

This special issue is dedicated to the children of Ukraine who have lost their lives or suffer physical and psychological trauma as a result of the war of aggression waged by the Russian Federation against Ukraine.

Introduction: background, aims, source material, and brief outlook on this special issue on Ukrainian Suržyk

This special issue of *Russian Linguistics* contains six articles on the mixed Ukrainian-Russian variety Suržyk. Five of them are in-depth analyses presenting the central findings of the research project *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*. The project was funded as a joint German-Austrian initiative under the D-A-CH program of the German Research Foundation (DFG), project no. 419468937 (headed by Gerd Hentschel, University of Oldenburg), and the Austrian Science Fund (FWF), project no. I 4189-G30 (headed by Tilmann Reuther, University of Klagenfurt). The abovementioned five articles are contributions by the Oldenburg research team¹ and are characterized methodologically by a corpus-based, quantitative approach.

The sixth article by our Ukrainian partner and colleague Tetiana Kuznietsova investigates the discourse and polemics surrounding *Suržyk* as reflected (primarily) in Ukrainian media (for more details, see below). This last study was made possible by two scholarships from the Volkswagen-Stiftung (project no. 9B893 / 9D158), awarded to war refugees from Ukraine, which enabled Tetiana Kuznietsova to spend two years at the Department of Slavic Studies at the University of Oldenburg.

It is not necessary here to provide a detailed introduction to the historical and sociopolitical background of the development of *Suržyk*. This is offered by Taranenko (2007), Hentschel and Taranenko (2021),² and in one of the papers by Gerd Hentschel (forthcoming a), in this special issue, which all include further references. There is also no need to explicitly outline the state of research here. This is addressed in part again in the last two papers just mentioned and also in this special issue, but from a different point of view, by Tetiana Kuznietsova (forthcoming). The other contributors to this volume also refer to relevant prior research in their respective contexts. For the state of research up to the first half of the last decade, see the anthology edited by Hentschel et al. (2014).

A key focus of this introduction is to provide an outline of the source material for the *Oldenburg Corpus of Ukrainian Suržyk* (hereafter OCUS) that plays a central role in the results presented throughout the special issue. The outline is also presented here to avoid repeating this information in

¹ The publications by the Klagenfurt partner are available here:
<https://tilmannreuther.academia.edu/research#papers>.

² This is the first publication from the above-mentioned project that utilizes the sociobiographical material collected therein.

each of the five articles that analyze the corpus. Nevertheless, only the most essential information is provided here regarding the data collection, corpus structure, and the respondents. A comprehensive and detailed documentation of the Oldenburg corpus of Ukrainian Suržyk, along with sociobiographical data, is available in Hentschel et al. (2026), hereafter OCUS-Doc, at <https://uol.de/slavistik/forschung/sprachwissenschaft/bereich-mischvarietaeten/okus-dok>.

The joint German-Austrian research project focused exclusively on the Ukrainian territory north of the Black Sea, i.e., the oblasts of Odesa, Mykolaïv, and Xerson.³ However, other material in the corpus covers a much broader region of Ukraine, incorporating data collected from central oblasts during two previous projects by the Oldenburg team, supported by Ukrainian colleagues. This material is also included in the corpus analyses in this special issue. The territorial scope of the three projects is reflected in the map presented in Hentschel and Taranenko (2021):



Map *The strength of the three codes in the oblasts* (Source: reproduced from Hentschel and Taranenko, 2021; the Roman numerals refer to the results of a cluster analysis)

Just a brief note on the map: the coloring of the oblasts reflects the extent of the presence of the three codes—Ukrainian, Russian, and Suržyk—in people’s everyday lives. The stronger the presence of Ukrainian in an oblast, the more the color shifts toward blue, which is blended into purple by red for Russian and into green by yellow for Suržyk. A gray shade indicates a certain balance between the three codes. The linguistic situation in oblasts assigned to the same cluster is more similar than in oblasts assigned to the other clusters. These data derive from responses provided by approximately

³ It was originally planned to include Zaporizžja oblast. However, due to the effects of the armed conflict in the neighboring Donetsk oblast (since 2014), data could not be collected there.

1,400 participants in the central regions and approximately 1,200 along the Black Sea coast, collected via closed-format surveys (i.e., written questionnaires⁴).

Whereas the collection of linguistic data in the three Black Sea oblasts took place in 2021 and to a lesser extent in 2020, the area north of it was already covered in earlier research projects led by Gerd Hentschel. First, in 2010 and 2011, family conversations in all these 11 oblasts were recorded and transcribed as part of the project *Inflectional Morphological Irregularities in “Contemporary” Contact Varieties of North Slavic Languages*, funded by the DFG, project no. 155014374. This project had a dual focus: firstly, on structural phenomena of language contact and mixing (and not on social aspects), and secondly, on other contact configurations in North Slavic (cf. Menzel & Hentschel, 2017).⁵ It was followed by another project in the same region: *Variability and Stability in a Mixed Substandard in Extensive and Time-Stable Language Contact: Ukrainian Suržyk between Ukrainian and Russian (in comparison with Belarusian Trasjanka)*, funded by the Fritz Thyssen Foundation, project no. 10.14.1.066. The data were collected in the first months of 2015, by recording semi-structured in-depth interviews. These interviews particularly explored the relationship between Ukrainian and Russian in Ukraine as well as both languages’ relationship with Suržyk while also probing respondents’ attitudes toward the three codes and their usage. The most recent project on the Black Sea coast likewise involved the collection of spoken language data from both family conversations and in-depth interviews, though in this case the interviews laid a greater focus on respondents’ language biographies. In both of the latter projects, in addition to structural phenomena, significant attention was paid to the social or sociobiographical background of the respondents and its impact on usage and attitudes (cf. Hentschel & Zeller, 2016, 2017; Hentschel & Taranenko, 2015, 2021; Hentschel & Palinska, 2022a; Zeller & Hentschel, 2024).

The entire corpus underpinning the analyses in this special issue contains roughly 730,000 word forms (tokens), about half of which come from family conversations, while the other half derive from interviews. We therefore speak of the *family corpus* and the *interview corpus*. Speech in the family conversations is certainly more spontaneous and thus more uncontrolled than in the interviews. However, the frequent use of Suržyk in the latter also demonstrates a high degree of informality in respondents’ speech. Passages of family conversations and, especially, interviews that hardly

⁴ The questionnaires can be accessed at <https://uol.de/slavistik/forschung/sprachwissenschaft/drittmittelprojekte/variabilitaet-und-stabilitaet-im-gemischten-substandard-surzyk> for the central region and at <https://uol.de/slavistik/forschung/sprachwissenschaft/hybridisierung-von-zwei-seiten> for the three oblasts on the Black Sea coast. They were completed by professional interviewers based on the respondents’ answers. In both cases, the closed-format interviews were organized and conducted by the company SOCIS, Kyiv, under the direction of Mykola Čurylov.

⁵ Additionally, the studies addressed Trasjanka as a Belarusian-Russian mixed variety, as well as the influence of Polish on Kashubian and Lemko.

contained any Suržyk speech, but only Ukrainian or (seldom) Russian (especially at the beginning of interviews), have not been included in the corpus (see below). The interviewers were instructed to speak informally themselves. In several analyses, this contrast in spoken genre is modeled as a potentially influential independent variable, possibly affecting the variation between Ukrainian and Russian elements and is thus statistically controlled for.

The corpus is thus heterogeneous, not only geographically and situationally, but also chronologically. These differences are considered in the quantitative analyses and the interpretation of their results. Of course, the situation in Ukraine, including the linguistic situation, had already changed considerably even before February 2022. To certain extent, this might have influenced linguistic patterns in Suržyk. However, due to its considerable degree of stabilization (see the five analytic articles in this volume) it is not to be expected that these changes have led to fundamental differences in Suržyk itself.

The corpus structure is essentially fourfold: (i) *Central-Family*, (ii) *Central-Interview*, (iii) *Black Sea-Family*, (iv) *Black Sea-Interview*. All four parts are of comparable size: (i) 172,322 tokens, (ii) 168,353, (iii) 199,846, (iv) 188,587.

The conceptual orientation of the projects targeting, on the one hand, central Ukraine (sub-corpora (i) and (ii)) and, on the other hand, the three Black Sea oblasts (sub-corpora (iii) and (iv)) differed in one important respect. The explicit object of study of the former was traditional, Ukrainian-based Suržyk. This variant of Suržyk had been discussed before in several articles and two monographs (cf. the contributions in Hentschel et al. 2014 and the bibliography therein). The materials collected in sub-corpora (i) and (ii) largely reflect this conceptual orientation. This orientation on traditional Suržyk in the corresponding two projects was motivated by the known fact that in the vast majority of the eleven oblasts the social position of Ukrainian (including its dialectal forms) in everyday life is relatively strong. The only clear exception is Xarkiv oblast (cf. Hentschel & Taranenko, 2021 and the map above). The project targeting the Black Sea region sought to capture both Ukrainian-based Suržyk and a potentially emergent Russian-based variant that may have developed in independent Ukraine after 1991. The inclusion of a possible Russian-based Suržyk was motivated by the fact that Russian is much more prevalent in everyday life along the Black Sea coast (cf. Hentschel & Taranenko, 2021). However, to anticipate the findings, the results for the latter in sub-corpora (iii) and (iv) were, unfortunately, disappointing (more details below in the description of how respondents were recruited). The directly following sections of this introduction provide basic background information on the two variants of Suržyk, whether attested or hypothetical.

The emergence of this Ukrainian-based Suržyk was initially driven by processes of approximation as large parts of the originally Ukrainian-speaking rural population migrated into cities where Russian predominated, particularly during the last 150–200 years of industrialization and urbanization (cf. Taranenko, 2007). However, the Ukrainian-based Suržyk soon developed “a life of its own”, as it was

passed down in families to the following generations. The contact with the two donor languages, of course, has remained, with a retreating impact of Russian at least in the last decade. In this variant, one donor language, Russian—being the language of power during its emergence—functions as a superstrate, while Ukrainian, as the basal source language, serves as the substrate.⁶

Around the turn of the millennium, the hypothesis emerged in linguistic research that a Russian-based Suržyk could develop in independent Ukraine, i.e., after 1991. Successive governments of independent Ukraine undoubtedly promoted Ukrainian, though with different intensity; thus, the Russian-speaking parts of the Ukrainian population were to some extent forced to shift toward Ukrainian, especially in public spheres like education—as far as they were able to. However, at least until the full-scale invasion of Ukraine by the Russian Federation and the associated atrocities committed by their troops in occupied territories, there was no official push in Ukraine to abandon Russian (cf. Hentschel & Brüggemann, 2015). Nevertheless, especially in American Ukrainian studies, the notion arose that Russian speakers might approximate Ukrainian, possibly leading to the emergence of a Russian-based Suržyk (cf. Bilaniuk, 2004, 2005; Flier, 2008), often called “Neo-Suržyk” in this special issue.

The most recent project informing this special issue was designed to investigate the potential coexistence of the old Ukrainian-based Suržyk and a potentially emerging, new Russian-based Neo-Suržyk. (For this reason, the project’s short title is *Hybridization from Two Sides?*) In the three oblasts on the Black Sea, the probability of encountering Russian-based Suržyk was hypothetically much higher. This territory, broadly settled only from the 19th century onward, was shown in a Tsarist-era census from around 1900 to have a dominant use of Russian (and Yiddish) in urban areas, and Ukrainian in rural areas (cf. Hentschel & Palinska, 2022a). Settlement patterns brought dialect-speaking Ukrainians from many regions together, likely producing a wide range of mixed local dialects that did not have the centuries-long social roots found in large parts of the center. In the 20th century, Russian spread further throughout the region (cf. the map above). In this area, unlike most of the center (see above), the proportion of the population that primarily speaks Russian is known to be far higher (cf. Hentschel & Taranenko, 2021).⁷ These conditions provided reason to expect that, if anywhere, both Ukrainian-based Suržyk (especially in more rural areas) and Russian-based Suržyk (perhaps in urban settings) might be recorded here.

Thus, the existence of the second variant—*Neo-Suržyk*—still needed to be verified. However, the initiators of the Black Sea project, Hentschel and Reuther (2020), already voiced doubts that a

⁶ One may even differentiate the role of Ukrainian in that dialectal Ukrainian can undoubtedly be characterized as a substrate, while the Ukrainian standard language functions as an adstrate, especially in the early phase of its development at the end of the 19th century and in the first decades of the 20th century. However, this differentiation is irrelevant here.

⁷ Of course, there are other oblasts along the eastern border with the Russian Federation that show a similar linguistic situation, such as the Donbas area and Zaporizžja. The three Black Sea oblasts, however, saw little to no armed aggression from the Russian Federation prior to 2022 (with the exception of the Crimean Peninsula), which made data collection feasible.

Russian-based Suržyk could develop into a comparably stable and distinguishable variant within the roughly 35 years of Ukraine's independence—an admittedly short time compared to the more than a century long development of Ukrainian-based Suržyk. As Thomason and Kaufman (1988) argue, mixing or merging phenomena between genetically closely related and structurally highly similar languages resemble analogous phenomena in dialect contact, i.e., dialect levelling. Furthermore, Trudgill (1986) assumes a time span of at least three generations for a new more or less stable dialect to occur in a constellation of dialect contact: much longer than 35 years.

An empirical answer to the question of a Russian-based Suržyk depends, first of all, on the meaning we attribute to the term *Suržyk* in our metalanguage. If we use it to refer to any form of mixing Ukrainian elements into fundamentally Russian speech, then a Russian-based Suržyk undoubtedly exists, in the sense of sporadic interlinguistic interference. Even with regard to Ukrainian-based Suržyk, some scholars still claim that it is merely a spontaneous, irregular—if not chaotic—or at least individually variable form of mixing for each of its speakers (cf. Masenko, 2014; Moser, 2016). However, about fifteen years ago, an opposing view emerged. For example, del Gaudio (2010) and Šumarova (2014) posit a considerable degree of regularization or complementary stabilization of Ukrainian and Russian “ingredients” competing within Ukrainian-based Suržyk. This would mean that the latter is developing into a fused lect in the sense of Auer (1999). Of course, it is more than probable that, in spite of such an emergent fusion, the Ukrainian-based variant of Suržyk may be accompanied by spontaneous mixing phenomena in speech, as is normally the case as long as both donor languages are in use in society, again according to Auer (1999).

These earlier studies by del Gaudio, Šumarova, and others, however, are based on informal observation, and their empirical basis remains unclear.⁸ To the best of our knowledge, no studies based even on informal observation have been conducted to date on a potential Russian-based Suržyk. When we address the latter in this special issue (especially in Hentschel, forthcoming a), we refer only to a (potential) variant that would exhibit at least some degree of stabilization.

Hentschel and Reuther (2020) further hypothesized that, if such a Russian-based Suržyk exists, the systemic differences from a Ukrainian-based variant may be gradual. Even if the former does not exist, there may be sub-variants of Ukrainian-based Suržyk that are structurally closer to Russian than others. Hentschel's (2024) study of the Suržyk lexicon, based on the same corpus (OCUS) as the analytical studies in this special issue, shows, first, that in the competition between corresponding Ukrainian and Russian lexical items (“translational equivalents”), the distribution has already stabilized for some items in favor of the former and for others in favor of the latter. For a third group, where variation can be observed, preferences broadly follow the gradual strength of Ukrainian (or,

⁸ The most comprehensive monographic study by del Gaudio (2010) at least provides the information that the observations were made in four oblasts belonging to the area covered by the first two projects underlying this special issue (cf. Hentschel, forthcoming a).

conversely, Russian) across the regions illustrated in the map above. Earlier, Hentschel and Palinska (2022b), working with this data set as well, identified regionally differentiated preferences for competing infinitival endings. In this case, however, the relatively stable distribution results from an interaction between Ukrainian dialectal features and the presence of Ukrainian or Russian in everyday use. The last two papers propose viewing Suržyk as a “mesolectal continuum” with many internal isoglosses, which often do not converge, as in dialect continua.

The sub-corpora (i) and (ii) for central Ukraine differ from those for the Black Sea coast (iii) and (iv) not only in their “target objects” (traditional Suržyk vs. traditional Suržyk and Neo-Suržyk), but partly also in the principles underlying respondent recruitment. A fundamental problem, irrespective of the sub-corpus, is how to identify suitable Suržyk speakers. Since no document or register records whether a person speaks Suržyk, random sampling is excluded from the outset.

The first project focused on Ukrainian-based Suržyk as used in family communication in central regions (i – *Central-Family*). The families were selected by O. Taranenko (Institute of Linguistics, National Academy of Sciences of Ukraine, Kyiv) on the basis of his assessment that they practice strongly mixed Ukrainian-Russian speech that corresponds to a widely prevalent type of mixed speech that he and the members of his team consider the Suržyk that had already been discussed in the years before. Eight families were chosen from eight of eleven oblasts. Preselection relied on suggestions from colleagues, and sample recordings were evaluated (see OCUS-Doc for details). Taranenko has studied the linguistic situation in Ukraine—and Suržyk in particular—from a sociolinguistic perspective since the late 1990s (cf. Hentschel et al. 2014, pp. 391–392). In other words, as in earlier informal studies, the data are based on speakers whom the researchers themselves, on the basis of long-term experience in everyday life in Ukraine, classify as Suržyk speakers. The targeted selection of respondents corresponds to the aim of the first project, namely a basic description of (here inflectional-morphological) structures of typical Ukrainian-based Suržyk. This is also reflected in the limited sociobiographical metadata: none of the 85 respondents identified Russian as their childhood language; instead, Ukrainian or Suržyk, or both codes, were reported. At the time of data collection, Russian was regularly used by only seven respondents, in addition to Ukrainian or Suržyk or both.

The second project, from which the sub-corpus *Central-Interview* (ii) derives, followed the same principle, with respondent selection again carried out by O. Taranenko. The preselection in the eleven oblasts, however, differed. In a survey of 1,400 randomly selected respondents, data were collected on the frequency of use of Ukrainian, Russian, and Suržyk (cf. Hentschel & Taranenko, 2015; Hentschel & Zeller, 2016, 2017). Those reporting frequent use of a linguistic “mixture with Ukrainian and

Russian words”⁹ and consenting to further participation were selected for in-depth interviews, yielding 68 respondents (see below for their linguistic characterization).

The recruitment procedure for the third project on the Black Sea coast, i.e., for (iii) *Black Sea-Family* and (iv) *Black Sea-Interview*, also followed this approach. Preselection of interviewees, and in this case also of families, was to be based on an analogous survey of 1,200 respondents. The questionnaire used in the second project served as the basis, with minor revisions. The closed-format interviews were conducted in February 2020. 151 respondents stated, first, that they use the mixed code often, if not very often, and, second, that they were willing to meet again for a longer conversation about linguistic issues in their lives and in Ukraine, possibly also in a family setting. The first condition was important in order to avoid unnecessary costs for speakers who use the mixed code only occasionally and are therefore likely to converse with the interviewer in Ukrainian or Russian, whether in the standard varieties or in more colloquial forms, which could include dialectal traits, too.

Shortly thereafter, the COVID-19 pandemic broke out, seriously jeopardizing fieldwork. When contacted again in the following weeks, only five of the 151 potential candidates agreed to meet with an interviewer previously unknown to them for an in-depth interview. Moreover, it proved impossible to invite the families of some of these candidates to have their conversations recorded with the candidate participating. In such cases, the candidate would not have been considered separately for the in-depth interviews.

New participants were extremely difficult to recruit, resulting in delays of several months. Special thanks are due to our collaborator and co-author Tetiana Kuznietsova (at the time based at the National University “Odesa Law Academy”) and our team member Olesya Palinska, without whom data collection under such adverse conditions would not have been possible. We reverted to a procedure similar to that used in the second project (Interviews in the Centre): respondents were preselected on the basis of individual recommendations by Tetiana Kuznietsova’s colleagues, her advanced students, and other contacts—for all three oblasts on the Black Sea coast. More than one hundred volunteers were recommended. They all had to fill out the same questionnaire that also underpinned the closed-format interviews just mentioned above. Thus, we obtained information on potential informants’ self-estimations. A total of 99 candidates were preselected for in-depth interviews, including the five remaining candidates from the first recruitment attempt. In addition, 15 families (five in each of the three oblasts) preselected by the Odesa team agreed to have their conversations recorded. They filled out the questionnaire as well, with the exception of sporadic participants who did not produce more

⁹ Note that in the forms to be filled out by the interviewers for the interviewees, we never used the term *Suržyk*. This mixed code and its speakers were highly stigmatized by many representatives of the political, cultural, media, and academic elites even before the Russian annexation of the Crimean Peninsula, the more localized military conflict in the Donbas region, and thus before the full-scale invasion of Ukraine by the Russian Federation (cf. Stavyc’ka, 2014; Kuznietsova, forthcoming).

than a couple of utterances (sentences), mostly very short ones (often young children). Here, in the family context, the questionnaire was completed after the conversation had been recorded.

The questionnaires of the 99 candidates for the interviews already revealed that the number for whom Russian was the language or, more precisely, one of the codes regularly used in their families during their childhood was very low: 17 of 99. Only eight of the 17 declared that Russian was the sole or dominant code in their families during their childhood. These eight respondents, of course, were the most probable speakers of a (potential) Russian-based Suržyk. However, it has to be taken into account that this is not necessarily the case if a Ukrainian Suržyk was widely in use, and maybe Ukrainian as well, in their broader contact situations (e.g., when speaking with grandparents in the countryside, peers and friends in the neighborhood, etc.). Be that as it may: already at this early (though heavily delayed) stage of the fieldwork it was becoming clear that the material for a Russian-based Suržyk (again: if it exists at all) would be at best relatively sparse.

The final selection was carried out by Tetiana Kuznietsova and Olesya Palinska. Tetiana Kuznietsova, stemming from the Northern area around Sumy on the Russian border, where Suržyk according to Hentschel and Taranenko (2021) is especially widespread, was among the first scholars in independent Ukraine who concerned themselves with the issue of Suržyk (Kuznietsova, 1999, 2001a, 2001b). Her later career brought her to Kyïv and Odesa, and she remained in both for many years. This means that she is personally acquainted with mixed Ukrainian-Russian speech in different parts of the country, all of which belong to the area where data was collected for OCUS. Olesya Palinska spent most of her life in L'viv. Both Kuznietsova and Palinska are bilingual, Ukrainian and Russian. The task was to select those candidates that use Ukrainian as well as Russian elements in their speech to a considerable degree. Which of the two left the clearer traces in the mixed speech was not principally relevant.

The two colleagues selected 70 out of the 99 recorded candidates to be transcribed for OCUS, i.e., for linguistic analysis.¹⁰ The non-selected candidates tended not to practice a clearly mixed speech, but rather either (colloquial) Standard Ukrainian or Standard Russian, with sporadic, as a rule lexical, interferences of the other language.

At this point, the codes used by the interviewers in the in-depth interviews in both areas, the center and the Black Sea coast, should be mentioned. As has been noted above, interviewers were asked to adapt linguistically to the interviewees. This aim was largely achieved. In the Center, interviews were conducted primarily in Suržyk: out of 68 interviews, interviewers spoke Suržyk in 36 cases, and in a further 14 they used Ukrainian alternating with Suržyk fragments. Only one interview (from Sumy

¹⁰ Note that all the material of all 96 was described roughly for content analysis. This was done for the Klagenfurt research focus. The respondents' linguistic biographies were the central aspects of interest for the colleagues in Austria.

oblast) was conducted in Russian, while the remaining 17 were conducted in Ukrainian, ranging from the standard literary language to, in most of these cases, fairly colloquial varieties.

Interviewers in the South also widely adapted to the speech of the respondents: of the 70 interviews included in the corresponding sub-corpus, interviewers spoke Suržyk in 28 cases, and in another 10 they switched between Ukrainian, Russian, and Suržyk in various combinations. In 15 interviews, questions were asked in Ukrainian (rather colloquial), and in 17 in Russian, mostly in its standard form.

As mentioned above, to the best of our knowledge there has been no research project to date with a controlled empirical component. Recruiting interviewers who were capable of adapting their speech to that of the interviewees was a similar problem to determining the latter, especially for the project on the Black Sea coast under COVID-19 conditions.

In this introduction we provide a rough overview (in Tables 1, 2, and 3) of the linguistic profiles of three groups of respondents: those comprising the sub-corpora (ii) Central-Interview and (iv) Black Sea-Interview as well as the participants in (ii) Black Sea-Family conversations.¹¹

Table 1 *Codes used frequently and regularly during childhood*

	Ukr	Sur	UkrSur	UkrRuss	SurRuss	UkrSurRuss	Russ	total
Centr-Int	26	31	4	0	0	0	7	68
BlSea-Fam	16	43	7	1	2	3	3	75
BlSea-Int	9	43	7	2	2	3	2	68
total	51	117	18	3	4	6	12	211

It suffices here to state that a total of only 25 respondents (of all those who answered this question) named Russian as the only, or one of two or three codes used in their families during their childhood. This is just over 10%. This underlines what has been said above: the expectation to find a Russian-based Suržyk in the linguistic materials of OCUS is low—to repeat: if it exists at all. The second important aspect to be mentioned about respondents’ linguistic background during their childhood concerns Suržyk: There can be no doubt that the Suržyk (to be precise: “the roughly balanced Ukrainian-Russian mixture”) that the respondents stated to have been in use in their families during

¹¹ The linguistic background of participants in (i) *Central-Family* was commented upon above, at least in an even shorter form due to the lack of differentiated sociobiographical information. - More details of the linguistic profile of respondents in the four sub-corpora are presented in some of the papers and especially in OCUS-Doc, together with a general sociobiographical description of the respondents that contributed to OCUS.

their childhood must have been a Ukrainian-based variant: Only 24 of the 211 respondents in Table 1 were younger than 20 years old, i.e., were born after 2000. The cited hypothesis of an emerging Russian-based Suržyk concerns the period after 1991, but definitely not the first years after the collapse of the Soviet Union. In spite of the fact that the political promotion of Ukrainian started soon after 1991, it is impossible that this had a sustainable influence on the linguistic preferences and codes of the majority of the population that used Russian as the preferred or only code during Soviet times.

Table 2 *Codes currently used frequently and regularly*

	Ukr	Sur	UkrSur	UkrRuss	SurRuss	UkrSurRuss	Russ	Total
Centr-Int	0	8	20	0	6	34	0	68
BISea-Fam	0	2	24	1	11	32	0	70
BISea-Int	0	3	18	4	8	35	0	68
total	0	13	62	5	25	101	0	206

Table 2 clearly shows that only a couple of respondents are “monocodal”, and if so, they named the “mixture”, i.e., Suržyk. While only roughly one in ten respondents stated that Russian belonged to their linguistic repertoire during childhood, two of three respondents did so for the time when the materials were being collected: a rise of roughly 55 percentage points. This means that Russian might have had a strong impact on the mixed code. On the other hand, roughly 4 out of 5 respondents named Ukrainian as one of their codes of frequent use. The higher importance of Ukrainian is confirmed by the answers to another question in the questionnaire: which code is used most of all.

Table 3 *The one code currently used the most*

	Ukr	Sur	Russ	total
Centr-Int	28	35	5	68
BISea-Fam	22	40	14	76
BISea-Int	21	34	13	68
total	71	109	32	212

The impact of the respondents’ linguistic background on their mixed speech, i.e., the linguistic phenomena at issue, will be discussed in more detail by several papers in this special issue.

To sum up the description of the respondents' linguistic profiles, it must be stated that on the one hand there is only a limited likelihood that the as yet not empirically attested Russian-based Suržyk can be found in the materials collected, at least to an extent that allows for a solid quantitative analysis. On the other hand, there is a considerable number of respondents for whom a frequent, if not very frequent use of Russian can be assumed.

In Tables 1, 2, and 3, the number of participants that answered the relevant questions varies between 206 and 212. Altogether, we have full sociobiographical information for 214 respondents. This of course does not mean that every respondent answered every question. For 85 respondents in the *Central-Family* sub-corpus there is only minimal sociobiographical information. No sociobiographical information is available for 17 respondents. Apart from two respondents whose questionnaires were lost, these are peripheral participants in family conversations (see above), for whom it was not necessary to record this information. Note that there were altogether 50 respondents that contributed fewer than 500 words, i.e., word forms (tokens). These persons, however, contributed only about 10,000 of the roughly 730,000 overall tokens in OCUS. There is a further group of 25 respondents who contributed more than 500 but less than 1,000 tokens; altogether they contributed roughly 20,000 tokens. In other words, the two groups contribute about 4% of the whole. In many analyses these groups are ignored when sociobiographical aspects have to be considered, in addition to a few others who did not answer the corresponding questions.

For the two sub-corpora *Central-Interview* and *Black Sea-Interview*, the goal was to ensure a balanced distribution according to certain sociobiographical parameters other than respondents' linguistic background: these were gender, age, and educational background. This aim has been achieved for the former sub-corpus, but it was not met to the same extent for the latter regarding the criteria of age and educational background. Again, the COVID-19 pandemic caused problems. Older individuals, in particular, were highly reluctant to participate over the entire "financeable" duration of fieldwork, especially in interviews. As a result, the Black Sea interview sub-corpus is disproportionately composed of data from younger people (up to about 30 years of age). Consequently, the proportion of respondents either still in higher education or who have completed it is also disproportionately high.

The less-than-ideal demographic composition of the respondent sample, particularly in terms of age and education, was the price to pay to prevent the Black Sea project's termination. The statistical methods applied in the analyses allow for these factors and the differences in the respondents' linguistic background to be controlled for.

A further difference between the central and the Black Sea sub-corpora is that in the former, no data were collected in villages. In neither region were data collected in metropolises (cities of around 1,000,000 or more inhabitants), due to their generally Russian-speaking populations. This applies to

four cities: Kyïv, Xarkiv, Dnipro, and Odesa. Although these cities in the meantime (may) have shifted toward Ukrainian to varying degrees, this shift is definitely incomplete. In a similar study on Belarusian Trasjanka, conducted by Gerd Hentschel and Bernhard Kittel in Belarus around 2010 (cf. Kittel et al., 2018), it proved virtually impossible to find willing interview participants in the capital Minsk within the time limits and financial resources available to the project. In the central Ukrainian regions, villages were excluded for a different reason: there was a consensus in earlier Suržyk research that the mixed variety was particularly widespread in mid-sized towns, while villagers tended to speak local dialects. To what extent these dialects are still in active use in the wide areas of central Ukraine today (the variance is likely to be considerable) and the extent to which they may themselves have been “Suržykized” (i.e., sustainably influenced by Russian) is largely unknown. The focus on localities with the best prospects for collecting language material in Suržyk was ultimately also driven by pragmatic and financial considerations. In contrast to the center, far more respondents per oblast had to be and could be recorded in the three oblasts along the Black Sea coast. This was necessary due to our intention to grasp the possible coexistence of the Ukrainian-based Suržyk and the potential Russian-based variant. In addition, the linguistic situation here and its socio- and contact linguistic history made the inclusion of villages worthwhile, i.e., the considerable differences between Russian- (and Yiddish-) speaking towns and Ukrainian-speaking villages at the turn from the 19th to the 20th century and the permanent immigration from Ukrainian- and Russian-speaking regions into all of the Black Sea oblasts.

On the basis of the data in OCUS, a controlled, empirically grounded description of central structural domains is presented in this special issue: A basic aim in all the analyses will be to determine to what extent there is variation between Ukrainian and Russian linguistic elements (in a broad sense) and to what extent this variation has already been reduced. Where variation can be observed, the question is, to what extent do intra-linguistic factors and to what extent do respondents’ social or sociobiographical characteristics play a role. This description will thus allow answers to be provided to the question of the strength of the impact of Ukrainian and Russian on the structural characteristics of Suržyk: At issue here are phenomena of gradual differences within the traditionally assumed and discussed Ukrainian-based Suržyk, the potential existence of a Russian-based Suržyk, and a possible gradual transition between them. Such a comprehensive investigation had been lacking to date. On the basis of OCUS, a pilot study on an inflectional morphological phenomenon was published by Hentschel and Palinska (2022b), a further study on derivational morphology and morphosyntax by Palinska and Hentschel (2022), and a study on a specific phonetic feature by Zeller (2022). Hentschel (2024) previously published a comprehensive study on the mixed Ukrainian-Russian lexicon.

As is the case with the four papers just mentioned, five of the papers in this special issue present the central findings of the third project OCUS is based upon, targeting the Black Sea coast, together with the findings based on the materials from two earlier projects targeting the Center of Ukraine.

Accordingly, the five studies are, first, rather comprehensive with a length of about 30 pages.¹² Second, they interact with each other. Mutual references are made explicit.

The central paper is Hentschel (forthcoming a). The article begins with a brief historical introduction on the emergence of the Ukrainian- and the potential Russian-based Suržyk, turns to theoretical considerations, and then formulates hypotheses about how the two variants of Suržyk might be distinguished in the analyses. These hypotheses are then tested against the empirical data, for theoretical reasons mainly on inflectional morphology. In other words, the empirical parts of the article synthesize all findings into a general conclusion reflecting the Oldenburg part of the project on the Black Sea coast. On the other hand, the historical and theoretical discussion serves as a point of reference for the other papers in the special issue.

Accordingly, Hentschel (forthcoming b), in his second paper, serving as the primary empirical input into his central paper, offers a detailed study of the distribution of Ukrainian and Russian endings in Suržyk for the central lexical parts of speech: nouns, adjectives, and verbs.

Zeller (forthcoming) proposes a detailed study of several relevant phonetic and phonological phenomena where Ukrainian and Russian exhibit clear differences. The interaction with morphology is discussed as well.

Hentschel, Zeller, and Palinska (forthcoming) contribute a study on vocatives and discourse markers, starting from the hypothesis that linguistic expressions in syntactic isolation may underlie different transfer regularities in language contact than expressions bound into sentence structure. Thus, the topic of *borrowing scale* is an issue for the paper.

Palinska and Zeller (forthcoming), last but not least, offer a study on selected morphosyntactic phenomena in Suržyk, phenomena that, again, exhibit contrasts between Ukrainian and Russian.

The special issue concludes with a contribution by Kuznietsova (forthcoming) of a completely different nature, not produced within the framework of the above-mentioned projects. We hope that this paper helps the reader who is hardly, or at best only slightly, acquainted with the linguistic situation in Ukraine and the corresponding public discourse to understand the sociocultural context in which Suržyk is embedded in Ukraine. It examines the sociopsychological perception and portrayal of Suržyk in Ukrainian media, by Ukrainian politicians, and in popular-scientific papers and articles by Ukrainian linguists.

This discourse falls roughly under what is generally known as purism. While linguistic purists typically fight against the “foreign” in the “native”, especially lexical borrowings, the situation here is somewhat different: Suržyk is often not even seen as “native” or “our own” by nationally-minded

¹² Note that in principle there are no page limits for papers, even in regular volumes of *Russian Linguistics*.

Ukrainian purists—or only with the greatest reservations. Years ago, the Ukrainian author Jurij Andrukhovych noted that Suržyk is viewed by many in Ukraine as the “incestuous child of Ukrainian-Russian bilingualism” (cf. Stavyc’ka, 2014). The linguistic influence of Russian in Ukraine, its political and social dominance, and correspondingly the—cautiously put—marginalization of Ukrainian, which has lasted for decades, if not centuries (cf. Danylenko & Naienko, 2019), has long been a highly emotional issue in Ukraine, long before 2014 or 2022. More than 30 years ago, Serbens’ka (1994) published her *Anty-Suržyk* in L’viv, a normative guide contrasting Suržyk elements or Russian elements in Suržyk and corresponding phrases with their equivalents in the “mother tongue” (*pidna mova*), i.e., in cultivated Standard Ukrainian. The article in this special issue by Tetiana Kuznietsova traces this discourse on Suržyk primarily over recent decades. She believes there are signs that the negative perception is beginning to shift—if only because speakers of the traditional, Ukrainian-based Suržyk are now viewed with considerable certainty as part of the national “us” rather than the “them” of the Russian occupiers. It seems highly unlikely that this would be the case with speakers of a Russian-based Suržyk—if such a variety exists.

Gerd Hentschel
Olesya Palinska
Jan-Patrick Zeller

April 2026

References

- Bilaniuk, L. (2004). A typology of Suržyk: Mixed Ukrainian-Russian language. *International Journal of Bilingualism*, 8(4), 409–425. <https://doi.org/10.1177/13670069040080040101>
- Bilaniuk, L. (2005). *Contested tongues: Language politics and cultural correction in Ukraine*. Ithaca.
- 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>
- Del Gaudio, S. (2010). *On the nature of Suržyk: A double perspective*. Sagner.
- Flier, M. (2008). Suržyk or suržyks? In G. Hentschel, & S. Zaprudski (Eds.), *Belarusian Trasjanka and Ukrainian Suržyk: Structural and Social Aspects of Their Description and Categorization* (pp. 39–56). BIS Verlag. [= *Studia Slavica Oldenburgensia*, 17].
- Hentschel, G. (2024). Ukrainian and Russian in the lexicon of Ukrainian Suržyk: Reduced variation and stabilisation in Central Ukraine and on the Black Sea coast. *Russian Linguistics*, 48(2). <https://doi.org/10.1007/s11185-023-09286-9>
- Hentschel, G. (forthcoming a). Hybridization from two sides? Ukrainian- vs. Russian-based Suržyk—Historical and theoretical aspects and empirical insights. *Russian Linguistics*, 50.
- Hentschel, G. (forthcoming b). Inflectional morphology in Ukrainian-Russian code-mixing—On endings and stems of main lexical word classes in Suržyk. *Russian Linguistics*, 50.

Hentschel, G., & Palinska, O. (2022a). 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, 259–290. <https://doi.org/10.1007/s11185-022-09259-4>

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

Hentschel, G., & Reuther, T. (2020). Ukrainisch-russisches und russisch-ukrainisches Code-Mixing. Untersuchungen in drei Regionen im Süden der Ukraine. *Colloquium: New Philologies*, 5(2), 105–132. <https://doi.org/10.23963/cnp.2020.5.2.5>

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

Hentschel, G., & Taranenko, O. (2021). Bilingualism or tricodism: Ukrainian, Russian and “Suržyk” in Ukraine. Analysis and linguistic-geographical mapping. *Die Welt der Slaven*, 66(2), 268–299. <https://doi.org/10.13173/WS.66.2.268>

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/slzw-2016-0039>

Hentschel, G., & Zeller, J. P. (2017). Aspekte der Sprachverwendung in zentralen Regionen der Ukraine. *Wiener Slawistischer Almanach*, 79(1), 37–60. <https://doi.org/10.5282/fqv8sk98>

Hentschel, G., Zeller, J. P., & Palinska, O. (forthcoming). On vocatives and discourse markers in Ukrainian Suržyk: Ukrainian or Russian? *Russian Linguistics*, 50.

Hentschel, G., Taranenko O., & Zaprudski S. (Eds.) (2014). *Trasjanka und Suržyk—gemischte weißrussisch-russische und ukrainisch-russische Rede. Sprachlicher Inzest in Weißrussland und der Ukraine?* Peter Lang. www.peterlang.com/index.cfm?event=cmp.ccc.seitenstruktur.detailseiten&seitentyp=produkt&pk=55043&concordid=58533

Kittel, B., Lindner, D., Brüggemann, M., Zeller, J.P., & Hentschel, G. (2018). *Sprachkontakt—Sprachmischung—Sprachwahl—Sprachwechsel. Eine sprachsoziologische Untersuchung der weißrussisch-russisch gemischten Rede „Trasjanka“ in Weißrussland*. Peter Lang.

Kuznietsova, T. V. (1999). *Mova bilinhval’noï sim’i u funkcional’nomu aspekti*. Dysertacija kand. filol. nauk: 10.02.01. Kyiv: Kyivs’kyj deržavnyj linhvistyčnyj universytet.

Kuznietsova, T. V. (2001a). Suržyk i movlennja. *Urok ukraïns’koï*, 6, 22–24.

Kuznietsova, T. V. (2001b). Sociolinhvistyčna sutnist’ suržyku. *Sučasni problemy movoznavstva ta literaturoznavstva*, 4 (Ukraïns’ke i slov’jans’ke movoznavstvo), 277–280.

Kuznietsova, T., (forthcoming): Suržyk in the cultural-political discourse of Ukrainian linguistics and media. *Russian Linguistics*, 50.

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

Moser, M. (2016). “Suržyk” čy “suržyky”? *Ukraïns’ka mova*, 1, 27–54.

Hentschel, G., Zeller, J. P., & Palinska, O. (2026). *Documentation of the “Oldenburg Corpus of Ukrainian Suržyk”*. <https://uol.de/slavistik/forschung/sprachwissenschaft/bereich-mischvarietaeten/okus-dok>

Palinska, O., & Hentschel, G. (2022). Regional’nye osobennosti ispol’zovanija ukraïnsko-russkoj smešannoï reči (suržika) i vlijanie dialektov: pristavki i predlogi VID / OT. *LingVaria*, 2(34), 229–253. <https://doi.org/10.12797/LV.17.2022.34.15>

Palinska, O., & Zeller, J. P. (forthcoming). Competing morphosyntactic constructions in Suržyk: Ukrainian or Russian? *Russian Linguistics*, 50.

Serbens'ka, O. (1994). *Antysuržyk*. Svit.

Stavyc'ka, L. (2014). Ein Blutschandekind der Postmoderne. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Suržyk—gemischte weißrussisch-russische und ukrainisch-russische Rede. Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 351–374). Peter Lang.

Šumarova, N. (2014). Der Suržyk im System nah verwandter Zweisprachigkeit: soziolinguistischer und linguistischer Aspekt. In G. Hentschel, O. Taranenko, & S. Zaprudski (Eds.), *Trasjanka und Suržyk—gemischte weißrussisch-russische und ukrainisch-russische Rede: Sprachlicher Inzest in Weißrussland und der Ukraine?* (pp. 307–326). Peter Lang.

Taranenko, O. (2007). Ukrainian and Russian in contact: attraction and estrangement. *International Journal of the Sociology of Language*, 183, 119–140.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

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

Zeller, J. P. (2022). Okannia and Akannia in Ukrainian-Russian mixed speech (“Surzhyk”). *Ukraïns'ka mova*, 2(82), 38–59. <https://doi.org/10.15407/ukrmova2022.02.038>

Zeller, J. P. (forthcoming). The impact of Ukrainian and Russian on sound patterns in Suržyk. *Russian Linguistics*, 50.

Zeller, J. P., & Hentschel, G. (2024). Die ukrainische Schwarzmeerküste. Sprachen—Nationalitäten—Identitäten. In R. Goldt, B. Wiemer, *Die Ukraine als Objekt russischer Großmachtansprüche. Sprachen, Identitäten und Diskurse* (pp. 21–63). Springer. <https://doi.org/10.26530/20.500.12657/86052>