infinitive dialect type and stem type.

The ANOVA analysis yielded the following results, cf. Table 6.

Table 6. ANOVA analysis, R^2, effect size.

Variable	**Chi sq**	**Df**	**Pr(> Chi sq)**
SIZE OF PLACE	4.077	3	0.25327
GROWN-UP-IN	4.364	3	0.22343
INF-DIAL-TYPE	198.487	4	< 2e-16 ***
FIRST CODE	2.535	1	0.11135
MOTHER TONGUE	0.549	1	0.45856
SEX	2.369	1	0.12376
EDUCATION	0.814	4	0.93659
AGE	3.612	1	0.05738
STEM-TYPE	236.029	4	< 2e-16 ***
WEIGHT OF CODE	2.290	1	0.130221
Signif. level: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
From the GLMM:			
Marginal R^2 / **Conditional R^2**			0.371 / **0.561**
Effect size: $f^2 = \frac{0.561}{1-0.561} = 1.27$			

Only infinitive dialect type and stem type show significant values, extremely significant ones in fact (cf. the right-hand column). The values for all other variables do not surpass the widely accepted critical value of 0.05. The GLMM yielded the same result. In addition, the conditional R^2 value of 0.561 means that 56% of the variation can be explained by the model presented. The effect size value (f^2) of 1.27 confirms a strong effect of the two significant variables (cf. Cohen, 1992, p. 157).

Two further aspects should be noted: Firstly, the two significant variables show more or less the same Chi-squared value, but STEM-TYPE displays a slightly higher one. This is due to the fact that the clearest differences between the areas of different INF-DIALECT-TYPEs are between the only *ty* type and all four others, and the bulk of the material comes from the latter areas. Secondly, if we repeat the ANOVA analysis for the oblasts with both endings in the dialectal base, which is sound due to the minor differences between them, only the variable STEM-TYPE again shows a highly significant Chi-squared value of 179.822. The corresponding value of the variable INF-DIAL-TYPE (Chi-square = 0.271) is far below significance, which confirms that the differences between them are minor and the variation in the corresponding area is largely determined by the STEM-TYPE.

6 Conclusion

The most general descriptive outcome of the analysis presented is that the ending *t'*, which corresponds to the default ending of infinitives in Standard Russian and is present in many traditional Ukrainian dialectal varieties, is by no means omnipresent in Ukrainian Surzhyk, and conversely, that the ending *ty*, which is the default ending in Standard Ukrainian and in other, mostly western Ukrainian dialects, is by no means a peripheral phenomenon. This contradicts the cited findings of Del Gaudio (2010). His data, however, are from four northern Ukrainian oblasts (often their northern parts), which immediately border with Belarus and Russia, i.e. Russian-speaking territories. It is precisely these oblasts that show the lowest share of *ty*, though even here its share cannot be ignored. As a matter of fact, the map presented above illustrates that the share of *ty* is lowest and thus that of *t'* highest in oblasts bordering Belarus and Russia, or in Ukrainian oblasts that are traditionally seen as Russian speaking (but could not be considered in our investigations). This is of course a clear indication of the long-standing impact of Russian in these regions.

Nevertheless, our analysis does not fully support Flier's (1998) view that infinitives with *t'* are a Russian trait in Surzhyk. Nor does it support the opposite view of Moser (Mozer, 2016) and Dubichynskyi et al. (2016) that *t'* in Surzhyk is a Ukrainian dialectal trait. Our results suggest that both are partially right and partially wrong, because they are too simplistic. The regional distribution of the two endings definitely influences their presence in the corresponding regional variants of Surzhyk. In regions where only *ty* has been fixed in dialects by Ukrainian dialectologists, this ending is by far the most frequent one in Surzhyk as well. However, one has to be aware of the fact that in these regions Ukrainian, not Russian, has always been dominant. Conversely, the same is true for regions where in dialects only *t'* has been recorded by Ukrainian dialectology. However, it is not only these regions where a very clear quantitative dominance of *t'* can be observed in Surzhyk. The most striking examples are Zhytomyr and Kharkiv. In the former oblast, the presence of the two endings in traditional dialects in the regions where data were collected is balanced but the share of *t'* is very high. In the latter oblast, the ending *ty* is even more widespread dialectally but in Surzhyk *t'* clearly dominates. This means that it is unfounded to generally classify *t'* in Surzhyk as a Ukrainian dialectal trait in areas of rural dialects, where it is traditionally not at all or only weakly represented.[19]

The search for other factors that influence the usage of the two endings has yielded just one more variable with a significant connection to the frequency of the two endings in Surzhyk, other than the infinitive dialect type: the morphologic-etymological affinity to Ukrainian or Russian of lexical verb stems. In fact, it has a somewhat stronger impact than the variable infinitive dialect type. For Russian verb stems, *t'* is almost obligatory in all regions except for those where Ukrainian dialects show only *ty*. On the other hand, in the latter regions the very high share of *ty* is reduced by half in Russian verb stems. If we restrict our analysis to the oblasts with both endings in the dialectal base, which is sound due to the minor differences between them, then the result of the quantitative analysis is that only this variable, stem type, is significantly correlated with the frequency of the endings in Surzhyk, in that Ukrainian verb stems are clearly less likely to be combined with *t'* than all other stem types, especially Russian and hybrid ones.

The quantitative model we presented for the frequency of the two endings *ty* and *t'*, in which only the variables infinitive dialect type and stem type play a significant role, explains 56 percent of the overall variation in their distribution. Scholars working without quantitative methods may object that almost half the variation remains unexplained. This is clearly not the case: There is another factor playing a role which cannot directly be measured (if at all, at least not on the basis of our data). This is the long-standing influence of Russian during the last century (at least), especially in the eastern regions of the country. However, there are clear, even quantifiable, indirect indications that this factor plays a role, although it itself cannot directly be quantified.

[19] For a distant spectator, this may seem to be an obvious scenario. The state-of-the-art in Ukrainian linguistics, as has been reported above, is different.

The most general is the fact that the highest share of the typically Russian *t'* is to be found on the northern and eastern periphery, bordering (mainly) Russian-speaking territories. This is observed most convincingly in the oblast of Kharkiv, which for decades has been a stronghold of the Russian language: *t'* dominates clearly in Surzhyk, although in the dialects in the areas considered it is clearly in the minority. There can be no doubt that at these locations the high frequency of *t'* is motivated by decades of Russian influence.

Given the fact that industrialization and the degree of urbanization in Ukraine broadly increases from the west to the east, an analogous decrease in the influence of dialect on people's speech in the same direction is plausible. In places where traditional dialects on the one hand show *t'* predominantly or exclusively, and Russian on the other hand is or used to be strong, the question of whether *t'* is of Russian or dialectal Ukrainian origin ultimately cannot be resolved by quantification. By analogy, however, it is sound to assume that dialectal influence is relatively important in regions with a more rural character, whereas in heavily industrialized and urbanized regions (and there are many in east-central Ukraine) dialectal influence must be considered weaker.

One should furthermore keep in mind the following: If a person (i) has a command of a Ukrainian dialect (or a strong dialectal background in his or her family) that shows a certain ending predominantly or exclusively and (ii) has a command of one or the other standard languages as well, in the majority of cases there is no way to determine whether, in his or her Surzhyk speech, he or she takes this ending in a concrete utterance from the dialect or from the standard language that shares the ending with the corresponding dialect. Of course, there may be cases where, for example, the pronunciation of the mixed utterance is clearly Ukrainian with the exception of the infinitive (with a Russian or common stem), which would be pronounced in a clearly Russian way, as if citing the Russian word form. Such utterances are extremely rare. In this case, the occurrence of the infinitive could definitely be described as an instance of spontaneous insertional code switching, whereas Surzhyk in general can be described as instances of code mixing of the type that Muysken (2000) calls congruent lexicalization. Given this background, the best one can do in such cases of variation is to determine the linguistic and social conditions and their interactions by using modern multivariate statistical analysis, illustrating which linguistic or socio-biographical features increase and decrease the token frequencies of two or more competing, functionally equivalent markers (or forms, words, constructions etc.), and then interpret these quantitative findings qualitatively on the grounds of our general knowledge of the linguistic and social embedding of the phenomenon at issue.

The distribution of *ty* and *t'* in Ukrainian Surzhyk across, between and within regions is definitely not chaotic, as is generally assumed by many scholars of Surzhyk (and Trasjanka), who deny the foreseeability of the use of one or the other of two (or more) competing variants in the mixed varieties and argue that there are no usage norms (e.g. Cychun, 2014; Masenko, 2014; Mechkovskaja, 2014; Mozer, 2016). There are obvious defaults for the usage of both endings, with regional differences. The majority of variation can be explained by only two measurable variables and a third which is not directly measurable but undoubtedly present. Of course, there is a certain amount of variation, even unforeseeable variation. In this regard, several things should be kept in mind: (i) even oral speech in standard varieties shows a certain degree of variation (cf. Lüdtke & Mattheier, 2005); (ii) speech in subvarieties in modern societies, of course, shows more variation than traditional rural dialects do, as long as their societies are relatively immobile (cf. Trudgill, 1986); (iii) if the donor languages (varieties) are actively used in the same mixed or fused variety then phenomena of spontaneous mixing and of stabilized fusion can overlap (cf. Auer, 1999). In this respect, the degree of variation in infinitive endings (in the two different areas) is rather small. The described deviations are the main symptoms of sporadic mixing, although they can mostly be traced back to differences in respondents' individual biographies.

The presented findings lend support to our proposal to view phenomena such as Surzhyk as mesolects in the sense of fusions of autochthone rural dialectal varieties and standard languages. In mesolects of this type, the amount of dialectal variation (between local dialects) is reduced.

Such varieties lose many autochthone dialectal traits (in our case some other endings of the infinitive, already poorly represented in few local dialects), but under certain social conditions preserve others. The latter may undergo a redistribution, functionally or, as in our case, regionally simplified under the influence of a standard language. The specifics of Ukraine consist in the fact that there are two standard languages at hand, of which one (Ukrainian) is traditionally more strongly represented in the western parts of the country and the other (Russian) further eastwards. In recent decades, due to the political promotion of the Ukrainian Standard language by the government of the country, there seems to be a tendency towards a real strengthening of Ukrainian in the population in an eastward direction (cf. Hentschel & Taranenko, 2021).

Nevertheless, regional differences remain. Proposing a regionally diversified Surzhyk does not contradict Flier's (2008) point that there is only one (Ukrainian-based) Surzhyk, thus strongly contradicting Bilaniuk's (2004) classification of several socially or socio-biographically conditioned variants of Surzhyk. Our approach is rather a synthesis of sociolinguistic and dialectological approaches complemented by the recognition of the role of traditional Ukrainian dialects in Surzhyk. The latter point has been neglected in the (often emotional and politically-coloured) discussion of this variety. Surzhyk arose in the course of rural-urban migration connected with industrialization. This means that we simply must expect a certain degree of regional variation in one (Ukrainian-based) Surzhyk, due to the impact of dialectal substrata. These areal contrasts are partially interwoven with social and political characteristics. The latter include the different roles of Standard Ukrainian and Russian in different areas and social groupings. One should, however, assume only one Surzhyk (on a Ukrainian base) with differentiations along a territorial continuum mirroring traditional dialectal diversity to a certain degree, as well as the different impact of the two standard languages widely present across Ukraine with varying strength. In other words, Surzhyk mirrors in an intertwined way both dialect levelling and (post)colonial hybridization.[20]

Students of Surzhyk (and Trasjanka), mostly explicitly or (more often) implicitly working with theoretical concepts from structuralism, tend to ask questions about the "system" and, as a rule, deny any systematicity in Surzhyk. Is Surzhyk a system, or can it be understood as a system when we conceive of it as mesolects? When trying to provide an answer to such a question, one must bear in mind that in linguistics the notion of a "system" of a language has a twofold meaning. Firstly, the notion stands for an encompassing abstraction that linguists construct from observable linguistic phenomena of regular character. Secondly, the notion refers to an internalized, cognitively real system in Chomskyan terms, in the sense of the competence of an idealized hearer-speaker: that what speakers can speak.

Given this background, what does it mean when we propose (as above) to regard Surzhyk as mesolects? As the plural number in "mesolects" indicates, we propose to assume that there is more than one "Surzhyk-type" mesolect in the territory of Ukraine, in the same sense as there are several traditionally assumed dialects. This leads us to the next question: Can one speak one such dialect? The answer is clear: One definitely cannot. Traditionally assumed dialects, whether in Ukraine, Poland or Germany, stand for abstractions based on specific, selected regional linguistic traits, To be precise, these abstractions are often enough pre-shaped by historical tribal or territorial-political categories (ancient principalities, kingdoms etc.). For Polish, this has been outlined by Dejna (1998) and for German by Löffler (1982). The basic systems that linguists may describe and speakers are competent to use and realize (cognitively real), are those with much smaller territorial scope (Pol. *gwara*, Ger. *Mundart*, Ukr. *hovirka*), for example villages or groups of settlements comprising an area of close linguistic intercourse. These regionally minor variants may then be subsumed under the larger units of traditionally acknowledged dialects on the basis of linguistic similarities and differences. Endeavours to model territorial regional variation within one language (or larger "dialects" within a language) in terms of diasystems, which trace back to Weinreich (1954), have fallen short (Auer & di Luzio, 1988; Chambers & Trudgill, 1998, pp.

[20] Cf. Britain (2002) who argues in favour of considering the interplay of social and spatial (regional) variation in variationist sociolinguistics.

32–44). Instead, one conceives dialectal diversity as a dialect continuum, i.e., as a continuum of different basic systems, whereby neighbouring systems overlap to a large degree in their linguistic characteristics, distant systems to a (much) lower degree. Analogously, proposing that Surzhyk be regarded as several mesolects is tantamount to assuming a mesolectal continuum named Surzhyk: Surzhyk with different regional variants.

Instead of looking for an abstract diasystem to model regional variation in the mesolectal continuum of Surzhyk, one should rather concentrate on investigating corresponding linguistic variables (as we did with the form of the infinitive[21]), and fix in which space one of two (or more) variants dominates (the regional differences outlined above), where there are tendencies towards positional variation (the differences regarding Ukrainian or Russian infinitive stems described above) and where there is free variation between variants. Clear quantitative differences in the usage of one or the other variant indicate regularization and thus structure building, not just quantitative differences of casual, spontaneous code-switching or -mixing.

References

Auer, P. (1999). From codeswitching via language mixing to fused lects: Towards a dynamic typology of bilingual speech. *International Journal of Bilingualism*, *3*(4), 309–332. https://doi.org/10.1177/13670069990030040101

Auer, P., & di Luzio, A. (1988). Introduction: Variation and convergence as a topic in dialectology and sociolinguistics. In P. Auer & A. di Luzio (Eds.), *Variation and convergence: Studies in social dialectology* (pp. 1–10). De Gruyter. https://doi.org/10.1515/9783110851601.1

Bilaniuk, L. (2004). A typology of surzhyk: Mixed Ukrainian–Russian language. *International Journal of Bilingualism*, *8*(4), 409–425. https://doi.org/10.1177/13670069040080040101

Brandes, O. (2015). *Die weißrussisch-russische gemischte Rede als Subvarietät im Vergleich zu russischen nicht-hochsprachlichen Varietäten und Weißrussisch: Phänomene der Flexionsmorphologie* [Doctoral dissertation, Universität Oldenburg]. ORBIS. http://d-nb.info/1073247368/34

Britain, D. (2002). Patterns of variation including change. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 603–637). Blackwell.

Chambers, J. K. (2002). Patterns of variation including change. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 349–372). Blackwell.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology: Cambridge textbooks in linguistics* (2nd ed.). Cambridge University Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cychun, H. (2014). Soziolinguistische, soziokulturelle und psychologische Grundlagen gemischten Sprechens. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 163–172). Peter Lang.

Danylenko, A., & Naienko, H. (2019). Linguistic russification in Russian Ukraine: Languages, imperial models, and policies. *Russian Linguistics*, *43*, 19–39. https://doi.org/10.1007/s11185-018-09207-1

Dejna, K. (1998). Gwara i jej stosunek do innych odmian języka ogólnonarodowego. In S. Gala (Ed.), *Teoretyczne, badawcze i dydaktyczne założenia dialektologii* (pp. 13–22). Łódzkie Towarzystwo Naukowe.

Del Gaudio, S. (2010). *On the nature of Surzhyk: A double perspective*. Sagner.

Dubichynskyi, V., Reuther, T., & Doleschal, U. (2016). Surzhyk: Leksychnyĭ aspekt (na materiali teleperedachi "Kliuchovyĭ moment", leksykohrafichnykh dzherel XIX–XXI st. i rezul′tativ anketuvannia). *Linguistica Copernicana*, *2016*(13), 335–363. http://doi.org/10.12775/LinCop.2016.017

Flier, M. (1998). Surzhyk: The rules of engagement. *Harvard Ukrainian Studies*, *22*, 113–136.

Flier, M. (2008). Surzhyk or Surzhyks? In G. Hentschel & S. Zaprudski (Eds.), *Belarusian Trasjanka and Ukrainian Surzhyk* (pp. 39–56). BIS-Verlag.

Hentschel, G. (2008). Zur weißrussisch-russischen Hybridität in der weißrussischen "Trasjanka". In P. Kosta & D. Weiss (Eds.), *Slavistische Linguistik 2006/2007: Referate des XXXII. Konstanzer Slavistischen*

[21] And as Palinska and Hentschel (Palins′ka & Khentshel′, 2022) did for the prefixes and prepositions *VID* / *OT*.

Arbeitstreffens, Männedorf bei Zürich, 18.–20. September 2006 und Referate des XXXIII. Konstanzer Slavistischen Arbeitstreffens, Potsdam, 4.–6. September 2007 (pp. 169–219). Sagner.

Hentschel, G. (2013). Zwischen Variabilität und Regularität, "Chaos" und Usus: Zu Lautung und Lexik der weißrussisch-russischen gemischten Rede. In G. Hentschel (Ed.), *Variation und Stabilität in Kontaktvarietäten: Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien* (pp. 63–99). BIS-Verlag.

Hentschel, G. (2014). "Trasjanka" und "Surzhyk" – zum Mischen von Sprachen in Weißrussland und der Ukraine: Einführung in die Thematik und Ausblick auf den Band. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 1–26). Peter Lang.

Hentschel, G. (2017). Eleven questions and answers about Belarusian-Russian Mixed Speech ("Trasyanka"). *Russian Linguistics*, *41*(1), 17–42. https://doi.org/10.1007/s11185-016-9175-8

Hentschel, G. (2018). Die weißrussische "Trasjanka" und der ukrainische "Surzhyk": Grundlegende quantitativ-qualitative soziolinguistische Ähnlichkeiten und Unterschiede. In S. Kempgen, M. Wingender, & L. Udolph (Eds.), *Deutsche Beiträge zum 16. Internationalen Slavistenkongress Belgrad 2018* (pp. 127–138). Harrassowitz Verlag. https://doi.org/10.2307/j.ctv11sn3zx.14

Hentschel, G., & Brüggemann, M. (2015). Gibt es in der Ukraine einen ukrainisch-russischen Sprachenkonflikt? Über das Miteinander, Gegeneinander und Durcheinander von Ukrainisch und Russisch. In E. Bakalova, T. Endrich, K. Shlyakhtovska, & G. Spodarets (Eds.), *Ukraine. Krisen. Perspektiven: Interdisziplinäre Betrachtungen eines Landes im Umbruch* (pp. 95–118). Wissenschaftlicher Verlag Berlin.

Hentschel, G., & Kittel, B. (2011). IAzykovaia situatsiia v Belarusi: Mneniia belorusov o rasprostranennosti iazykov v strane. *Sotsiologiia*, *2011*(4), 62–79.

Hentschel, G., & Palinska, O. (2022). The linguistic situation on the Ukrainian Black Sea coast – Ukrainian, Russian and Suržyk as "native language", "primary code", frequently used codes and codes of linguistic socialization during childhood. *Russian Linguistics*, 46(3), 259–290. https://doi.org/10.1007/s11185-022-09259-4

Hentschel, G., & Taranenko, O. (2015). Die Sprachenlandschaft der zentralen Ukraine: Ukrainisch, Russisch, 'Surzhyk': Verwendung – Kompetenz – nationale Positionierung. *Die Welt der Slaven*, *60*(2), 248–275.

Hentschel, G., & Taranenko, O. (2021). Bilingualism or tricodalism: Ukrainian, Russian and "Surzhyk" in Ukraine: Analysis and linguistic-geographical mapping. *Die Welt der Slaven*, *66*(2), 268–299.

Hentschel, G., & Zeller, J. P. (2016). Meinungen und Einstellungen zu Sprachen und Kodes in zentralen Regionen der Ukraine. *Zeitschrift für Slawistik*, *61*(4), 636–661. https://doi.org/10.1515/slaw-2016-0039

Hentschel, G., & Zeller, J. P. (2017). Aspekte der Sprachverwendung in zentralen Regionen der Ukraine. *Wiener Slavistischer Almanach*, *79*(1), 37–60.

Hentschel, G., Zeller, J. P., & Tesch, S. (2014). *Das Oldenburger Korpus zur weißrussisch-russischen gemischten Rede: OK-WRGR*. http://www.uni-oldenburg.de/ok-wrgr

Hrytsenko, P. (Ed.). (2015). *Dialekty v synkhroniï ta diakhroniï: Tekst iak dzherelo linhvistychnykh studiĭ*. KMM.

IAzyki Ukrainy. (n.d.). In *Wikipedia*. https://ru.wikipedia.org/wiki/%D0%AF%D0%B7%D1%8B%D0%BA%D0%B8_%D0%A3%D0%BA%D1%80%D0%B0%D0%B8%D0%BD%D1%8B

Jaspers, J. (2017). Diglossia and beyond. In O. Garcia, N. Flores, & M. Spotti (Eds.), *The Oxford handbook of language and society* (pp. 179–198). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190212896.013.27

Kamusella, T. (2018). Russian: A monocentric or pluricentric language? *Colloquia Humanistica*, *2018*(7), 153–196. https://doi.org/10.11649/ch.2018.010

Kyïvs′kyĭ mizhnarodnyĭ instytut sotsiolohiï. (2019). *Dumky i pohliady naselennia shchodo vykladannia rosiĭs′koï movy v ukraïns′komovnykh shkolakh i nadannia nepidkontrol′nym terytoriiam Donbasu avtonomiï u skladi Ukraïny: Berezen′ 2019 roku*. https://www.kiis.com.ua/?lang=ukr&cat=reports&id=832&page=2

Levchuk, P. (2020). *Trójjęzyczność ukraińsko-rosyjsko-polska Ukraińców niepolskiego pochodzenia*. Księgarnia Akademicka. https://doi.org/10.12797/9788381382854

Löffler, H. (1982). Gegenstandskonstitution in der Dialektologie: Sprache und ihre Differenzierungen. In W. Besch, U. Knoop, W. Putschke, & H. E. Wiegand (Eds.), *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung* (Vol. 1.1, pp. 441–463). De Gruyter.

Lüdtke, J., & Mattheier, K. J. (2005). Variation – Varietäten – Standardsprachen: Wege für die Forschung. In A. N. Lenz & K. J. Mattheier (Eds.), *Varietäten – Theorie und Empirie* (pp. 13–38). Peter Lang.

Masenko, L. (2014). Surzhyk im System umgangssprachlicher Formen des Ukrainischen. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 327–342). Peter Lang.

Matviias, I., Zakrevs′ka, IA., & Zales′kyĭ, A. (Eds.). (1984–2001). *Atlas ukraïns′koï movy* [AUM]. Naukova dumka.

Maznichenko, IE., Makedon, V., Sharabanova, S., & IAlovnycha, I. (Eds.). (2019). *Ukraïns′kyĭ pravopys* [UkrPrvp]. Naukova dumka.

Mechkovskaja, N. (2014). Die weißrussische Trasjanka und der ukrainische Surzhyk: Quasi-ethnische, russifizierte Substandards in der Geschichte der sprachlichen Situation. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 53–89). Peter Lang.

Menzel, T., & Hentschel, G. (2017). *Flexionsmorphologische Irregularität im innerslavischen Sprachkontakt: Sprachinhärente Präferenzen oder politisch-soziale Dominanz: Russisch vs. Weißrussisch / Ukrainisch – Polnisch vs. Kaschubisch / Lemkisch.* Peter Lang.

Meyerhoff, M. (2002). Communities of practice. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 526–548). Blackwell.

Mozer, M. (2016). "Surzhyk" chy "surzhyky"? *Ukraïns′ka mova*, *2016*(1), 27–54.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing.* Cambridge University Press.

Palins′ka, O., & Khentshel′, G. (2022). Regional′nye osobennosti ispol′zovaniia ukrainsko-russkoĭ smeshannoĭ rechi (surzhika) i vliianie dialektov: Pristavki i predlogi *VID* / *OT*. *LingVaria*, *17*(2(34)), 229–253. https://doi.org/10.12797/LV.17.2022.34.15

Preston, D. R. (Ed.). (1999). *Handbook of perceptual dialectology* (Vol. 1). John Benjamins. https://doi.org/10.1075/z.hpd1

Schmidt, J. E., & Herrgen, J. (2011). *Sprachdynamik: Eine Einführung in die Regionalsprachenforschung.* Erich Schmidt.

Stavyc′ka, L. (2014). Ein Blutschandekind der Postmoderne. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 343–370). Peter Lang.

Taranenko, O. (2004). Surzhyk. In V. Rusanivs′kyĭ & O. Taranenko (Eds.), *Ukraïns′ka mova: Entsyklopediia* (pp. 665–668). Ukraïns′ka entsyklopediia.

Taranenko, O. (2007). Ukrainian and Russian in contact: Attraction and estrangement. *International Journal of the Sociology of Language*, *2007*(183), 119–140. https://doi.org/10.1515/IJSL.2007.007

Taranenko, O. (2013). Variantnist′ vs. stabil′nist′ u strukturi ukraïns′ko-rosiĭs′koho "surzhyku" (URS): Sukupnist′ idiolektiv vs. sotsiolekt. In G. Hentschel (Ed.), *Variation und Stabilität in Kontaktvarietäten: Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien* (pp. 27–61). De Gruyter.

Taranenko, O. (2014). Ukrainisch-russischer Surzhyk: Status, Tendenzen, Bewertungen, Prognosen. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Surzhyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 265–287). Peter Lang.

Trudgill, P. (1986). *Dialects in contact.* Blackwell.

Vseukraïns′kyĭ perepys naselennia. (2001). http://ukrcensus.gov.ua

Weinreich, U. (1954). Is a structural dialectology possible? *Word*, *10*(2–3), 388–400. https://doi.org/10.1080/00437956.1954.11659535

Woolhiser, C. (2011). "Belarusian Russian": Sociolinguistic status and discursive representations. In R. Muhr (Ed.), *Non-dominating varieties of pluricentric languages: Getting the picture* (pp. 11–49). Peter Lang.

Zaprudski, S. (2007). In the grip of replacive bilingualism: The Belarusian language in contact with Russian. *International Journal of the Sociology of Language*, *2007*(183), 97–118. https://doi.org/10.1515/IJSL.2007.006

Zeller, J. P., Taranenko, O., & Hentschel, G. (2019). Language and religion in Central Ukraine. *International Journal of the Sociology of Language*, *2019*(260), 105–130. https://doi.org/10.1515/ijsl-2019-2050

ZHovtobriukh, M., Volokh, O., Samiĭlenko, S., & Slyn′ko, I. (1980). *Istorychna hramatyka ukraïns′koï movy: Navchal′nyĭ posibnyk*. Vyshcha shkola.

Acknowledgment
We are grateful to Ekaterina Danilava and István Fekete for advice on the statistical analysis and to Katherine Bird for help with the English text. Remaining errors are ours.

The publication was financed by the funds of the projects supported by the Fritz-Thyssen-Foundation (no. 10.14.1.066) and Deutsche Forschungsgemeinschaft (no. 155014374 and no. 419468937).

The authors declare that they have no competing interests.

Both the authors participated equally in preparing conception and academic editing of this article.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (http://creativecommons.org/licenses/by/3.0/pl/), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Authors 2022

Publisher: Institute of Slavic Studies, Polish Academy of Sciences

Publishing history: Received 2022-03-22; Accepted 2022-06-20; Published 2022-12-28.