Sonderdruck aus/Offprint from

DIE WELT DER SLAVEN

Internationale Halbjahresschrift
für Slavistik

Herausgegeben von
Daniel Bunčić (Köln), Susanne Frank (Berlin),
Schamma Schahadat (Tübingen), Monika Wingender (Gießen)

Jahrgang 66 (2021), Heft 2

Harrassowitz Verlag · Wiesbaden

Einsendung von Artikeln (bitte Typoskriptregeln und Dokumentvorlage anfordern oder von http://weltderslaven.de/ herunterladen) bzw. Rezensionsanfragen an: redaktion@weltderslaven.de

Eine Verpflichtung zur Besprechung oder Rücksendung zugesandter Bücher kann nicht übernommen werden. Rezensionen nur nach Rücksprache.

Inhaltsverzeichnisse aller Jahrgänge unter:
https://weltderslaven.de/

Die Zeitschrift wurde 1956 im Harrassowitz-Verlag gegründet und erscheint seit Band 61 (2016) wieder bei Harrassowitz.
Von Band XVII (1972) bis XXI (1976) erschien sie im Böhlau-Verlag (Köln/Wien) und von Band XXII (1977) bis LX (2015) im Verlag Otto Sagner (München).

Alle Artikel haben einen anonymen Peer-Review durchlaufen.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://www.dnb.de/ abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at https://www.dnb.de/.

Gesetzt aus der Campus Garamond von MacCampus und der Hypatia von Adobe
Druck und Verarbeitung: Hubert & Co., Göttingen
Gedruckt auf alterungsbeständigem Papier
Printed in Germany

https://www.harrassowitz-verlag.de/

ISSN 0043-2520

DOI Zeitschriftenreihe 10.13173/0043-2520
DOI Titel 10.13173/WS.66.2

I. Artikel

II. Rezensionen

Bilingualism or tricodalism: Ukrainian, Russian and "Suržyk" in Ukraine

Analysis and linguistic-geographical mapping

Gerd Hentschel, Oleksandr Taranenko

The study aims to show that the different usage frequencies of Ukrainian, Russian and Suržyk in everyday life in Ukraine can be described and mapped as a continuum of "multicodalism". This sets it apart from previous representations that usually assume a historical-political demarcation of sub-areas and consequently tend to indicate a linguistic bisection of the country into a Ukrainian-speaking area encompassing the west and the central region, and a Russian-speaking area in the east and south. The approach proposed here employs empirical data on usage of the three codes as the basis for its structuration. Starting from the mean frequencies of code usage in individual oblasts, a cluster analysis groups these into larger areas and clarifies their internal variation. This result is not just a more continuous picture of the linguistic landscape in Ukraine, but also a more coherent one.

Keywords: multilingualism, linguistic cartography, language contact, Ukrainian and Russian studies

Hentschel, Gerd, University of Oldenburg, Slavic department, gerd.hentschel@uol.de.

Taranenko, Oleksandr, O. O. Potebnja Institute of Linguistics, National Academy of Sciences of Ukraine, Kyïv, oataran99@gmail.com

This paper is based on two grants: (1) *Variability and stability in a mixed substandard in extensive and time-stable language contact: The Ukrainian-Russian Suržyk in Ukraine*, funded by the Fritz-Thyssen-Foundation (no. 10.14.1.066), (2) *Hybridization from two sides: Ukrainian-Russian and Russian-Ukrainian code-mixing in the context of the sociolinguistic situation in Southern Ukraine along the Black Sea coast*, funded in an Austrian-German cooperation by the *FWF – Der Wissenschaftsfonds* (no. I 4189-G30) and the *Deutsche Forschungsgemeinschaft* (DFG, no. 419468937).

Die Welt der Slaven 66 (2021) 2, 268–299 DOI: 10.13173/WS.66.2.268

1. Introduction: A mixed subvariety as a Ukrainian legacy of Russian colonialism[1]

Leaving aside a multitude of minority languages that are not relevant to this study, Ukraine is considered bilingual with Ukrainian and Russian. Furthermore, a form

[1] This paper is the product of two projects (a) *Variability and stability in a mixed substandard in extensive and time-stable language contact: The Ukrainian-Russian Suržyk in Ukraine*, funded by the Fritz-Thyssen-Foundation (no. 10.14.1.066), headed by Gerd Hentschel, Oldenburg; (b) *Hybridization from two sides: Ukrainian-Russian and Russian-Ukrainian*

of mixed Ukrainian-Russian speech has emerged from the centuries long contact between the two related and structurally very similar East Slavic languages Ukrainian and Russian. This mixed speech is generally established in the country under the name "Suržyk" and has moved away from being a colloquial language and into academic and cultural-political discourse. The term was originally used to denote a low-grade mixture of different grains or flour (cf. Taranenko 2007).[2] This already touches upon the negative connotations of Suržyk that have grown into outright stigmatisation among some nationally-oriented elites in Ukraine and to a certain extent this negative view is transferred to speakers of Suržyk (cf. Stavyck'ka 2014). On the whole, the majority of the population has a more relaxed attitude. Although many people have a sceptical to deprecatory view on Suržyk, the number of moderate if not decided friends is just as high. Attitudes are particularly positive in the central areas of the country (Hentschel & Zeller 2016, 652–657). Bilaniuk (2018) also identifies a tendency towards a less negative evaluation of the mixed speech in Ukraine.

Spontaneous forms of speech mixing involving Ukrainian and Russian can be considered to have started in the second half of the seventeenth century when Ukraine east of the Dnipro, including today's capital Kyïv, fell under the rule of Moscow. From that point onwards Russian was the language of power in the territory of Ukraine and dominated both politically and socially until the end of the Soviet Union (cf. Danylenko & Naienko 2019). The other language of power in the area of interest that had been established over the last two centuries, Polish, was pushed into Ukraine's western extremities during the eighteenth and nineteenth centuries, and completely ousted from Ukraine (apart from among the Polish minority) after the Second World War. In spite of the much older dominance of Russian, it was only during the phase of industrialisation in the nineteenth century, when apart from the westernmost regions around L'viv the rest of Ukraine was under Russian rule, that the socio-political preconditions emerged for a certain stabilisation of the mixed Ukrainian-Russian speech. This process accelerated during the period of Soviet industrialisation, which was accompanied

code-mixing in the context of the sociolinguistic situation in Southern Ukraine along the Black Sea coast, headed by Tilmann Reuther (Klagenfurt) and Gerd Hentschel (Oldenburg), which is being sponsored in an Austrian-German cooperation by the *FWF – Der Wissenschaftsfonds* (no. I 4189-G30) and the *Deutsche Forschungsgemeinschaft* (DFG, no. 419468937) in cooperation with Oleksandr Taranenko (Kyïv) and Tatjana Kuznecova (Odesa). We are grateful to Olesya Palinska (Oldenburg) for various comments on a previous version of the paper and to Katherine Bird (Berlin) for the translation. Remaining errors are ours. Maps have been created by Thomas Robbers (Oldenburg).

[2] There is a similar situation in neighbouring Belarus (cf. Hentschel 2017; 2018) where the mixed speech is called "Trasjanka". This will be referred to below for purposes of comparison.

by increasing urbanisation, leading to mixed Ukrainian-Russian speech becoming a mixed variety in a narrow sense. Two factors played a decisive role. First, the rural population, who mainly spoke local Ukrainian dialects, moved permanently to the towns where Russian dominated, not least because it was spoken by many people in socially and politically prominent leadership positions. The rural-urban migrants adapted to the language norms of their urban environment, facilitated by, on the one hand, the genetic and structural closeness of both languages, and on the other hand, the lack of necessity for complete adoption: even if adoption was only partial, understanding was effortless (cf. Taranenko 2014: 274–276). Second, the rural-urban migrants practised mixed speech with their children and grandchildren which promoted its stabilisation (cf. Hentschel 2017: 19–21). Initially this meant that "classic" Suržyk was a variety with a Ukrainian base (substrate) and a Russian superstrate, while the latter dominated during the Soviet era in its standard form. The new Ukrainian standard language, which did not revive the old Ukrainian written tradition lost during the eighteenth century, first emerged during the nineteenth century but only received a measure of state support in 1917/18 and during the 1920s. For the entire Soviet period that followed, Ukrainian was only peripheral, e.g. to be found in the education sector. Ukrainian can be considered an adstrate of Suržyk.

Suržyk is of course a highly variable code. It is well known that in mixed speech, or more generally in mixed codes, stabilising mixing (fusion) and spontaneous mixing overlap in the speech for as long as the donor codes are in use in society. For Ukrainian and Russian in Ukraine this remains the case up until today.[3] However, many representatives of the Ukrainian cultural elite with a nationalist orientation consider Suržyk to be the legacy of Russian colonial rule, as an "incestuous child of bilingualism" as the Ukrainian writer Ju. Andruxovyč called it without making this view his own (cf. Stavyc'ka 2014).

After Ukraine became independent following the break-up of the Soviet Union in 1991, the emergence of a somewhat different form of mixed, now Russian-Ukrainian speech can be observed. Newly-independent Ukraine strongly promoted Ukrainian in society, education and the media (cf. for example Taranenko 2014: 271). This meant that speakers who prior to 1991 had exclusively and preferentially used Russian now often had to turn to Ukrainian, at least to some degree or in certain types of communication. The speech of these people is sometimes now mixed Russian-Ukrainian. However, not least because of its recent origin (over the course of the last almost 30 years) it is not yet possible to expect a stabilisation of such a "Neo-Suržyk" on a Russian basis.

The following study will illuminate how the three codes Ukrainian, Russian and Suržyk are distributed across different regions of the country as a means of

[3] On the stabilisation of linguistic structures cf. Taranenko (2013; 2014), Menzel & Hentschel (2017), Hentschel (2018).

communication in everyday life.[4] It remains beyond question that Ukrainian-Russian Suržyk, even in today's more or less stabilised form, does, of course, not share the same "status" as the other two recognised languages. It primarily functions as a distinct subvariety of speech.[5] Taking a structural-quantitative view (token and type frequencies), "old" Suržyk as practised in the central region of the country is clearly more similar to Ukrainian than Russian in all respects, in contrast to Trasjanka, even in lexis (cf. Menzel & Hentschel 2017, Zeller 2018, Hentschel 2018). If it were not for the previously cited nationalist-political reservations, one could refer to Suržyk in this regard as a socially produced variety, a social dialect of Ukrainian, in spite of all its Russian traits.

The following analysis focuses on the multicodalism of the population by including Suržyk and taking, as in Hentschel & Taranenko (2015), a quantitative sociolinguistic[6] approach as a prelude to subsequent qualitative analyses. As such and in common with all studies of this kind, a certain degree of relativity is characteristic. First, the frequency of usage of the three codes is determined by respondents' self-assessments and second, the boundaries between Suržyk and the other donor languages are based on respondents' subjective assessments. Although less obvious, surveys that do not include Suržyk also encounter a similar problem to the latter (see below and Kulyk 2010: 392).

2. Previous representations of the distribution of Ukrainian, Russian and Suržyk

One of the more differentiated representations in recent years is the study by the Kyiv International Institute of Sociology (KIIS[7] 2003). This is, as it seems, the only

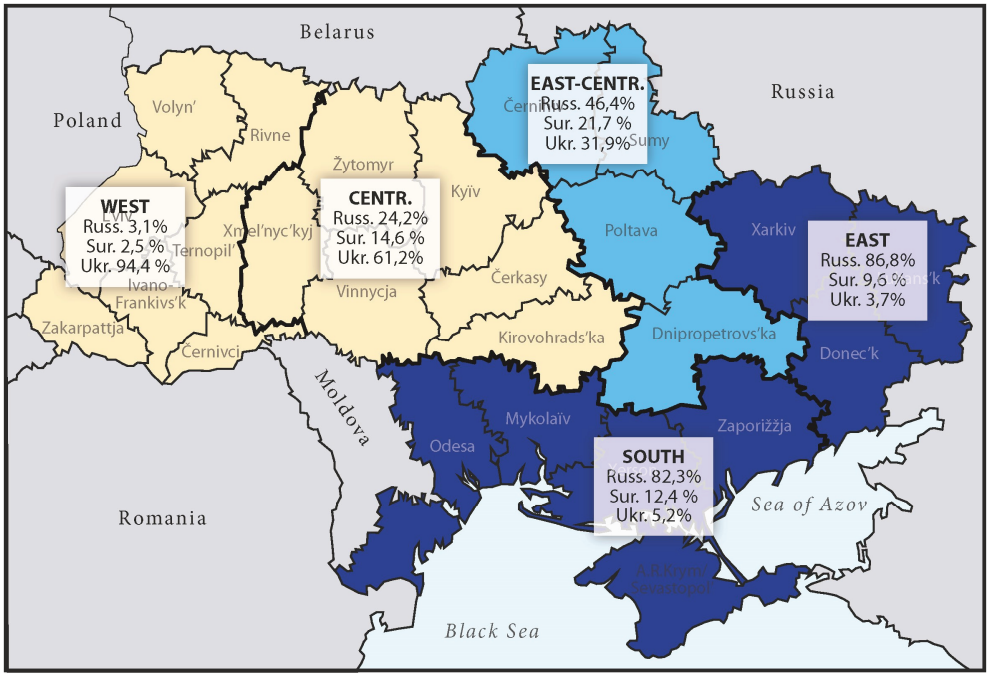
[4] The aim here is not yet to differentiate between traditional Suržyk and a presumed "Neo-Suržyk". (In any case, the latter would be quantitively less important.) The project currently being conducted that provided the data for this study (cf. Section 3) is also investigating the question of whether the presumed Russian-based "Neo-Suržyk" as addressed e.g. by Flier (2000) should generally be positioned on the same level as traditional Suržyk. The hypothesis is that the former, as a much younger phenomenon, is much more an instance of spontaneous mixing than the latter and that it is sometimes coloured more strongly by Ukrainian and sometimes more strongly by Russian with nuanced distinctions rather than a clear boundary. Of course, this also depends on individual speakers' basic orientation towards Ukrainian and Russian as well as situational mechanism of "style-shifting" due to the strong structural similarities of the two languages.

[5] An informal written form of Suržyk can be found in the internet (cf. Kuroxtina 2009, Ševčuk-Kljuževa 2016). Bilaniuk (2018) refers to a further instance of Suržyk online.

[6] Note that we use the term "sociolinguistic" as an adjective corresponding rather to the sociology of language. In this paper we do not refer to linguistic elements or construction, which would be sociolinguistic in a narrow sense: the influence of social phenomena on the use of linguistic units.

[7] Cf. the web site: https://www.kiis.com.ua/?lang=eng.

study in Ukraine that in addition to Ukrainian and Russian also considered Suržyk as a third code. However, this phone survey can be viewed critically for our purposes because the respondents could answer in Ukrainian or Russian, but whether they missed the "goal" of Ukrainian or Russian and presented mixed speech, i.e. Suržyk, was decided by sociologists (not linguists). A decisive disadvantage of this method is that many people in Ukraine are bi- if not tricodal, which means that depending on the communicative context they can switch between at least one of the two recognised languages and Suržyk (cf. Hentschel & Taranenko 2015, Hentschel & Zeller 2017). Consequently, the KIIS (2003) survey is likely to only record those speakers of Suržyk whose linguistic repertoire is limited to this one code.



Map 1: The distribution of the three codes according to principal usage (according to KIIS 2003)

Similar to many other "linguistic maps" of Ukraine, Map 1 illustrates an underlying areal bisection: on the one side the east (in the broad sense), subdivided into the eastern centre and the east (in the narrow sense), with Russian as the dominant language employed[8] and, on the other side, the west (in the broad sense), with the

[8] We do not concern ourselves with so-called native languages as in many other surveys. When questioned about native language, the large majority of answers are of symbolic significance with a more or less loose coupling to actual usage. In countries like Ukraine and Belarus, asking about the native language consistently yields different results to asking

western centre[9] and the west (in the narrow sense). It is not difficult to recognise a historical-political division in this linguistic divide: the east (in the broad sense) substantially corresponds to the region of Ukraine that came under the rule of Moscow in the second third of the seventeenth century, and the south corresponds to the region that since the end of the eighteenth century was a stable part of the Tsarist empire, although there had not been any East Slavic settlements (neither Ukrainian nor Russian) there previously.[10] A further disadvantage of this representation is that the two major regions with their two or three sub-regions were obviously predetermined as a historical-political schema in order to calculate average values for the subdivisions.

Two more recent surveys should also be mentioned. The first was conducted by the respected Razumkov Centre[11] (2016a) in December 2015,[12] the second was again conducted by KIIS, this time in 2019. Both refer to roughly the same communicative space by asking about "the language used for communication at home" or which language is used with close family. Neither considers Suržyk. Answers are given on the scale: "only Ukrainian – mainly Ukrainian – Ukrainian and Russian (roughly equal) – mainly Russian – only Russian". The territorial structure used in these surveys varies partially from that of the KIIS 2003 survey. The survey by the Razumkov Centre (2016a) assumes "only" one centre that covers the western centre from KIIS (2003) as well as the eastern centre but without the Dnipro Oblast which is assigned to the east. In addition, the east also includes the Zaporižžja Oblast which in KIIS (2003) was assigned to the south. The 2019 survey by KIIS at least nominally uses the same territorial differentiation although some details remain unclear: Is the Dnipro Oblast considered part of the centre as in KIIS (2003), i.e. eastern centre, or part of the east as in Razumkov Centre (2016a)? And does Zaporižžja Oblast belong to the east or the south?[13] Further-

about which language is used (principally) (for Ukraine see Hentschel & Taranenko 2015: 271; for Belarus see Hentschel & Kittel 2011: 110–117).

[9] KIIS just refers to the western central region as "the centre", which we have retained in the map.

[10] This last area is sometimes also referred to as "New Russia", which is actually an old term but if used today would rather express a Russian sense of entitlement, which in the case of the Crimea was also violently implemented.

[11] Cf. the web site: http://razumkov.org.ua/en/.

[12] Razumkov Centre (2016b) is a further survey conducted by this institution in November 2016. We do not consider it here because it is less differentiated than the survey conducted in the previous year and the results are similar.

[13] This is actually unusual in KIIS internet publications. In another survey from 2020 which is not concerned with language, the assignment of oblasts to regions is very similar to that in Razumkov Centre (2016a) with the exception of Xmel'nyc'kyj. In the latter it is assigned to the central region, in the former to the west: http://kiis.com.ua/?lang=ukr&cat

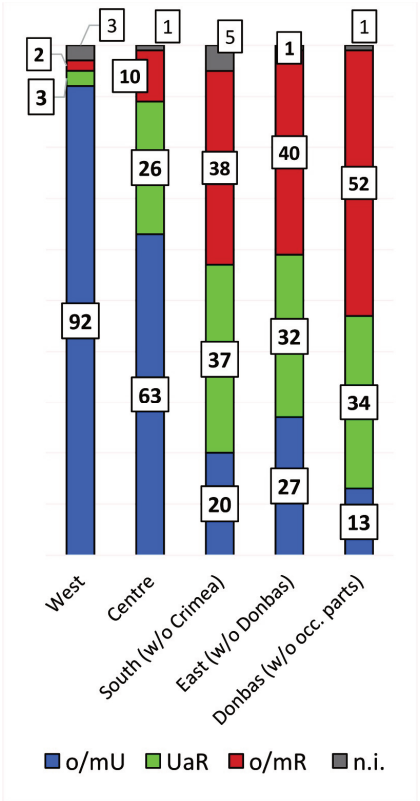


Figure 1a: The distribution of Ukrainian and Russian in % according to Razumkov Centre (2016a)

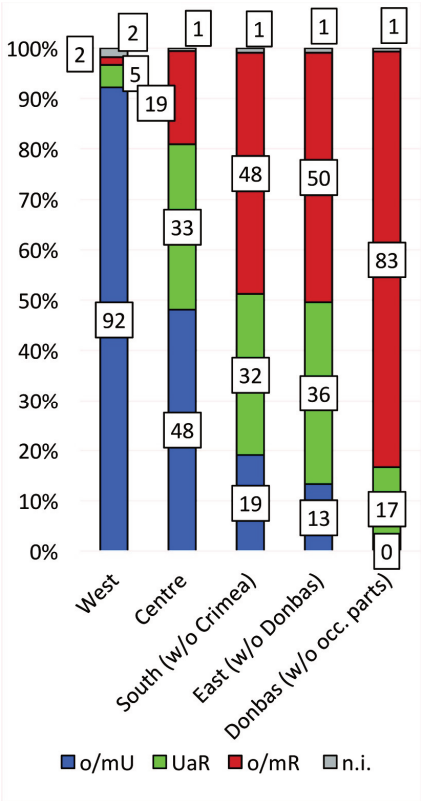


Figure 1b: The distribution of Ukrainian and Russian in % according to KIIS (2019)

more, the question arises of why in neither survey the central region of Ukraine is divided into one part east of the Dnipro river and one part west of it, given that in KIIS (2003) very different results for the two areas were presented.

The last question addresses the general problem of all three analyses mentioned here: "large regions" are presented with varying degrees of transparency as to their historical and/or political motivation. The differences in usage frequencies in smaller areas, such as the single oblasts of these regions, are subsumed under average values. In our study we take a different approach. Of course, we also employ a politically-motivated structuration. We generalise the results from the different survey locations with regard to the administrative unit of the oblast. But their borders only have the status of a coordinate system. The oblasts are then clustered on the basis of the frequencies of code usages there. This means that there are no predefined large regions, instead the clusters indicate regional structuration on the

=reports&id=958&page=1 (accessed 14th July 2020). This means that KIIS (2003) and KIIS (2019) employ very different structurations.

basis of frequency of code usage. In order to clarify the differences and similarities between our approach and the traditionally predefined large regions, we briefly present the results of the surveys by the Razumkov Centre (2016a) and KIIS (2019) in Figure 1a and 1b.[14]

It is unnecessary and would even be wrong to discuss the quantitative differences between the two surveys here in detail. It can be assumed that there are (partially not specified) differences in how the surveys were conducted and how the data are presented. Nevertheless, three trends are visible in both approaches. First, with values of over 90 percent for "only/mainly Ukrainian" the west is both stable and clearly distinct from all other regions. In the south and the east (as well as in the parts of Donbas that are not Russian occupied and that we cannot consider further) the usage frequency of Ukrainian as the sole or principal medium of communication at home or with the family declines very sharply, falling to values around 20 percent in the 2015 survey and even lower in 2019 (in Donbas zero). The centre of the country takes a middle position between the last three mentioned regions and the west. Its values are closer to those of the west in Razumkov Centre (2016a) than in KIIS (2019).

For Russian as the sole or principally used language, the comparison of the two studies shows a mirror image increase from west to east and south that is particularly striking in the unoccupied Donbas.

The central question remains, however, of which of the (three or five) options provided in the two surveys were chosen by respondents who speak Suržyk at home but because of the survey design could not give this answer. This is not only the case for people who usually only speak Suržyk and do not have "adequate" command of both standard languages,[15] but also for those who practise Suržyk in addition to at least one of the two standard languages. Quantitative analyses such as Hentschel & Taranenko (2015: 254) have shown that there are large numbers of such people, many of whom have a university education (cf. Hentschel & Zeller 2017: 45). Recently, informal-qualitative studies have confirmed these findings (Bilaniuk 2018: 298). The latter also describes how over the last 10–15 years the thoroughly negative public image of Ukrainian has at least partially changed for the positive. This makes it even more surprising that Ukrainian surveys no longer consider Suržyk.

[14] In KIIS (2019) the figures for the categories "only Ukrainian" and "mainly Ukrainian" (and analogously for Russian) vary, in Razumkov Centre (2016a) they do not. For better legibility in the chart we combine the categories "only …" and "mainly …" into one: o/mU or o/mR, read: "only or mainly …" vs. UaR, read "… and …", i.e. roughly balanced between the two. The category "n.i. – no information" denotes other languages or missing answers.

[15] Some Ukrainian and Belarusian linguists have even employed "inadequate" competence in Ukrainian or Belarusian in their definitions of Suržyk and Trasjanka. For a critique of this approach, see Hentschel (2017: 22–23).

It is striking that in both surveys from the last five years, and most strongly in the more recent one, a very large proportion of respondents claim to use Ukrainian and Russian to a comparable extent at home and in their families! It is well known that in independent Ukraine, Ukrainians who previously spoke Russian now at least partially have to turn to Ukrainian in public spaces. Without doubt there are also families who switch between Ukrainian and Russian. On the other hand, the private and informal nature of such constellations leads to occurrences of code switching and code mixing that are anything but rare. Whatever the case, respondents who primarily use Suržyk in their homes and families are forced by the predefined options available in the two recent surveys cited here to hide their practice in other categories. It remains unclear how they made their choices: they could have been motivated by political correctness or they assess their Suržyk as being closer to Ukrainian or Russian, or as being balanced between the two (cf. Hentschel & Zeller 2017: 51–53), or it could just be by chance.

3. The analysis

3.1 The data

The data were collected by the two projects named in footnote 1. Both investigations focused on Ukrainian Suržyk, and so-called closed interviews on the respondents' sociolinguistic background were conducted by the company Socis.[16] To this end, respondents were interviewed and the interviewers filled in the questionnaires. The first project, with data collection in early 2014, focused on eleven central Ukrainian oblasts, including one bordering on the west (Xmel'nyc'kyj) and one bordering on the east (Xarkiv). As already mentioned, these two oblasts have been assigned to the west or the east, rather than the centre, in various areal structurations (cf. the maps below in Section 3). It was initially planned to include the two eastern oblasts in the so-called Donbas (Luhans'k, Donec'k) but already in 2014 it was not possible because of the brewing conflict. The centre has, however, always been considered the area where Suržyk is most widely spread, which was most recently confirmed by Hentschel & Taranenko (2015) (see below).

In metropoles of the central areas, such as Kyïv, Xarkiv and Dnipro, Russian traditionally dominates in everyday public life, apart from official occasions where Ukrainian is expected to be used. In the countryside of the central region, rather Ukrainian in dialectal form is in use, presumably with a certain amount of Russian interference. How far the latter reaches is unclear, because traditional dialectology usually ignores the influence "from outside". Along with the limited project bud-

[16] These closed interviews also served to select respondents with the appropriate usage profile in the three codes to be invited for an open interview which was or will be analysed qualitatively. The material collected in the opens interviews also forms the basis for corpus-linguistic analyses.

get this fact prompted the exclusion of metropoles and villages from data collection in the central area. Thus, cities of every size (apart from the above-named metropoles that are considered Russian speaking) were surveyed as well as towns with at least the status of a "place with urban character", since it was deemed most likely to encounter Suržyk there. Between 100 and 150 respondents were interviewed in each of the eleven oblasts, the total number was 1,400.

For the second project, on the Black Sea region, a very similar survey with a partially identical questionnaire was conducted in March 2020. However, this project had a slightly different objective. While the project on central Ukraine focused on the old, Ukrainian-based Suržyk, the second project on the south, mainly considered Russian speaking, looked at both types of mixed Ukrainian-Russian speech. As for the centre, Russian-speaking metropoles were excluded, which here only applied to the city of Odesa. Since the oblasts in the Black Sea region do not count as areas where old autochthon East Slavic dialects were spoken (with the exception of some northern and north-western border regions of the Odesa and Mykolaïv Oblasts), villages were also included. The three Black Sea oblasts (see below) were surveyed.[17] Roughly 400 randomly-selected respondents were interviewed in each oblast.[18] They are aged between 18 and over 80 years with an even distribution up to the age of roughly 70 years. The proportion of males and females is virtually equal. All levels of education are present in proportions close to those of the general population.

Surveys on language use in multilingual societies, and therefore this survey too, do not "measure" actual language use. Such an "objective" quantification of highly variable mixed forms of speech derived from two structurally similar and closely related donor languages, like Suržyk, would not be theoretically and methodologically trivial. Surveys are based on respondents' "subjective" estimations and statements. Hentschel & Taranenko (2015: 265–271) discuss this "subjectivity" for the linguistic situation in Ukraine. Following their discussion, it can be assumed that on the whole the respondents' estimations match the reality of language use well, and that misestimations in either direction cancel each other out. Of course, a further problem may arise, e.g. from statements that are politically-emotionally motivated and may even contradict the respondent's self-assessment. In a situation of conflict, as is currently the case in Ukraine with regard to Russia, it is occasionally advisable to be particularly cautious in interpreting quantitative observations. The following analysis draws attention to such occasional problems of interpreta-

[17] Zaporižžja Oblast, which lies further east on the Sea of Azov, was not included because while the project was being planned the armed conflict was threatening to spill over into this oblast. It was not considered sensible to collect linguistic data (family conversations, interviews) as a follow-up to the initial survey in an area close to a war zone.

[18] Both questionnaires can be found on the website of Oldenburg University at https://uol.de/slavistik/forschung/sprachwissenschaft/hybridisierung-von-zwei-seiten.

tion. Generally, however, we simply speak of "usage" while being aware that statements about the usage of the three codes are meant.

3.2 The strength of the single codes and their combinations

3.2.1 The respondents' nationality and choice of interview language

The nationalities of the respondents in the central region and the Black Sea region are very similar. In the three Black Sea oblasts, 90 percent said they had Ukrainian nationality, in the centre it was 96 percent; correspondingly, 10 percent and 4 percent respectively gave their nationality as Russian. Respondents with other nationalities were not included in the survey. The pattern is similar among the respondents' parents and spouses. In the Black Sea region, 80 percent of respondents had parents with Ukrainian nationality and 13 percent with Russian; in the central region the values were 88 percent for Ukrainian nationality and just under 10 percent for Russian. The remainder of the parents had other nationalities (slightly more often in the south than in the centre) or the answers were missing. The respondents' spouses were Ukrainian nationals to 88 percent in the south and to 95 percent in the centre, and Russian nationals to 8 and 4 percent respectively; the very small remainder had other nationalities.

While the composition of both groups of respondents is very similar in regard to the proportions of Ukrainian and Russian nationals, there were large differences in the choice of interview language.[19] In the central region 83 percent of respondents chose Ukrainian as the interview language and thus 17 percent chose Russian, in the Black Sea oblasts only 46 percent chose Ukrainian and 54 percent Russian. However, in two oblasts of the central region there was also a preference for conducting the interview in Russian: in the Dnipro Oblast 65 percent chose Russian, and in Xarkiv Oblast even more clearly with 83 percent.[20] In the Black Sea region there is a pronounced correspondence between nationality and the choice of interview language: whereas respondents with Ukrainian nationality chose Ukrainian and Russian in equal proportions for the interview, 92 percent of the 122 respondents with Russian nationality chose to conduct the interview in Russian. In the central region, the "markedness relations" were reversed: 85 of respondents with Ukrainian nationality chose Ukrainian as the interview language. (Three-quarters of those who chose Russian were from Dnipro and Xarkiv Oblasts.) In contrast, the small number of respondents with Russian nationality

[19] The individual questions were posed by an interviewer who entered the answers into a Ukrainian or Russian questionnaire. Please note that there were many questions other than those on the use of codes (e.g. for linguistic biographies, attitudes, etc.), which will be analysed in future papers.

[20] In all the other central oblasts surveyed in 2014, the interviews were conducted in Ukrainian in at least 80 percent of cases.

(57) chose the Ukrainian and Russian questionnaires in equal proportions. The choice of language for the interview already provides an initial indication for the greater importance of Russian in the south and (partially) in the east,[21] even if nearly half the respondents chose the Ukrainian questionnaire, while in Xarkiv Oblast in 2014 it was only 20 percent.

3.2.2 The primary codes

The primary code is understood here to be the code the respondents chose in answer to the question: *Which of the following codes do you use most often in everyday life?* The answer could be given on the five-point scale illustrated in Figure 2.

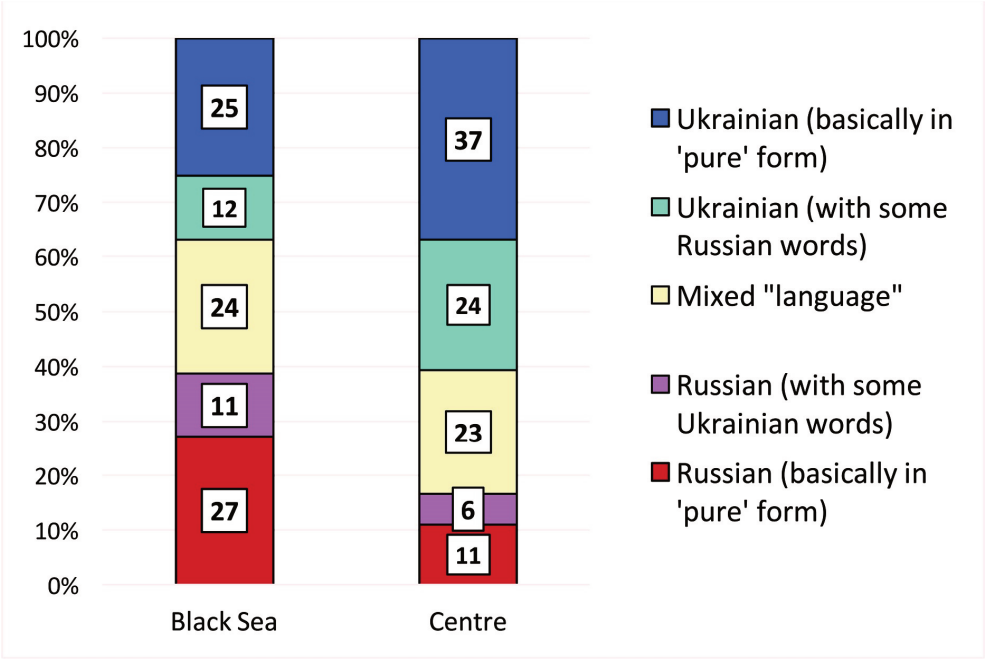


Figure 2: Primary codes on a five-point scale

Before presenting the results, a few words on the scale are necessary. A five-point scale that differentiates two sub-cases of Ukrainian and Russian, namely "with some words" of the other language or without, was chosen in order to avoid the appearance of an "inflationary" spread of Suržyk. The prevalence of Suržyk is already higher in our surveys than in others, which in our minds is connected to our avoidance of the stigmatising term "Suržyk" in the questionnaire itself in fa-

[21] Of course, Hentschel & Taranenko (2015) also confirm a gradually decreasing importance of Ukrainian from west to east for central Ukraine, but also point to a similar one from centre to periphery (see below).

vour of the more neutral "mixed language".[22] It is likely that the common stigmatisation of Suržyk as a style of speaking among the uneducated, if not "un-Ukrainian Soviet people", but also the understandable endeavours of norm-conscious teachers of the Ukrainian language to maintain a so-called language culture, prevent many contemporaries from admitting to their usage of Suržyk. Using terms with negative connotations in surveys is more than questionable. The five-point scale was chosen here to avoid Ukrainian speech with various, sporadic (ad hoc) borrowings from Russian or Russian speech with such borrowings from Ukrainian being evaluated critically and thus increasing the values for Suržyk. If the "convinced" speakers of the supposedly "pure" national languages are compared with the "uncertain" speakers in terms of purity, that means for Ukrainian and Russian in both regions, then we find that in the Black Sea region nearly one quarter of respondents and in the central region just over one quarter state that they speak one of the recognised languages in such a way that they use some words from the other. We will not elaborate on this point here, since an additional study of actual speech data will be conducted for further differentiation.[23] In the following discussion we restrict ourselves to a three-level differentiation derived from the above five-point scale: Ukrainian, Russian and Suržyk. The latter encompasses only relatively balanced or strongly mixed speech. The proportion of respondents in both regions who consider this their primary medium of communication is between one fifth and one quarter.

Figure 2 above already shows (not surprisingly) a greater importance of Russian in the Black Sea region and that Ukrainian is more prevalent in the central region,[24] while the proportion of Suržyk is roughly equal in both regions. However, Hentschel & Taranenko (2015) already discovered great variation in five sub-regions of the central area, illustrated in Map 2.[25]

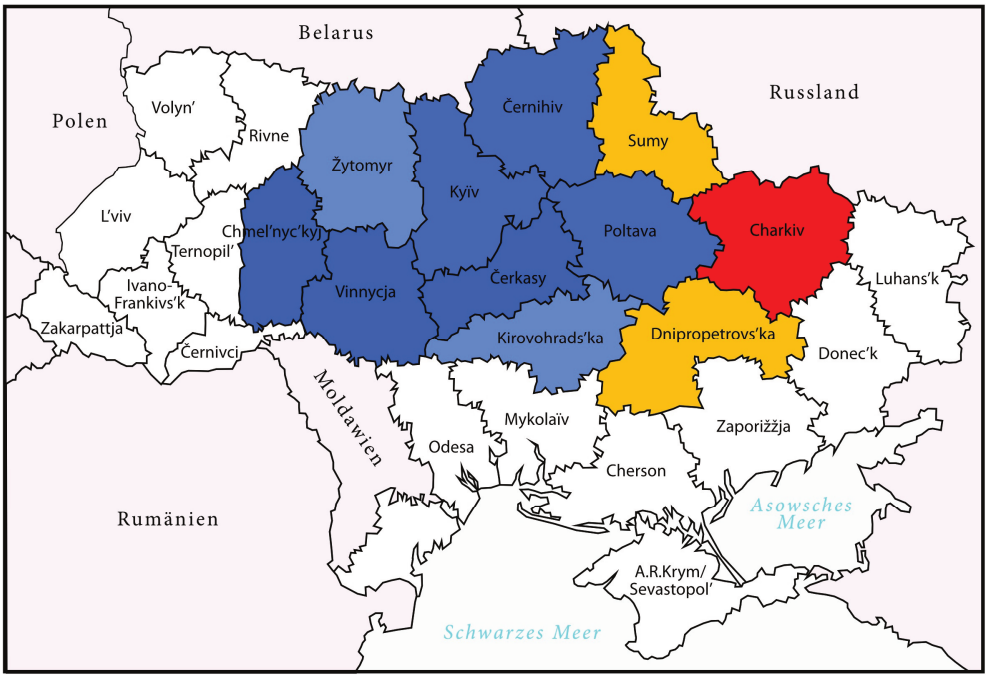
These five sub-regions of the central area will be compared with the three Black Sea oblasts on the three-level scale. The left-hand side of Figure 3a excludes vil-

[22] Using the term "mixed language" in the questionnaire of course does not mean that we consider Suržyk a special language, quite on the contrary, see below. Hardly any of the respondents have any linguistic knowledge and the colloquial term "language" (Ukrainian *mova* / Russian *jazyk*) is known to denote phenomena that linguists would refer to as "speech", "regional dialect", "social dialect", "register" etc. There would be no point in confronting non-specialists with such terms. Similarly, the use of the term "word" in the questionnaire is also more colloquial, since at least with inflected elements "word form" would be more accurate.

[23] This is planned for a later point in the project.

[24] In this article, "centre" always refers to the area studied by the first project in footnote 1 and described above.

[25] The division into five regions is based on the usage frequencies of the three codes. This reasons for this differentiation are described in more detail below.



Map 2: The weighted prevalence of the three codes according to Hentschel & Taranenko (2015)

A – Xmel'nyc'ky, Vinnycja, Čerkasy
B – Kyïv, Černihiv, Poltava
C – Žytomyr, Kirovohrads'ka obl.
D – Sumy, Dnipropetrovs'ka obl.
E – Xarkiv

lages in the south (as previously mentioned, villages were not surveyed in the central region), whereas Figure 3b, which only represents the south, includes villages.

There is no great divergence between the three Black Sea oblasts: roughly one third of the respondents give Ukrainian as their primary language, between 40 and 50 percent give Russian. Russian dominates, but not to the same extent as in KIIS (2003).[26] The more recent surveys by the Razumkov Centre (2016a) and KIIS (2019) also indicate a strengthening of Ukrainian in the south.[27] Greater differences between the three oblasts are visible for Suržyk: only one in twelve respondents in Mykolaïv name it as their primarily code, one in three respondents in

[26] The particularly low proportion of Russian in Odesa Oblast arises because the metropole, i.e. the city of Odesa, is not included in the two projects focussing on Suržyk because the metropoles are generally considered Russian speaking.

[27] The likely strengthening of Ukrainian in southern and eastern regions over the last two to three decades is not the subject of this study but will be investigated in a further one that is also part of project.

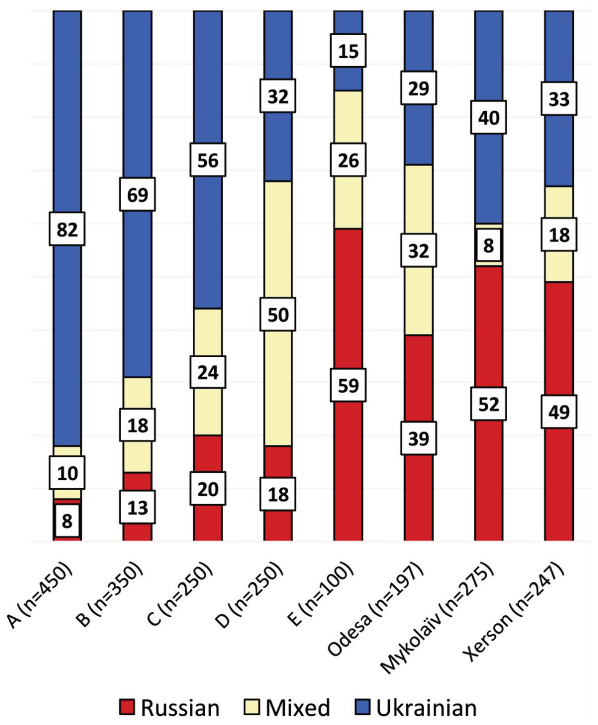


Figure 3a: The primary codes in the three Black Sea oblasts and the five central regions in %

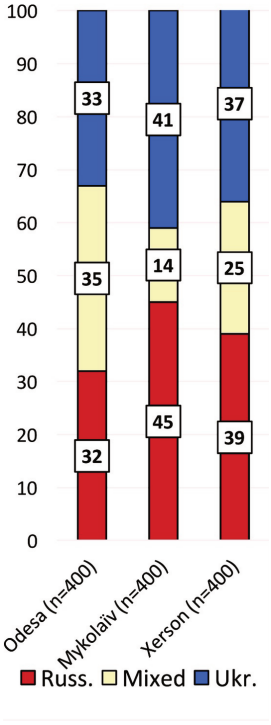


Figure 3b: The primary codes in the three Black Sea oblasts (including villages) in %

Odesa, and in Xerson nearly one in five. Suržyk is mentioned more often in our survey, which is related to the survey design discussed above.

On the whole, the three Black Sea oblasts studied are most comparable with central region E, which only encompasses Xarkiv Oblast. However, in the latter region Ukrainian plays a much smaller role as primary code, being even less common than Suržyk. The Odesa Oblast is similar to central region D (Sumy, Dnipro), at least loosely with regard to the presence of Suržyk, which is, however, stronger in region D. The values for Ukrainian are almost the same. Russian is much weaker in Dnipropetrovs'ka Oblast, not only in comparison to Odesa but also to Xarkiv. This indicates the questionability of linguistically assigning Dnipro, like Xarkiv, to the east. The regions A, B and C of the western centre according to Hentschel & Taranenko (2015) are distinct from all others because in these three Ukrainian is named as the primary code by the overall majority of respondents. The differences here are of a gradual nature and are expressed in a declining importance of Ukrainian and a corresponding increase in the values for Russian and Suržyk. Although these findings are comparable with those of KIIS (2003), their presentation of a "western centre" as a linguistic unit does appear to be highly questionable. This is

even more so the case with the broader unit of the "centre" in the two other above-mentioned surveys (Razumkov Centre 2016a, KIIS 2019), which also includes region D from Figure 3a.

Finally, it should be noted for the south that the proportion of Suržyk increases when the data from the villages are included, as seen in Figure 3b. The proportion of Russian sinks, which underlines the importance of Russian in the cities.

3.3 Multicodalism – the coexistence and combination of the three codes, speaker types and their areal distribution

Focusing on the primary code as that which is principally used, as in the previous section and the above-cited Ukrainian surveys (which are restricted to the domestic-private realm) can, of course, only provide limited insights into the respondents' everyday linguistic life in their places of residence. This is also the case if, as the above-cited Ukrainian sociologists have recently done, the two codes recognised as standard languages are considered to be equivalent in the surveys. Hentschel & Taranenko (2015: 253–255) suggested a different approach for the central areas of Ukraine that focused more on the coexistence of the codes in the different areas and less on constructed, binary complementarity. They could show that most speakers in central Ukraine use at least two codes in everyday life, a smaller number use all three. Furthermore, they were able to identify fine gradations in the frequency of usage, also for Suržyk. Their approach, as well as that reported here, was not restricted to the domestic-private realm but everyday life in general, including work life.

Map 2 above illustrates the regional similarities and differences identified by Hentschel & Taranenko (2015) in their analysis. The five areas that were modelled are thus not based on an external (political-historical motivated) a priori subdivision of Ukraine, but – as previously mentioned – use the oblasts as a "coordinate system".[28] Rather, the five regions are the result of a cluster analysis based on the frequency of designating the codes as being used all the time or used often. In contrast to the customary representations, Hentschel & Taranenko (2015) consider all three codes in drawing a picture of everyday linguistic life in the centre of Ukraine that presents forms of monocodalism and roughly equally weighted multicodalism, both with their different expression across oblasts and regions.

Turning first to the "coexistence" of the codes (leaving aside the combinations practised by individuals), the picture at the Black Sea coast of the coexistence of the three codes is very similar to that described by Hentschel & Taranenko (2015:

[28] Of course, there are differences within the oblasts, but capturing them fully would imply an enormous empirical effort, regardless of the motivation for wanting to find out.

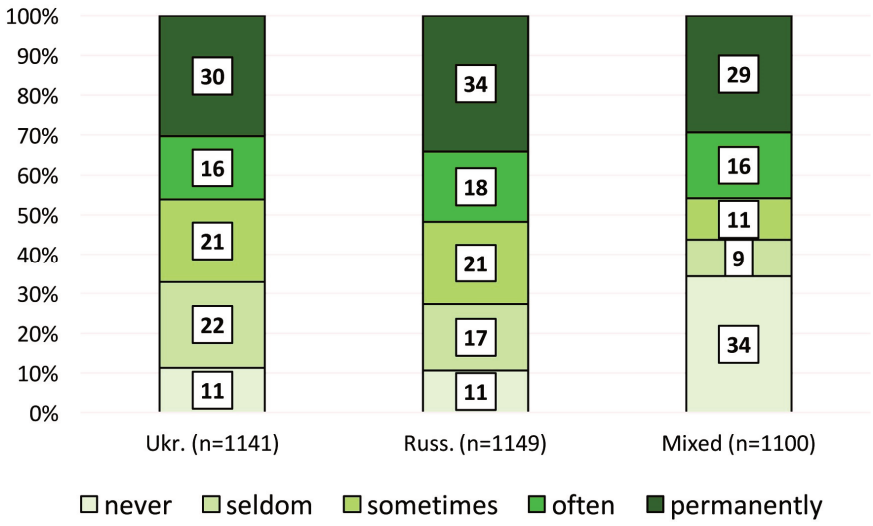


Figure 4: The coexistence of the three codes in the Black Sea region and their intensity of use

253) for the central region.[29] The only difference is that in the south the values for Ukrainian and Russian are broadly similar, whereas they were weaker for Russian in the central region (cf. Figure 4).

All three codes are each used by around one half of the respondents often, if not always. The latter is the case for around one third of the respondents. Russian is more frequently described as being used often than Ukrainian or Suržyk. The most striking finding, however, is that only one in ten respondents say that they never use either of the standard languages. For Suržyk the proportion is around one third. The different number of cases, *n*, in the three columns can be explained by the possibility to answer "hard to say". These "negative" answers were excluded from the frequency calculations shown in Figure 4 (see below for more details). This option was chosen twice as often for Suržyk (exactly 100 times) as for the two recognised languages. This can also be viewed as a reflex prompted by a kind of shame associated with admitting to using the often stigmatised mixed speech.

The numbers in Figure 4 clearly show that speakers often combine codes, i.e. they choose according to situation, conversation partner, etc.[30] It still remains to be seen which combinations in which proportions they use, that is how individuals "combine" codes, or in other words to what extent their communication is multilingual or, better, multicodal.

[29] In this initial overview of the relations there is no need to repeat those findings here. Similarities and differences will become apparent below.

[30] See Hentschel & Zeller (2017) with regard to the central region. A forthcoming study will present findings for the Black Sea region.

In what follows a more strongly differentiated model of bi- and tricodalism in Ukraine will be developed than that presented in Hentschel & Taranenko (2015). It proceeds from the usage of the three codes by each individual speaker and not from the codes' general incidence in regions. In addition, the model also aims to adequately consider usage frequencies for the codes that are less than "often". Respondents can not only "overrate" their usage frequency for a code but they can also "underrate" it, e.g. for reasons of "political correctness". The model should thus consider as much of the data as possible. The focus is therefore on different forms of individual combinations, or types of speakers, and less on the coexistence of (often or permanently used) codes in the oblasts and regions. In a second step, the frequency of these speaker types in single oblasts will be used to sketch a map of multicodalism or rather a map of the multicodalism of speakers in central Ukraine and the Black Sea region.

A cluster analysis (using SPSS 26) will initially investigate which constellations of multicodalism can be used as a basis for grouping speakers (clusters) when all five stated levels of frequency of code usage are considered: from always to never. The clusters are subsets of respondents who, in terms of their frequencies of usage of the three codes, are maximally homogenous but maximally different from all other subsets. Given the large number of respondents (over 2,000 in total), a cluster centre analysis using k-means (Quickcluster) is appropriate. In this method the number of clusters to be investigated (but of course not the type) is determined a priori (cf. Bühl 2019: 635–696). The number of clusters can be arrived at by, for example, qualitative considerations. We decided on seven, based on the logical possibilities for using three codes: three monocodal clusters, three bicodal clusters and one tricodal cluster. The test will discover which constellations are most prominent, based on each of the five-level frequency statements for each of the three codes. The clusters constructed with this procedure represent certain speaker types, that is types of speakers with a specific profile for their extent of usage of the three codes. These do not necessarily have to correspond to the seven logical possibilities, they could also be gradations of one part of the seven combinations, e.g. if one or more of the logically possible combinations are not empirically confirmed.

The cluster analyses for the speaker types were first conducted separately for the Black Sea region and central Ukraine, because – as previously mentioned – the selection of respondents for each of the projects providing the data was slightly different: in central Ukraine villages were excluded.

The following picture emerges (Table 1), initially separate for the Black Sea region (left) and central Ukraine (right):

How often Ukr.?	*How often Russ.?*	*How often Mixed?*	*Type*	*%*	*How often Ukr.?*	*How often Russ.?*	*How often Mixed?*	*Type*	*%*
4.92	2.01	1.60	U	23	4.90	1.50	1.65	U	19
1.99	4.94	1.45	R	26	2.17	4.82	1.80	R	12
4.61	4.16	1.31	U-R	5	4.66	3.20	2.41	U-R	13
1.57	1.54	4.96	S	4	2.43	2.36	4.75	S	11
3.79	1.74	4.72	S-U	7	4.51	2.22	4.09	S-U	30
1.79	4.06	4.40	S-R	10	2.49	4.24	4.43	S-R	6
3.67	3.69	4.44	S-U-R	24	4.32	4.15	4.13	S-U-R	10
Black Sea				*N* = 1,030	*Centre*				*N* = 1,400

Table 1: Clusters of speakers based on usage frequencies for the three codes: the two large regions compared

In these two separate analyses the dependent variables were the five-level usage frequencies for the three codes given by the respondents (who constitute the independent variable). If a code is never used, it receives the value of 1; if it is always used, the value is 5. The column of numbers below the question "How often …?" are the mean frequencies for that code. The combination of all three mean frequencies in each row of the table constitutes the so-called cluster centre. In the first (left-hand) part of the table for the Black Sea region, the value in the top left cell is 4.92. This means that the respondents in this cluster use Ukrainian almost all the time. In the cell to the right is the value for the average frequency of using Russian, here 2.01, i.e. rarely. In the next cell to the right is the average for Surzhyk; at 1.6 this is even lower than that for Russian and is close to non-usage (in other words: very many respondents in this cluster do indeed not use it). We classify this cluster, or rather speaker type, as virtually "monocodal" Ukrainian speakers, even though at least some occasionally, but on average rarely, resort to Russian and even less often to Surzhyk.

By and large the clusters correspond to the seven logical possible conditions of the codes being used often if not permanently. The cluster U-R in central Ukraine is a more obvious exception. Here the mean for Russian is nearly 1.5 points less than that for Ukrainian and closer to that for Surzhyk. The relation between Ukrainian and Russian is more evenly balanced in the type U-R in the Black Sea region. This means that the cluster U-R in central Ukraine is not as clearly distinguished from the type S-R as at the Black Sea coast.

When the three means in each row are interpreted in this fashion, then the seven constellations mentioned above can be confirmed as speaker types: (1) three mainly monocodal, (II) three mainly bicodal and (III) one mainly tricodal type. The type is determined by the highest mean values; in each of the seven clusters

at least one mean is greater than 4 (these cells are marked green). If the cluster is rather multicodal, two or three means are either greater than 4 or at least close to it (the latter are marked light grey).

It should be noted that not all of the respondents from the Black Sea region were considered: of the 1,200 respondents, those 1,030 were included who explicitly answered all three questions on the usage frequency of the three codes, even if they stated that they never use one or more code. If one of the three questions was not answered, the respondent was excluded from this analysis. It is not possible to automatically equate a missing answer with non-usage of a code. A missing answer may be motivated by the desire to not mention that one uses a certain code, with whatever frequency.[31]

In order to better compare the data from the Black Sea region with that from central Ukraine in the further analyses, it is necessary to exclude a subset of respondents from the Black Sea oblasts. As previously mentioned, no data was collected in villages of the central region. Therefore, only respondents from settlements that have as least as many inhabitants as the smallest settlement included from the central region were considered in further analysis. The threshold is 4,000 inhabitants. This reduced the number of respondents to 2,029 (Black Sea: 629, central region 1,400) and brought the number of respondents per oblast in the Black Sea region closer to that in central Ukraine.

How often Ukr.?	*How often Russ.?*	*How often Mixed?*	*Type*	*N*	%
4.86	1.62	1.89	U	437	21.5
2.00	4.88	1.26	R	298	14.7
4.70	3.42	1.49	U-R	154	7.6
2.41	2.33	4.80	S	213	10.5
4.43	1.82	4.44	S-U	225	11.1
2.29	4.46	3.96	S-R	241	11.9
4.43	3.52	3.93	S-U-R	461	22.7
		total:		2,029	100.0

Table 2: Clusters of speakers based on usage frequencies for the three codes: Pooled data from the two regions

A new cluster analysis was conducted, but this time without differentiating between the Black Sea and the central region (Table 2). The results are very similar to those of the two independent models for the Black Sea region with slightly

[31] We can only speculate about why answers to these three questions are only missing in the Black Sea region but not in the central region. Data on the central region were collected in early 2014, at a time when relations between Ukraine and Russia were not exactly harmonious but neither the tragic events at the Maidan, the fall of Janukovyč's government, the annexation of the Crimea by Russia nor the war-like events in the east of Ukraine were salient issues. As already mentioned, the data from the Black Sea region were collected in early 2020.

different mean values for the cluster centres. More importantly, the analysis yielded the same speaker types.

In general, the findings can be characterised as follows: monocodal speakers oriented to either Ukrainian or Russian (U and R) constitute "only" a good third of respondents, with speaker type U being one and a half times more widespread than type R. On average, speakers of these two types rarely if ever use the respective other "recognised" code or Suržyk. Speakers oriented to both languages, the type U-R, who at best only use Suržyk peripherally, are relatively rare and make up only 8 percent of all respondents. The majority of respondents of the type U-R tend to slightly favour Ukrainian. We can interpret these three types as one block with a purist orientation, characterised by a marked "abstinence" from Suržyk. In total, these "purist" speakers oriented to the national languages comprise roughly 42 percent of speakers.

This means that, conversely, speakers for whom Suržyk plays a relatively important role form the majority. The types S, S-U, S-R and S-U-R together have a proportion of 58 percent even though, as indicated above, Suržyk was and is stigmatised by large parts of the elites in the country.[32] It is therefore not unreasonable to assume that some respondents minimised or omitted their usage of Suržyk, which would mean that it is even more prevalent than presented here. Of these four speaker types, S-U-R is the most widespread, accounting for (similar to the monocodal type U) more than a fifth of all respondents. The other three types with "S" are all around the same size, each accounting for a good tenth of respondents, including the monocodal type S. In surveys that do not include Suržyk as an answer category, respondents of monocodal type S as well as those of type S-U-R (altogether one third of respondents) would have to give inaccurate answers to questions about their domestic or family language. In this situation, speakers of the types S-U or S-R could choose Ukrainian or Russian, but these tend to be used more in public or formal contexts and less at home.

If the speakers of the two national languages are added in a similar fashion, regardless of whether they are mono- or multicodal, a similar picture emerges with roughly 63 percent speaking Ukrainian and 59 percent Russian. Therefore, all three codes are present in Ukrainian society to a similar extent, although Suržyk is of course a decidedly "low variety" that nevertheless is used with a similar frequency to Ukrainian and Russian. And one last point: in clusters where Ukrainian and Russian are combined as extensively used codes (mean > 3), it is Ukrainian that gets the higher value.

[32] Which is why it is certainly not unreasonable to assume that some respondents minimised or neglected to mention it use, which means that is probably more widespread than is presented here.

The situation we have just described is of course a generalisation for all the oblasts included. Hentschel and Taranenko (2015) already established considerable differences between the oblasts of central Ukraine in their previously mentioned simpler model, but they could also cluster the oblasts. The next step is therefore to take the results of the clustering of individual speakers and to differentiate them according to oblast. First, we adopt a descriptive perspective on the situation in the 14 oblasts considered. Figure 5 presents the share of the seven speaker types in each oblast. The oblasts are presented in a rough order, starting with a high share of Ukrainian on the left. This share decreases from left to right in favour of an increasing share of Suržyk in the four constellations mentioned (S, S-U, S-R, S-U-R), represented by some brown background in the columns, from very light ochre to dark brown. The latter again decreases when moving to the right in favour of an increasing share of Russian. Ukrainian and Russian alone or the combination of both without Suržyk are represented by blue, red and green

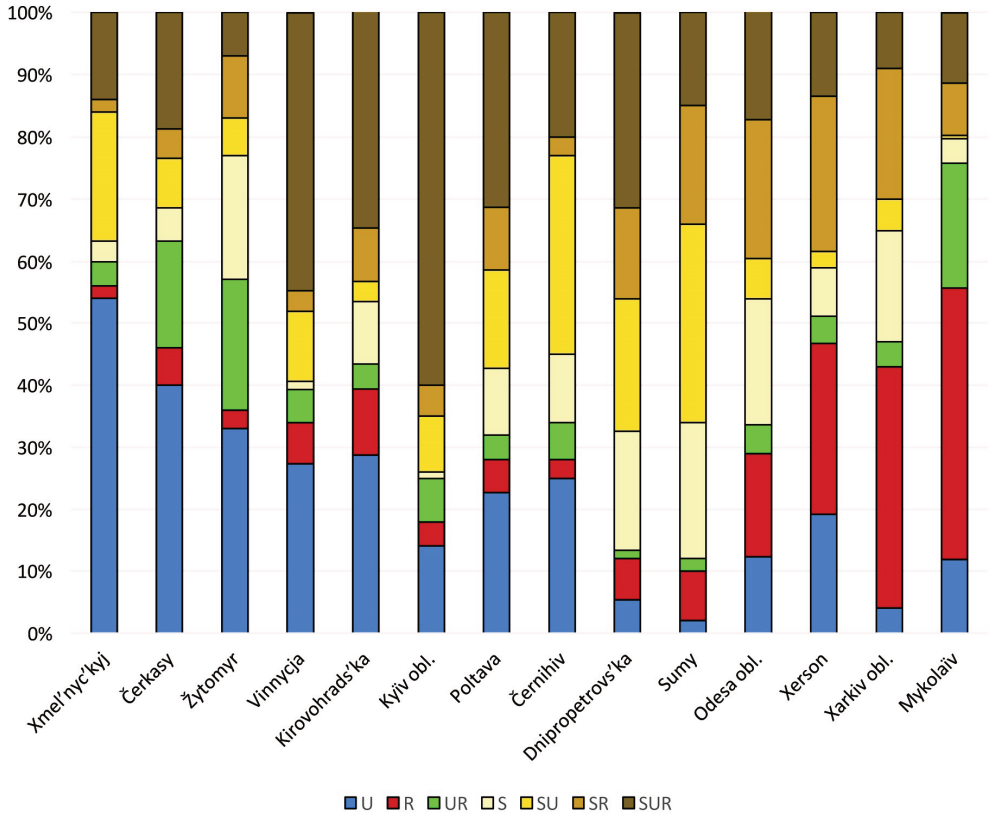


Figure 5: Mono- and multicodalism in the oblasts

Please note: Brown background, from very light ochre to dark brown, means Suržyk or with Suržyk as one of the codes.

in the columns. Geographically speaking, one can roughly observe an increasing presence of Suržyk moving from the west eastwards, reaching a peak in the oblasts of Sumy and Dnipro (Dnipropetrovs'ka oblast). The presence of Suržyk again decreases eastwards and southwards (see Figure 5).

Only some selected observations shall be made at this point: It is remarkable that in ten oblasts Suržyk plays a major role, being frequently spoken by half of the respondents or even more, usually in combination with at least one of the other two codes. This is most striking in Dnipro and Sumy Oblasts where only one in ten respondents does not often speak Suržyk in everyday life. Only in Xmel'nyc'kyj, Čerkasy and Žytomyr Oblasts (rather central oblasts in the west of the territory studied) as well as in Mykolaïv (the middle oblast in the Black Sea region) does Suržyk shift more to the background, being regularly spoken by less than half of the respondents. Mykolaïv stands out with a surprisingly high proportion of speakers who say they do not use Suržyk, the combination S-U being virtually absent (with a proportion of less than one percent), the lowest rate of Suržyk in the whole territory under investigation. The two neighbouring oblasts on the Black Sea coast show a considerably higher proportion of Suržyk. Xarkiv stands out with its relatively high proportion of monocodal Russian speakers. Together with Xarkiv the three Black Sea oblasts are traditionally considered Russian-speaking regions in Ukraine (cf. Map 1 on KIIS 2003). The latter is not completely incorrect, but also not completely correct either. The other two codes in different constellations are present to a high degree. The proportion of monocodal Russian speakers in Xarkiv and Mykolaïv reaches a maximum of "only" 40 percent. These figures are certainly comparable with those of the Razumkov Centre (2016a) or also of KIIS (2019). Both our surveys and the two just mentioned find similar proportions of monocodal Ukrainian speakers in these four oblasts, generally under 20 percent. While the surveys by the two Ukrainian sociological institutes in the last five years suggest a share of speakers with an individually balanced relation of Ukrainian and Russian usage of roughly 40 to 50 percent, our data give reason to believe that a large number of Suržyk speakers are hidden in these figures. To what extent their Suržyk is "coloured" more by Ukrainian or by Russian can only be elucidated in so far as the proportion of respondents who are more strongly oriented to Suržyk and Russian, and thus whose Suržyk should be "more Russian-like", is rather high in comparison with the more western and northern oblasts.

The other ten oblasts apart from Xarkiv and the three Black Sea oblasts have only one common characteristic: the lack of a large number of monocodal Russian speakers. Otherwise, they are very heterogenous. The last selected observation to be made concerns one of them – Kyïv: here the majority of respondents (about 60 percent) says they use all three codes more or less evenly with high frequency.

Only the neighbouring oblast Vinnycja comes close to that (with little more than 40 percent).

The descriptive picture of the situation in the 14 oblasts presented in Figure 5 should, however, be formulated both more generally and more precisely: One of the central aims of the study is to describe the map of the Ukraine in terms of the three central linguistic codes: Ukrainian, Russian, and Suržyk, in other words in terms of the weight of the three codes in the area considered. Figure 5 does not mirror the following aspect, which will be illustrated by a simple example: For speakers who state that they use one code very often, say Ukrainian, but the other codes at best rarely (the speaker type U), Ukrainian plays a much more important role than for speakers who claim a very frequent use of not only Ukrainian but also of the two other codes, here Russian and Suržyk (the speaker type S-U-R). As a matter of fact, the meaning of a specific value given in a frequency judgement (here: "very often") in these two cases is different, depending on the combination it occurs in when three codes are asked for. This aspect will be considered in the further analysis.

As a first step, rating points were calculated for the codes in each oblast. For each monocodal speaker of the types U, R or S, the oblast received three points for the corresponding code. For each bicodal speaker, the oblast received one and a half point for each of the codes spoken. For every tricodal speaker, the oblast received one point for each of the three codes. In other words, there are maximally three points per speaker, to be allocated to the one, two or three codes that the speakers use regularly and frequently. Due to the different numbers of respondents in the oblasts, the points were normalised in a second step so that they sum to 100 and thus can be treated analogously to percentage values. These values represent an index of the weight of the three codes, taking into account the mono- and multicodalism of the speakers.

The values for each code in each oblast calculated in this way were then subjected to a cluster analysis. For a small number of cases – here the 14 oblasts – a hierarchical cluster analysis (cf. Bühl 2019, 635–696) is appropriate. The number of clusters calculated ranged from 2 to 7. The results are illustrated in Table 3.

If we were to differentiate just two clusters, then Xerson, Xarkiv and Mykolaïv would be grouped together and would thus be separated from all 11 other oblasts (the last column on the right). In the next step, for three clusters, a further three oblasts would be separated from these 11: Sumy, Diniporpetrov'ska and Odesa, leaving a remainder of eight oblasts. When differentiating four clusters, Mykolaïv becomes distinct from Xarkiv and even the neighbouring Xerson; more distinct from any one out of the eight Xmel'nyc'kyj, Čerkasy, Žytomyr, Kirovohrads'ka, Kyïv, Poltava, Vinnycja, Černihiv and most distinct from the remaining seven oblasts. The latter form a rather compact block that would be divided only in the

next step of a five-cluster solution. In this case Xmel'nyc'kyj and Čerkasy would be separated from the other six.

This procedure could be continued up to 14 "clusters", each of which would contain only one oblast. Thus, applying Ockham's razor, clusters become meaningless. As Table 3 illustrates, with seven clusters, the three southern oblasts and Xarkiv would each already form their own cluster. Consequently, we decided to adopt the five-cluster solution in maps or figures for illustrating the continuum character of the linguistic situation in Ukraine. With five clusters, only one oblast constitutes its own cluster. For reasons of clarity, the five clusters will be denoted hereafter by Roman numerals from I to V.

The proximity relations between oblasts and clusters of oblasts can better be illustrated by graphical means. The cluster analysis applied yields a dendrogram (a type of tree diagram) illustrating distance and similarity between the cases (oblasts) (Figure 6).

Clusters / Oblast	7	6	5	4	3	2
1: Xmel'nyc'kyj	1	1	1	1	1	1
2: Čerkasy	1	1	1	1	1	1
3: Žytomyr	2	2	2	1	1	1
4: Kirovohrads'ka	2	2	2	1	1	1
5: Kyïv	2	2	2	1	1	1
6: Poltava	2	2	2	1	1	1
7: Vinnycja	2	2	2	1	1	1
8: Černihiv	2	2	2	1	1	1
9: Dnipropetrovs'ka	3	3	3	2	2	1
10: Sumy	3	3	3	2	2	1
11: Odesa	4	3	3	2	2	1
12: Xerson	5	4	4	3	3	2
13: Xarkiv	6	5	4	3	3	2
14: Mykolaïv	7	6	5	4	3	2

Table 3: Combinations of scaled distance clusters (from 2 to 7 clusters)

The location where branching occurs is measured in Figure 6 on the horizontal line at the top on a scale from 0 to 25 indicating the degree of distance. So, the first branching at distance 25 singles out the block of Xerson, Xarkiv and Mykolaïv from all other blocks, thus mirroring the two-cluster solution in the right-hand column of Table 3 above. At distance 19, the three-cluster solution, the oblasts Odesa, Sumy and Dnipropetrovs'ka branch from the remaining eight (listed first in the vertical list of oblasts), and so on. The branching continues to the final point of the 14 oblasts. Thus, it is easy to see where the situation is most similar. However, not just the pairing under one branch should be considered, but also the distance at which that branching occurs is important too: There are, for example, three pairs of oblasts that are only separated at the minimal distance of 1: Kirovohrads'ka and Kyïv, Žytomyr and Poltava, Dnipropetrovs'ka and Sumy. The everyday usage of the three codes by the respondents resident in these three pairs of oblasts is maximally similar. The cluster analysis yields two other pair of oblasts under one last branching: Xmel'nyc'kyj and Čerkasy as well as Xerson and Xarkiv. Here, however, the branching takes place at dis-

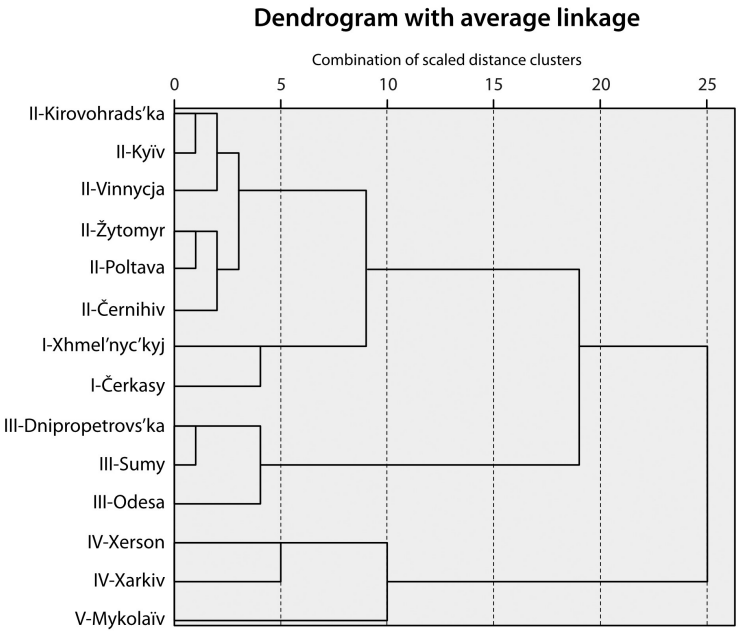


Figure 6: Dendrogram of the hierarchical cluster analysis based on weighted frequency of speaker types

tance 4 and 5, respectively. Thus, the oblasts within the latter two pairs differ more strongly from each other than those in the three pairs of oblasts first mentioned. Moreover, the situation between Xmel'nyc'kyj and Čerkasy or Xerson and Xarkiv is more heterogenous than that in the first six oblasts at the top of the list. The differentiation between the latter starts at distance 3, and not at 5 as with Xerson and Xarkiv.

Unfortunately, it is not possible to directly draw conclusions about similarity relations between clusters containing few elements (oblasts) just from their vertical order in the dendrogram produced by the statistics program SPSS. Thus, for example, cluster I in the five-cluster solution, i.e. Xmel'nyc'kyj and Čerkasy, is by no means more similar to cluster III than to cluster II.

These relations between clusters and even the finer similarities and differences within in the clusters can be visualised in a map. Map 3 fulfils this function. The five areal clusters, noted to left of the oblasts' names by Roman numerals, remain recognisable, as can be seen in the map.

Each of the fourteen oblasts has a specific colour that corresponds to the quantitative findings of this study, which were already presented in Table 3 and the dendrogram in Figure 6 but without mirroring all nuances of similarities and differences. In Map 3, however, each oblast is represented according to both its own specific sociolinguistic profile and its cluster membership, which is mirrored by the colouring. Of course, the contrasts in this map are certainly not as clear as

Map 3: The internal differentiation of the four areal clusters

in those, for example, from KIIS (2003) or those that could easily be drawn from the studies by Razumkov Centre (2016) or KIIS (2019). This is mainly because the latter proceed from a historical-political spatial division and then calculate averages for three of the five large regions. Consequently, the regionally large differences, especially in the highly heterogenous central area, are glossed over and the large regions are only apparently more clearly differentiated.

We took a different approach. As mentioned above, for each oblast we calculated a weighted measure of the stated usage of each code in everyday life. These three values for each oblast formed the input for the cluster analysis. The sociolinguistic spatial clusters of the oblasts and the corresponding Map 3 are thus directly based on analysis, and not on "a priori" shaped sub-areas of the Ukraine. Figure 7 illustrates the 42 values on which the cluster analysis and Map 3 are based. These results are only presented here because now we can order the oblasts according to their similarity and distance relations since these were the output of the cluster analysis.

The values determine the (continuous) proportion of blue (cyan) for Ukrainian, yellow for Suržyk and red (magenta) for Russian in the colouring of the oblasts in Map 3. The clearest blue tone (Ukrainian) can be found in cluster I, most clearly in Xmel'nyc'kyj and Čerkasy, the latter acquiring a slightly lighter shade of blue, due to a somewhat decreasing weight of Ukrainian. The oblasts of cluster II tend more or less clearly towards an achromatic grey, most distinctly in the oblasts of

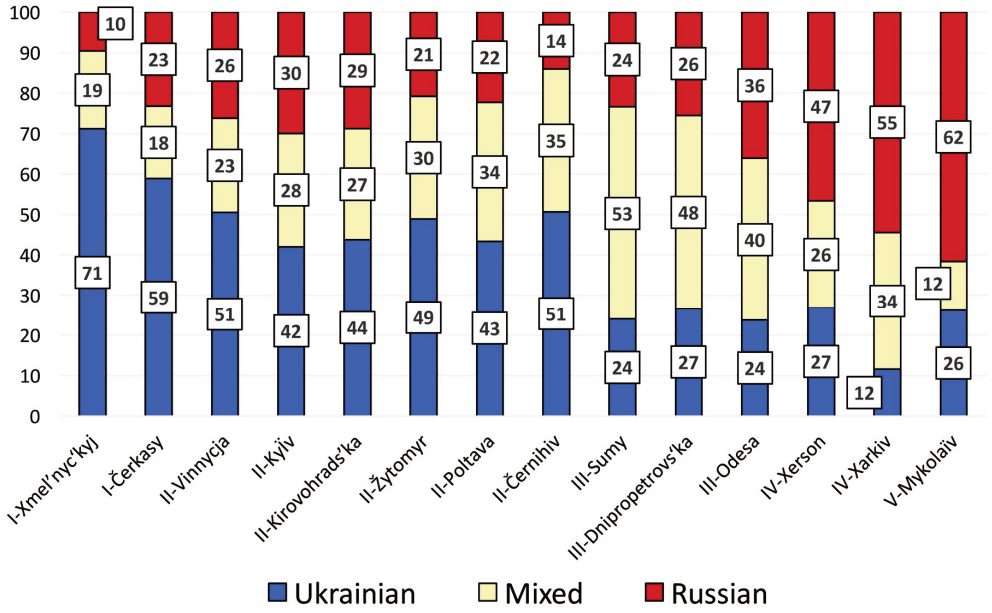


Figure 7: The "weight" of the three codes in clusters and oblasts on a scale from 0 to 100 points

Kyïv and Kirovohrads'ka. This indicates a fairly balanced strength of all three codes. Vinnycja and Žytomyr exhibit a slight blue tone mirroring resemblance to neighbouring Xmel'nyc'kyj and Čerkasy. Černihiv in this cluster takes a slight green colouring, due to a greatly reduced strength of Russian compared with the other oblasts in this cluster. The oblasts in cluster III exhibit tones of ochre, due to a strong share of Suržyk. Odesa here turns into a brown tone on grounds of a higher share of Russian. Due to the general dominance of Russian in the three oblasts of clusters IV and V, these have a clearly red tone. Compared with the oblast of Xarkiv in light red, almost turning into pink, Xerson (both IV) adopts a slight violet tone, due to a clearly doubled share of Ukrainian. Mykolaïv (V) is indeed violet since Suržyk is weak here and Ukrainian twice as strong.

4. Conclusion

Winston Churchill is attributed with saying that he only trusted statistics that he had falsified himself – a false attribution to the former Prime Minister of Great Britain.[33] Nevertheless, a profound scepticism towards statistical analyses as proposed here is widespread among scholars who themselves do not apply quantitative methods and extends to studies on the linguistic situation in Ukraine and Belarus (cf. Moser 2016). At least, the "cited" exaggeration conceals the often ne-

[33] We are grateful to an anonymous reviewer for drawing our attention to: https://falschzitate.blogspot.com/2017/09/ich-trauekeiner-statistik-die-ich.html (accessed: 30 Nov 2020).

glected aspect that quantification of observations depends heavily on the premises and methods of data collection and statistical analysis that – if they are explicitly stated – should be taken into account when interpreting the results. The sociolinguistic model of everyday multicodalism in Ukraine that is proposed here does not, at least in its cartography, yield a "clear and rounded" picture of larger subareas. Other surveys achieve this by proceeding from historical-politically-derived structurations in order to then calculate average frequencies for the codes (usually primary codes, if not "native languages") without considering the substantial differentiation within these areas. The historical-political background is nevertheless reflected in our analysis, too, as Map 3 indicates. The spatial clusters IV and V are part of the south and the east (in the broad sense), that differ from the rest of Ukraine in all surveys (including KIIS 2003, cf. Map 1). In any case, the area covered by clusters IV and V corresponds largely to the area of Ukraine where Slavic settlements from both Ukrainian and Russian territory were established relatively late. In this area, we also found Russian to be particularly strong but, compared with the other investigations mentioned above, considerably diminished when taking weighted multicodalism into account. However, Odesa from the south groups together with Dnipropertovs'ka oblast and Sumy into cluster III, when (weighted) multicodalism is considered. The clusters III, IV and V cover the vast majority of the territory that is grouped by KIIS (2003) outside the "centre", i.e. in the east (including an "eastern centre") or south of the Ukraine, together with Poltava and Černihiv. The latter two, however, group together with the central regions further westward when – as in our analysis – not concentrating on the primary language and not ignoring Suržyk.

This territory, our clusters I and II, is heterogenous.[34] This is not surprising. On the one hand, from the mid-seventeenth century and especially at the end of the eighteenth century, rule from Moscow (intermittently from St. Petersburg) pushed ever further westward into regions of permanent Ukrainian settlement; the westernmost region around L'viv fell under the rule of Moscow only from 1939–1941 to 1945 and then for nearly half a century. On the other hand, the reasons for the differently weighted mono- and multicodalism among speakers are of course not only historical-political. The respondents' sociobiographical characteristics and the correlating sociodemographic composition of the oblasts also play a considerable

[34] It is even more heterogenous than in Hentschel & Taranenko (2015). Its internal structure is slightly different from that in the study six years ago, even though both studies basically have the same database for the central region. The method chosen at that time was less sensitive in the sense that the multicodalism was observed, but only on the basis of the usage levels of "always" and "often". In particular, the more recent study confirms the relations between the "old" regions A to C (now clusters I and II) on the one hand and the oblasts in D (here cluster III, without Odesa, which was not considered in 2015) on the other and, in general, the heterogeneity of the central Ukrainian region.

role. Among academics, for example, a marked orientation towards the two national languages in everyday life, or at least one of them, is more likely in comparison with other respondents. Correspondingly, the proportion of "pure" Suržyk-speaking academics is likely to be much smaller. The rural-urban divide and, cross-cutting it, the level of industrialisation also play a part. Hentschel & Zeller (2017) have already described the relationships between respondents' language orientation and such sociobiographical characteristics for central Ukraine. However, not only sociobiographical characteristics in the narrow sense are important, but also political attitudes. Using the data from central Ukraine, Hentschel & Zeller (2016: 652–657) were able to identify four types of attitudes towards Suržyk that will influence whether someone practises Suržyk and whether they will admit to it. More general political attitudes, e.g. towards Russia's policy regarding the Crimea and Donbas, can have an effect on more Russian-speaking Ukrainian citizens in terms of their usage of Russian, or at least in terms of statements about their frequency of using Russian. Such dependencies and the associated differentiations within and between the clusters will be investigated in a further study.

The map of Ukrainian-Russian multilingualism or multicodalism is more colourful and complex if we base our models and maps on the multicodalism of individuals. The contrast between large regions such as the west, the centre (perhaps differentiated into a western and an eastern part) and the south of course do not disappear, but they are weakened considerably. As opposed to coarse contrasts between large regions that ultimately suggest a spatial-complementary dichotomous opposition Ukraine/Ukrainian vs. Russia/Russian, which some scholars seemingly prefer (e.g. Moser 2016), the model proposed here offers a gradual differentiation, a continuum of sociolinguistic nature, in which the multicodalism of the individuals, not least with the inclusion of Suržyk, creates a far stronger linguistic cohesion of Ukraine.

References

Bilaniuk, Laada. 2018. Purism and Pluralism. Language use trends in popular culture in Ukraine since independence. In: *Harvard Ukrainian Studies* 35, no 1–4 (2017–2018), 293–309.

Bühl, Achim. 2019. *SPSS – Einführung in die moderne Datenanalyse ab SPSS 25*. 16. Auflage. Hallbergmoos.

Danylenko, Andrii & Naienko, Halyna. 2019. Linguistic russification in Russian Ukraine: Languages, imperial models, and policies. *Russian Linguistics* 43. 19–39.

Flier, Michael. 2000. The rules of engagement. *Harvard Ukrainian Studies* 22. 113–136.

Hentschel, Gerd. 2017. Eleven questions and answers about Belarusian-Russian mixed speech ("Trasyanka"). *Russian Linguistics* 41(1). 17–42.

Hentschel, Gerd. 2018. Die weißrussische „Trasjanka“ und der ukrainische „Suržyk“: Grundlegende quantitativ-qualitative soziolinguistische Ähnlichkeiten und Unterschiede. In: Kempgen, Sebastian & Wingender, Monika & Udolph, Ludger (eds.), *Deutsche Beiträge zum XVI. Internationalen Slavistenkongress, Belgrad 2018*, 127–138. München.

Hentschel, Gerd & Kittel, Bernhard. 2011. Weißrussische Dreisprachigkeit? Zur sprachlichen Situation in Weißrussland auf der Basis von Urteilen von Weißrussen über die Verbreitung „ihrer Sprachen“ im Lande. *Wiener Slawistischer Almanach* 67. 107–135.

Hentschel, Gerd & Taranenko, Oleksandr & Zaprudski, Siarhej (eds.). 2014. *Trasjanka und Suržyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* Frankfurt am Main.

Hentschel, Gerd & Taranenko, Oleksandr. 2015. Die Sprachenlandschaft der zentralen Ukraine: Ukrainisch, Russisch, ‘Suržyk’. Verwendung – Kompetenz – nationale Positionierung. *Die Welt der Slaven* LX. 248–275.

Hentschel, Gerd & Zeller, Jan Patrick. 2016. Meinungen und Einstellungen zu Sprachen und Kodes in zentralen Regionen der Ukraine. *Zeitschrift für Slawistik* 61(4). 636–661.

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

KIIS. 2003. *Languages in Ukraine.* https://ru.wikipedia.org/wiki/File:Languages_in_Ukraine.PNG (accessed 12 Aug 2020).

KIIS. 2019. *Dumky i pohljady naselennja ščodo rosijs'koï movy v ukraïns'komovnyx školax i nadannja nepidkontrol'nym terytorijam Donbasu avtonomiï u skladi Ukraïny.* https://kiis.com.ua/?lang=ukr&cat=reports&id=832 (accessed 12 Aug 2020).

KIIS. 2020. *Koho ukraïnci xotily b bačyty predstojatlem ob'jednanoï cerkvy.* http://kiis.com.ua/?lang=ukr&cat=reports&id=958 (accessed 12 Aug 2020).

Kulyk, Volodymyr. 2010. Gespaltene Zungen: Sprache und Sprachenpolitik in der Ukraine. *Osteuropa* 60(2–4). 391–402.

Kuroxtina, Tat'jana N. 2009. Suržyk v internete. In: *Vestnik Moskovskogo Universiteta: Serija 9 (Filologija)* 2. 101–112.

Menzel, Thomas & Hentschel, Gerd. 2017. *Flexionsmorphologische Irregularität im innerslavischen Sprachkontakt: Sprachinhärente Präferenzen oder politisch-soziale Dominanz. Russisch vs. Weißrussisch/Ukrainisch – Polnisch vs. Kaschubisch/Lemkisch*. Frankfurt am Main.

Moser, Michael. 2016. “Suržyk” čy “Suržyky”. *Ukraïns'ka mova* 2016(1). 27–54.

Razumkov Centre. 2016a. Identyčnist' hromadjan Ukraïny v novyx umovax: Stan, tendenciï, rehional'ni osoblyvosti. *Nacional'na bezpeka i oborona* 2016(3–4)/161–162. 2–104.

Razumkov Centre. 2016b. Konsolidacija ukraïn'skoho suspil'stva: Vyklyky, možlyvosti, šljaxi. *Nacional'na bezpeka i oborona* 2016(7–8)/165–166. 2–97.

Stavyc'ka, Lesja. 2014. Ein Blutschandekind der Postmoderne. In: Hentschel, Gerd & Taranenko, Oleksandr & Zaprudski, Siarhej (eds.), *Trasjanka und Suržyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?*, 351–374. Frankfurt am Main.

Ševčuk-Kljuževa, Ol'ha. 2016. *Suržyk v Internet-prostori (na materiali social'nyx merež Vkontakte, Facebook, Instagram)*. Kyïv. https://er.nau.edu.ua/handle/NAU/19171 (accessed 22 Jul 2020).

Taranenko, Oleksandr. 2007. Ukrainian and Russian in contact: Attraction and estrangement. *International Journal of the Sociology of Language* 183. 119–140.

Taranenko, Oleksandr. 2013. Variantnist' vs. stabil'nist' u strukturi ukraïns'ko-rosijs'koho 'suržyku' (URS): Sukupnist' idiolektiv vs. sociolekt. In: Gerd Hentschel (ed.), *Variation und Stabilität in Kontaktvarietäten: Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien* (Studia Slavica Oldenburgensia, 21), 27–62. Oldenburg.

Taranenko, Oleksander. 2014. Ukrainisch-russischer Suržyk: Status, Bewertungen, Tendenzen, Prognosen. In: Hentschel, Gerd & Taranenko, Oleksandr & Zaprudski, Siarhej (eds.), *Trasjanka und Suržyk – gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?*, 265–288. Frankfurt am Main.

Zeller, Jan Patrick. 2018. Zum ukrainisch-russischen Sprachkontakt: Phonische Variation im ukrainischen „Suržyk" im Vergleich mit der weißrussischen „Trasjanka". In: Kempgen, Sebastian & Wingender, Monika & Udolph, Ludger (eds.), *Deutsche Beiträge zum XVI. Internationalen Slavistenkongress, Belgrad 2018*, 365–374. München.