

Hybridization from two sides? Ukrainian- vs. Russian-based Suržyk

Historical and theoretical aspects and empirical insights

Abstract

Ukrainian-Russian language contact over many generations has produced a mixed code commonly known as *Suržyk*. Historically, *Suržyk* developed from Ukrainian rural speakers' approximation of the Russian language, which dominated in urban areas as the language of power — thus, a Ukrainian-based *Suržyk*. Specific to *Suržyk* is that, over time, it was not only used by its speakers in communication with Russian-speaking interlocutors — typically representatives of authority — but also in communication among themselves and with their children. Millions of Ukrainian citizens make regular use of it, at least in certain communicative settings.

During the Tsarist and Soviet periods, Russian was the socially and politically dominant language, especially in the eastern and southern parts of the country. Ukrainian was pushed to the margins of the state and society until 1990. After Ukraine's independence, a certain reversal began: Ukrainian was politically promoted and required by public institutions (administration, media, educational institutions). This development recently gave rise to the hypothesis of the emergence of a Russian-based *Suržyk* — Russian speakers' approximation of Ukrainian.

To date, almost no large-scale empirical investigations have been conducted — not even on the “old” Ukrainian-based *Suržyk*, let alone on the possibly emerging Russian-based variety. The present study, along with four others in this special issue of *Russian Linguistics*, addresses this gap through a quantitative analysis of a corpus comprising over 700,000 word forms that examines different linguistic subsystems. The central question of this study is whether a structural distinction can be drawn between a Ukrainian-based and a Russian-based *Suržyk*. The underlying hypothesis is that, for speakers of a Ukrainian-based *Suržyk*, elements from more closed linguistic subsystems — especially inflectional morphology — should show clear Ukrainian features, whereas in a Russian-based *Suržyk*, inflectional morphology should more clearly follow Russian patterns.

A Russian-based *Suržyk* could not be demonstrated — not in the data on inflectional morphology, nor in other structural domains. The most striking phenomenon observed was the relatively stable realization of Ukrainian word stems with Ukrainian endings. Gradational deviations could only be detected in a handful of combinations occurring in inflectional grammatical categories (e.g., case, gender, number), which can plausibly be explained with preference theory. A considerable degree of variation between Ukrainian and Russian endings was found only with Russian word stems. On the basis of this variation, speaker groups can be distinguished according to the degree to which they adapt Russian stems to Ukrainian inflectional morphology. In this context, sociobiographical factors — particularly the degree of speakers' everyday

exposure to Russian — play a role. Nevertheless, the Ukrainian-based Suržyk can vary in its degree of Russification depending on the speaker, which is most visible in the treatment of Russian stems, not only in their inflectional morphology. Lexically, by contrast, Russian influence is strong.

1 Brief historical background

Suržyk is a highly colloquial mixed language variety found in Ukraine. Its emergence is the result of the centuries-long political and social dominance of the Russian language in Ukrainian territory.¹ The western part of Ukraine was largely exempt from this sustained dominance, particularly the region surrounding the metropolis of L'viv, which only fell to the Soviet Union after the end of World War II and thus came under Russian rule for the first time in its history — including the sociopolitical dominance of the Russian language.

In contrast, this dominance has much deeper roots in the eastern and southern regions: the so-called Left Bank of Ukraine (east of the Dnipro River), as well as Kyiv on the opposite side, fell under Moscow's control in the second half of the 17th century; the Black Sea coast, previously not part of the East Slavic linguistic area, followed at the end of the 18th century, falling under the reign of St. Petersburg. As a result of the so-called Partitions of Poland between 1772 and 1795, and later the Congress of Vienna in 1815, large parts of the Right Bank of Ukraine — previously dominated linguistically by Polish — came under the rule of the Russian Empire. The notable exception was the western part of Ukraine, which became part of the Habsburg Empire and, after World War I, with slightly shifted eastern borders, was once again part of Poland.²

In Ukrainian territories governed by Moscow or St. Petersburg, Russian was always the language of power. This situation did not change even as a modern Ukrainian literary or written language began to take shape in the 19th century, a process that was subject to numerous repressive measures until the early 20th century. The 1920s in the early Soviet Union were more favorable to the development of a Ukrainian standard language. Later, and not only during Stalin's purges at the end of the 1930s, Ukrainian — despite nominal tolerance — again came under political and social pressure.³

Until the end of the Soviet Union, it was virtually impossible to pursue even an average professional or societal career without solid knowledge of Russian, or without approximating Russian from genetically and

¹ The terms “Ukrainian territory” or “territory of Ukraine” refer to the geographical area that belong to independent Ukraine. This area corresponds to the territory of the Ukrainian Soviet Socialist Republic (UkrSSR) as of 1958, when Crimea was incorporated into the UkrSSR under Nikita Xruščev.

² A concise overview is provided by Kappeler (2019).

³ A comprehensive account of the disadvantages — and in many cases suppression and harassment — faced by the Ukrainian language at the hands of Russian or Soviet authorities over the centuries is offered by Danylenko & Naienko (2019).

structurally quite similar Ukrainian or its regional dialectal variants. For more prominent positions in the state and society, excellent command of Russian was a “*conditio sine qua non*”.

2 Suržyk or Suržyks — in synchrony and short-term diachrony

The acquisition of Russian — or even a complete shift to Russian⁴ — was less of a problem for the educated elites, provided they had the will to do so. However, the abovementioned linguistic approximation to Russian as the language of power was — at least to a certain degree — also required of the rural Ukrainian population, in particular if, in the process of industrialization and the accompanying urbanization, they migrated into cities and towns. Their starting point was the broad spectrum of local dialects.

These processes had already begun during the Tsarist period and reached further peaks during the pre-World War II Soviet era and again after the large-scale destruction during the war. In rural areas, forced collectivization under the Soviet planned economy certainly played a role. Taranenko (2007) points out that these conditions led rural-urban Ukrainian migrants to adopt or develop a mixed form of Ukrainian-Russian speech, in which they increasingly integrated elements of Russian into their primarily Ukrainian — i.e., mainly dialectal — speech. Because of the high degree of structural similarity, particularly in the realm of inflectional grammatical categories as well as morphological and syntactic structures in general, the massive “import” of Russian elements — lexical as well as inflectional and morphosyntactic — was structurally unproblematic (see below). There are hardly any constraints on the borrowing of non-lexical elements in contact between closely related and structurally similar languages, which is more similar to the interaction between local dialects of a single language (cf. Thomason & Kaufman, 1988, pp. 65–77; Thomason, 2015, pp. 28–29). Mixtures between such languages therefore resemble dialect leveling. Other factors contributing to the influence of Russian on Ukrainian included Russian being the mandatory language of instruction, the command language in the armed forces, and the obligatory language in technical and scientific domains — with few exceptions in the academic context, such as the languages and literatures of the Soviet republics (cf. Kravciv & Kubijovyč, 1973; Dzijuba, 2012). For the majority of the population, the Ukrainian standard language can be understood as an adstratum or lower-ranking superstratum, situated between the dialectal substratum and the dominant superstratum Russian. Ukrainian was always taught in Soviet times to a certain degree and with significant differences in scale depending on school type and region

There is no doubt that Ukrainian speakers approximated Russian in communication with Russians in various situational contexts even before the end of the 19th or during the 20th century. However, the

⁴ This shift was, at least in the 18th and 19th centuries, less a shift away from a standardized Ukrainian — since such a standard only emerged in the 19th century — and more a shift away from Polish or even French. The extent to which members of the elites were influenced by a dialectal Ukrainian substratum likely varied and is not of concern here.

specificity that developed in the course of the 20th century was that this mixed form of speech was not only practiced in communication with Russians or Russian speakers but also, first, among Ukrainians themselves and, second, with their children.⁵ The latter forms the basis for the various stabilizations of Ukrainian-Russian language mixing, as presented Hentschel and Palinska (2022a), Palinska and Hentschel (2022), Zeller (2022), Hentschel (2024), in other contributions in this volume⁶ and in this study.

This, so to speak, traditional manifestation or default variant of Suržyk can, at least from a diachronic perspective, be understood as Ukrainian-based Suržyk: speakers (primarily of dialectal varieties) of Ukrainian transferred, more or less intensively, elements from Russian into their speech. The first generation of these speakers would then, for example, upon returning to their villages and speaking with their parents or grandparents, be able to effortlessly revert to the locally common dialect, though occasional use of “modern” Russicisms acquired in the city — especially for specific aspects of Soviet urban life — would not be excluded. For their children, i.e., the second generation of Suržyk speakers, and even more so for the next, this is fundamentally different because their parents as a rule spoke to them in mixed Ukrainian-Russian from birth. The parents did so in order to facilitate their children’s access to Russian, at least via their own approximation of it, because they themselves had experienced rural dialects as an obstacle to social advancement and, in some cases, had already lost their autochthonous dialectal competence. This is, of course, not a unique experience in Europe. Contact with the original dialectal speech will depend on the degree to which the children maintain ties to their parents’ home villages — bearing in mind that mother and father do not necessarily come from the same rural area.

The early linguistic socialization of children born to first-generation Suržyk speakers⁷ — and to their children and grandchildren (now likely two or three, if not four generations) — already takes place within the mixed variety of Ukrainian-based Suržyk and thus is not directly based on any specific, local rural speech norm. Typically, it is only upon entering school that these children are confronted with the challenge of at least approximately, and ideally clearly, separating Ukrainian and Russian in their speech. The acquisition of standard Ukrainian and Russian can best be understood as the development of filters that suppress certain elements when Ukrainian is required and others when Russian is expected (see Hentschel,

⁵ In conversations with Ukrainians, some speakers of Russian would, of course, spontaneously approximate Ukrainian, depending on their proficiency in Ukrainian and the specific situation. In such cases, this was primarily motivated by the need to be understood. This, however, did not lead to any kind of fused lect in the sense of Auer (1999), as there was no social pressure to adapt to Ukrainian.

⁶ Cf. Zeller (forthcoming), Hentschel (forthcoming), Palinska and Zeller (forthcoming), Hentschel, Zeller, & Palinska (forthcoming).

⁷ First-generation speakers of Suržyk are, of course, not tied to a specific point in time after, say, the end of the 19th century. The decisive criterion is the timing of their migration from villages to towns and cities. Thus, the Suržyk of later first-generation speakers is — on the basis of their local Ukrainian dialect — shaped not only by Russian, but possibly also by the Suržyk of earlier first-generation speakers in their surroundings.

2017).⁸ It now becomes apparent that already the second generation of Suržyk speakers no longer has a “clear” Ukrainian base in the same sense as the older generation. For the younger generation the base was already mixed, though, of course with a more or less clear dominance of Ukrainian. So much may be said initially about Ukrainian-based Suržyk, if we arbitrarily — but not without reason — date its origins as a mass phenomenon at latest to the early Soviet period.

In Suržyk research, the question has been raised as to how many Suržyks should be distinguished — that is, whether there are different types.⁹ Bilaniuk (2004, pp. 415–422) proposes a classification of five distinct variants of Suržyk: (i) Urbanized peasant S., (ii) Village peasant S., (iii) Sovietized Ukrainian S., (iv) Urban bilinguals’ S., and (v) Post-independence S.

Bilaniuk’s variant (i) corresponds to the aforementioned Suržyk of Ukrainian rural-to-urban migrants, i.e., the first generation of Suržyk speakers. Variant (ii) refers to a rural version, practiced by its speakers particularly in interactions with Russian-speaking authorities (cf. the earlier note on agricultural collectivization). Variant (iii) concerns two versions of the Ukrainian standard language that emerged when the 1920s “Kharkiv orthography” (“Skrypnykivka”) — which deviated significantly from the Russian standard — was replaced by the more Russified “Orthography of 1933” or “Khvyliia orthography” during the 1930s. The latter was created and politically enforced by Soviet language planning (cf. Karunyk, 2017; Horbyk & Palko, 2017), with the intention that the Ukrainian standard language would increasingly converge with the Russian standard, for example through borrowings, neologisms formed by Russian patterns, and so on (cf. Skopnenko, 2012; Masenko et al., 2005). The same concept of targeted development applied to Belarusian. The political goal was ultimately a merging of the two “smaller” East Slavic languages into Russian. Regarding this variant (iii), one may question the logic of categorizing this planned, top-down written-language phenomenon together with the dynamic, primarily spoken, bottom-up phenomena of Ukrainian-Russian language mixing. Variant (iv) roughly corresponds firstly to the Suržyk of later generations of Suržyk speakers, derived from (i) or (ii), who already grew up in towns, if not cities. (Bilaniuk explicitly mentions Kyïv.) There, they developed a competence in both Ukrainian and Russian on top of the Suržyk of their first linguistic socialization. Secondly, this fourth variant, according to Bilaniuk (2004, pp. 419–420), also comprises Russian speakers trying to speak Ukrainian. Bilaniuk describes this variant of Suržyk, which lacks a “single directionality”, as the most diffuse one, ranging from code-switching to code-mixing in Auer’s (1999) continuum. A stabilization or, with Auer, fusion is at least implicitly denied. The heterogeneity, Bilaniuk alludes to, may, however, be due to her metalinguistic

⁸ Matras (2009, pp. 61–66) describes this in relation to other bilingual or multilingual constellations in the linguistic socialization of children.

⁹ In Ukraine, there are also other forms of code-switching and code-mixing, for example between Ukrainian and Polish. These are conditioned by very different historical and sociolinguistic factors and are also much more regionally restricted. Masenko (2019) rightly argues that the term *Suržyk* should be reserved exclusively for the mixture of Ukrainian and Russian.

categorization. Variant (v) refers to Russophone city dwellers of the post-Soviet period, especially from the eastern or southern *oblasts* (regional administrative divisions) or from cities in the East, who, in independent Ukraine, increasingly use Ukrainian in public situations where they would have used Russian during Soviet times. This variant, thus, rather exclusively embraces speakers with a linguistically Russian background, whereas variant (iv) does so only partially. The decisive difference seems to be based in chronology and area. The reasons for an increased use of (“intended”) Ukrainian may include practical necessities resulting from state language policy, or personal motivations such as expressing loyalty to the Ukrainian state. In this final case (v), Russian would clearly be the starting language from which an approximation toward Ukrainian occurs. Notably, this variant, limited by Bilaniuk to post-Soviet times, would be of very recent origin. Here, as for variant (iv), Bilaniuk speaks of code-switching and code-mixing in the sense of Auer. For (v) the implicit exclusion of fusion is plausible, in view of the short time period assumed for it. This is different for (iv), at least for people with a Ukrainian background.

In his study “Suržyk or Suržyks?”¹⁰, Flier (2008) argues against distinguishing Bilaniuk’s three (or, with the inclusion of the negligible variant (iii), four) first variants as different types of Suržyk. Without going into the details of his argument, one may agree with his view because the potential speakers of these variants do not form strictly discrete groups (communities of practice — cf. Meyerhoff, 2002), neither synchronically nor (even less) diachronically over the past 100 years. While certain features may be typical of the different social contexts Bilaniuk (2004) describes, they are rarely exclusive. Instead, these phenomena tend to overlap significantly across social and communicative contexts, potentially with significant differences in the frequency of occurrence. From the perspective of the variationist paradigm, it would make more sense to assume that relevant social or sociobiographical parameters affect variation (cf. Hentschel & Palinska, 2022a).

However, Flier’s acceptance of Bilaniuk’s variant (v) — a Russian-based Suržyk — is to be viewed more critically. He claims that the term “Suržyk” should be reserved for hybrids with clearly differentiated “matrix and embedded languages” (Flier, 2008, p. 54). He thus acknowledges two types of Suržyk: one with Ukrainian as the matrix language and the other with Russian; the remaining language would in each case be the embedded one. If we take the matrix vs. embedded language distinction in the sense of Myers-Scotton (2002) seriously — not as a pre-theoretical notion — this becomes debatable. The prevailing type of code-mixing in Suržyk is so-called congruent lexicalization, typical for situations involving genetically and structurally similar codes, including the mixing of a standard language with one or more of its dialects (cf. Muysken, 2000). Switching between or, better, mixing of elements from either language faces almost no syntagmatic restrictions and can occur even within word boundaries. A matrix-embedded distinction as defined by Myers-Scotton (2002) is not feasible here. With closely related and structurally similar languages, there are too few syntagmatic structural differences to permit a matrix/embedded language

¹⁰ Moser (2016) presents an article under the same title, in Ukrainian. He concludes that the individual Suržyks of speakers have nothing in common; see below in the Conclusions, Fn. 54.

differentiation. According to Muysken (2000), in cases of congruent lexicalization, only quantitative analysis can determine which language is dominant. In such cases, Muysken suggests using the term “dominant language” for the quantitatively prevailing one in a stabilized mixture.¹¹ That said, this does not rule out that, over generations — for example, within families or among individuals — a Suržyk that was originally clearly Ukrainian-dominated may gradually become more Russian-colored if Russian is stronger in the broader social context of contact or if there is social pressure. Instead of a “matrix language turnover” in the sense of Myers-Scotton (2002, pp. 247–249), one could rather speak of a gradually progressing “dominant language shift.”

The observations above on at least diachronic Ukrainian-based Suržyk regarding congruent lexicalization as a basic type of mixing and the possibilities for a quantitative shift toward Russian apply similarly to a Suržyk resulting from an approximation by native Russian speakers to Ukrainian — especially in cases where these speakers now use the mixed code even though they still have good command of Russian (and maybe some command of Ukrainian), notwithstanding certain accent features, particularly phonetic ones.

A crucial point in this discussion is the difference in age between the two potentially distinct types of Suržyk: A socially widespread Suržyk with Russian as the starting language would only be just over thirty years old, meaning there could not yet have been significant generational transmission. In contrast, the older, diachronically Ukrainian-based variant has existed for at least a century. One may hypothesize that this older variant is more stable, while the newer one is far more variable (cf. Hentschel & Reuther, 2020). According to Trudgill (1986), new mixed dialects tend to stabilize over three or four generations.

3 On the relevance of linguistic structural levels: hypothesis formation

Matras & Bakker (2003, pp. 1–22) define as mixed languages those that show a so-called lexicon grammar split. Simply put, this means that one language contributes the lexicon while the other provides the grammar. It is important to note that the distinction between lexicon and grammar should be viewed as gradated across multiple dimensions, meaning that in different constellations of mixed and source languages the split can occur at different points on a scale ranging from maximally grammatical to maximally lexical. This model, which Matras and Bakker (2003) illustrate using structurally-typologically and genetically diverse mixed languages, cannot be transferred “in pure form” to a mixed variety like Suržyk. As noted above, the differences between Ukrainian and Russian grammar are rarely categorical but instead pertain primarily to the expression of inflectional-morphological categories. In most cases, they involve differing distributions of fundamentally common endings, and only in few instances, different specific endings not present in the other language. Ukrainian and Russian syntax differ hardly at all (for some of several contrasts in morphosyntax, see Palinska and Zeller forthcoming).

¹¹ The term matrix language is used rather vaguely — not only by Flier but also by Thomason (2015) and Grenoble (2010) in reference to Suržyk — and often merely in the sense of “dominant language.” This conflates the structural conceptual distinction between matrix and embedded (subordinate) language.

The differences in the lexicon are more pronounced. The clearest cases involve borrowings from other languages (e.g., Ukr. *dax* from German *Dach* via Polish *dach* vs. Russ. *kryša*, a so-called inherited word, all meaning ‘roof’) or different developments in form-meaning relationships among historically common lexemes (e.g., Ukr. *mova* vs. Russ. *jazyk*, both ‘language’). Other differences are subtler, such as morphophonological variants of common inherited words (e.g., Ukr. *ridnyj* vs. Russ. *rodnoj*, both ‘native, own’; Ukr. *sluxaty* vs. Russ. *slušat’*, both ‘listen’). Still, a large area of overlap remains, ranging from inherited or jointly borrowed words with (phonologically governed) phonetic differences (e.g., Ukr. *podruha* vs. Russ. *podruga*, ‘female friend’ or Ukr. *maslo*, with [ɔ] in last, unstressed syllable, and Russ. *maslo*, with [ə], both ‘butter’) to complete phonetic identity (e.g., Ukr./Russ. *tam*, ‘there’).

An additional phenomenon must be noted: Ukrainian and Russian stems differ sometimes only in certain grammatical category combinations such as case-(gender)-number constellations, while in others, they are identical or indeterminate (e.g., Ukr. *kit* vs. Russ. *kot*, both ‘male cat — nom.sg.’, but *kota* in both Ukr./Russ. for gen.sg., neutralizing the stem contrast — see Hentschel 2024, 6 and forthcoming for more detail). Thus, lexical differences concern stem differences that affect either entire inflectional paradigms or parts of them.

A useful scheme for distinguishing between Ukrainian- and Russian-influenced Suržyk may lie in an observation by Cyxun (1998), although his focus was on the (at least diachronically) Belarusian-based Trasjanka, the Belarusian counterpart to Suržyk.¹² Cyxun (1998; similarly, Liskovec, 2006) argues that the lexicon of Trasjanka is almost entirely Russian,¹³ the inflectional morphology is mixed, and the phonology and phonetics are Belarusian. Hentschel (2013), Hentschel and Zeller (2014), and Zeller (2015) have confirmed this point with various corrections or specifications,¹⁴ but the tendency — i.e., the hierarchy Phonetics/Phonology < Morphology < Lexicon regarding the degree of Russian influence — remains valid. Thus, Trasjanka would be a case in which Russian, despite a historically Belarusian base, is the quantitatively dominant language (see Hentschel, 2017). The latter is largely conditioned by the fact that Belarusian is rarely spoken in Belarus, with Russian and Trasjanka being predominant in everyday communication (Hentschel & Kittel, 2011). Ukrainian, by contrast, is much more widespread in Ukrainian

¹² After the election of A. Lukašenka as President of Belarus, efforts to politically strengthen the role of the Belarusian language (analogous to the role of Ukrainian in an independent Ukraine) quickly dissipated. Therefore, the social or sociopolitical conditions for a Russian-based Trasjanka to emerge were not given. While individual cases of Russophones approximating Belarusian may exist in the few remaining Belarusian-speaking families, there is no Russian-based Trasjanka as a “group language” of originally Russian-speaking people turning to Belarusian and using an approximative version of the latter for mutual communication.

¹³ Hentschel (2024) demonstrates that the influence of Russian on the lexicon of Suržyk is by no means as far-reaching as in the case of Belarusian Trasjanka.

¹⁴ On the one hand, for example, the more “grammatical” function words are far less influenced by Russian than “strictly semantic” lexical items, and, on the other hand, even phonology and phonetics show traces of Russian influence.

society (Hentschel & Taranenko, 2021) and has been promoted by language policy over the last three decades.

Hypothetically, in a case of “dominant language shift,” a quantitative middle ground could be reached, raising the question of whether the two types of Suržyk are even distinguishable in principle. Such a quantitative midpoint would nevertheless be structurally asymmetric: If Ukrainian is the diachronic base language, then more closed subsystems such as inflectional morphology should show a quantitative dominance of Ukrainian elements (see below and Hentschel forthcoming). In a Russian-based Suržyk, the reverse should hold. In more open subsystems or only partially structured linguistic domains like the lexicon, things are less predictable. One may assume, however, that basic lexical items and function words would, in general, much better reflect the basic language of a mixed or fused lect than all other, purely “semantic” lexical items. This is not at issue here, because an exhaustive lexical investigation would require a much larger corpus than the corpus of roughly 730,000 word forms that constitutes the empirical base for this and all other analytic studies in this volume (see below and Hentschel, 2024).

Based on the above discussion, the following hypotheses can be formulated:

(H1): In a Suržyk with Ukrainian as the historical base, the “sound patterns” should be more Ukrainian. After all, many Ukrainians who are grammatically and lexically proficient in Russian still show a Ukrainian-influenced accent. This does not exclude the possibility that certain phonetic-phonological Russicisms may have become established across generations.¹⁵ Conversely, in a Suržyk with Russian as the base — that is, a relatively recent approximation by Russian speakers to Ukrainian — the “sound patterns” should be clearly more Russian. Even if phonic¹⁶ phenomena might be considered merely surface phenomena of a Russian or Ukrainian accent and not as indications of a fully developed “fused lect” in Auer’s (1999) sense, this question is addressed in Zeller (forthcoming). Zeller’s work (2015, 2018, 2022) so far suggests, however, that this aspect may be more complex in cases of contact between closely related languages that form part of a (historical) dialect continuum.

The transition from spontaneous mixing to a conventionalized mixture is more likely to be observed in structurally deeper phenomena — in the context of East Slavic language contact, especially in inflectional morphology. The lexicon, as a maximally open subsystem, and morphosyntax are more susceptible to

¹⁵ Zeller (2015) demonstrates this for the generationally developed Trasjanka based on Belarusian.

¹⁶ We use “phonic” following Weinreich (1953) as an umbrella term for both phonetic and phonological phenomena. As shown in Hentschel and Palinska (2022a), Suržyk should be understood as a mesolect in the wide area of Ukraine positioned between standard language(s) and fine-grained local dialects (subdialects). This mesolect is also locally differentiated, but on a coarser scale than traditional rural dialects. The linguistic features of local mesolect variants are shaped both by old dialects and by the social relevance of the two standard languages, with locally varying power dynamics among the three elements. Thus, the idea of a unified phonological system of Suržyk does not apply. Weinreich’s (1954) diasystematic approach — deriving one phonological system on the basis of structuralist (or even more abstract generativist) phonemes for a dialect continuum — has, as is well known, failed. We use the terms “phonics” or “sound patterns” similarly to “phonic”.

external influence and are less conclusive for investigating hybridization from both sides. However, connections between inflectional morphology and phonology/phonetics remain informative.¹⁷

Building on Cyxun (1998) and Liskovec (2006) — who note the intermediate susceptibility of inflectional morphology in the Belarusian-based Trasjanka to Russian influence — Menzel and Hentschel (2017) offer another quantitative and qualitative insight. Alongside Suržyk and Trasjanka, they examine colloquial Lemko¹⁸ and colloquial Kashubian in Poland — a minority and a regional language respectively. While Russian was or still is the dominant political and social language in Ukraine and Belarus, respectively, Polish serves that role in Poland. For inflected lexemes, they observe a strong tendency toward what Eugenio Coseriu (1967) calls “lexical solidarities”¹⁹ — here more freely interpreted as “tendential etymological-morphological solidarity”.²⁰ This principle implies that even in conventionalized mixed codes, stem-ending combinations reflect the etymological origins of both donor languages²¹ to a considerable degree: stems from the dominated or dialectal base language rarely take inflectional endings from the dominant “roof” language (in Kloss’s 1987 sense). These imported inflectional endings tend to appear with stems also borrowed from the dominant language, though they may also be morphologically adapted to the base language.²² Thus, solidarity according to donor or recipient language origin acts as one of several probabilistic factors.

From this, a second hypothesis follows:

(H2): In a mixed variety, word stems from the base language should, all else being equal, occur significantly more often with inflectional endings from that same language. Stems borrowed from the dominant language, by contrast, should more frequently take inflectional endings from the dominant language, but they can also exhibit a considerable degree of adaptation to the inflectional patterns of the

¹⁷ On the lexicon, see Hentschel (2024, forthcoming) and Hentschel, Zeller, & Palinska (forthcoming) on discourse markers; for morphosyntax, see Palinska & Zeller (forthcoming).

¹⁸ Lemko emerged from the southwestern fringes of the East Slavic — or, if one prefers, Ukrainian — dialect continuum and is generally not recognized as a separate language in Ukrainian linguistics nor by many politically Ukrainian-leaning speakers themselves (see Rusačenko & Koxančyk 2019; Myxal’čuk 2023). This parallels the refusal of some “Great Russians” to recognize Ukrainian (“Little Russian”) as an independent language (cf. Florinskij, 1900; Ljapunov, 1919).

¹⁹ In German, *fressen* is the neutral term for when animals feed, while *essen* is used for when humans eat. The pejorative use of *fressen* for people is based on this solidarity — or rather, on breaking it.

²⁰ Matras (2015, pp. 48, 66ff.) refers to this phenomenon in mixed codes as “morphological compartmentalization”.

²¹ The application of the term *donor language* is not unproblematic in cases such as Ukrainian Suržyk. What is certain is that this “mixed” subvariety exhibits Ukrainian and Russian features, elements, and constructions across various structural levels of language. In a Suržyk (clearly) based on Ukrainian, only Russian could be regarded as the donor language and Ukrainian as the recipient language. Conversely, in a clearly Russian-based “Neo-Suržyk,” only Ukrainian would qualify as the *donor language*, with Russian serving as the *recipient language*. However, since the question of the actual “existence” of such a Neo-Suržyk remains unresolved, I will for the time being use the term *donor language* in the broader sense, applying it to both Ukrainian and Russian. I will return to this issue in the final section.

²² Grečko & Palinska (2019) observe the same tendency for “solidarity” between root morphemes and derivational morphemes in both Suržyk and Trasjanka.

dominated language. For the question of Ukrainian- versus Russian-based Suržyk, this implies: In a historically Ukrainian-based mixed variety, it is expected that Ukrainian endings will predominantly occur with Ukrainian and common stems, while Russian stems will exhibit a comparatively higher frequency of Russian endings. Conversely, in a Russian-based Suržyk, it is expected that Russian endings will predominate with Russian and common stems, whereas Ukrainian stems will show an increased frequency of Ukrainian endings.

A complication arises from the fact that Ukrainian and Russian are genetically closely related, and many lexemes or stems are nonspecific, i.e., common to both contact languages. These offer an interesting case: Among historically Ukrainian-based Suržyk speakers, such stems should be treated more like Ukrainian stems in terms of inflectional ending combinations; among Russian-based Suržyk speakers, more like Russian ones.²³

Of the two hypotheses above, this contribution primarily investigates H2 — i.e., inflectional morphology — because of its central relevance to the question of whether there are two distinct Suržyks or whether the linguistic data instead indicate gradual differences, as suggested by the notion of a “dominant language shift”.

Methodologically, two analytical tools are available: First, an apparent-time analysis, focusing on respondents’ age. Second, the sociolinguistic background of respondents must be considered — specifically, the role played by the two source languages in their lives. Is the respondent’s sociobiographical background more “Ukrainian” or “Russian” in linguistic terms? Naturally, it must be considered that such differences may not be strictly dichotomous but gradational instead (and may vary across people’s lifetimes), which would in turn influence the relative weighting of Ukrainian and Russian components in Suržyk. Data on both age and sociolinguistic background were collected from the respondents who provided language data.²⁴

4 Analyses

4.1 Origin or Affinity of Stems and Endings in Inflected Word Classes

As outlined above, inflectional morphology plays a special role in the question of whether one should assume a Russian-based Suržyk in addition to an undoubtedly existent Ukrainian-based one. Phrased differently the question is whether for some respondents, Russian influence is structurally more fundamental than Ukrainian.

If a quantitative analysis is to provide answers to the questions posed above and test the central hypothesis,

²³ Hybrid stems constitute a special case and are not considered here. See Hentschel (forthcoming).

²⁴ For details on the data set — i.e., language data collection and sociobiographical data — see the *Introduction* of this special issue of *Russian Linguistics* (Hentschel, Palinska, & Zeller, forthcoming, and OCUS-Doc: Hentschel et al., 2026).

it cannot merely consider the frequencies of morphological endings from Ukrainian or Russian. Rather, the aim must be to examine to what extent Ukrainian, Russian, and common as well as possibly hybrid²⁵ *stems* are combined with Ukrainian or Russian *endings*. Common *endings* are excluded from the analysis, as neither provides evidence for the core question addressed here — namely, the extent and depth of Ukrainian or Russian impact on Suržyk or the possibility of hybridization from both sides. Hentschel (forthcoming) investigates this issue for selected constellations of morphological categories (e.g., specific case-number-gender or tense-number-person/gender combinations), for which clear contrasts between Ukrainian and Russian exist, specifically within the lexical word classes of nouns, adjectives, and verbs. Significant differences were found among these constellations or groups in the quantitative ratio of Ukrainian to Russian endings — further differentiated by the stem types: Ukrainian, common, Russian.²⁶

Table 1, reproduced here from Hentschel (forthcoming) for ease of presentation, provides illustrative examples:

Table 1 *Constellations of Ukrainian, Russian, and common stems and endings*

Stem	Ending	Example ^a	Constellation
Ukr.	Ukr.	a) <i>narodženn'-a</i> 'birth' b) <i>maju-t'</i> 'have' c) <i>cikav-i</i> 'interesting'	NOM.SG.NEUTR 3PERS.PRES.PL NOM.PL
Ukr.	Russ.	a) <i>bat'k-ov</i> 'parents' b) <i>ču-l-i</i> 'heard' c) <i>povs'akdenn-ov</i> 'everyday'	GEN.PL PAST.PL GEN.SG.FEM
Ukr.	Common ^b	a) <i>mov-a</i> 'language' b) <i>kaza-l-a</i> 'say' c) <i>ostann-ij</i> 'last'	NOM.SG.FEM PAST.FEM.SG NOM.SG.MASC
Russ.	Russ.	a) <i>stran'-e</i> 'country' b) <i>zavisi-t</i> 'depend' c) <i>obšč-ij</i> 'common / shared'	DAT.SG.FEM 3PERS.PRES.SG NOM.SG.MASC
Russ.	Ukr.	a) <i>stix-y</i> 'poem' b) <i>prepodaje-Ø</i> 'teach' c) <i>ũkusn-e</i> 'delicious'	NOM.PL 3PERS.PRES.SG NOM.SG.NEUTR

²⁵ Hybrid stems consist mostly of a Russian root and Ukrainian affixes (cf. Grečko & Palins'ka, 2019). They will only be considered in passing below, because they are too infrequent to formulate generalizations (cf. Hentschel forthcoming).

²⁶ The issue of how to distinguish Ukrainian, Russian, common, and hybrid stems or endings will not be further addressed here. This issue is discussed by Hentschel (forthcoming) and in greater detail OCUS-Doc: Hentschel et al. (2026).

Russ.	common	a) <i>bol'nic-a</i> 'hospital' b) <i>ščita-l-a</i> 'think' c) <i>izvestn-yj</i> 'famous'	NOM.SG.FEM PAST.FEM.SG NOM.SG.MASC
common	common	a) <i>kuxn'-a</i> 'kitchen' b) <i>zna-l-a</i> 'know' c) <i>dobr-yj</i> 'good(hearted)'	NOM.SG.FEM PAST.FEM.SG NOM.SG.MASC
common	Ukr.	a) <i>l'ud'-my</i> 'people' b) <i>žyve-mo</i> 'live' c) <i>molod-e</i> 'young'	INSTR.PL 1PERS.PRES.PL NOM.SG.NEUTR
common	Russ.	a) <i>mam-oj</i> 'mom' b) <i>xoče-t</i> 'want' c) <i>nov-oje</i> 'new'	INSTR.SG.FEM 3PERS.PRES.SG NOM.SG.NEUTR

^a a) noun, b) verb, c) adjective. Note that in the examples for verbs in the 3PERS.PRES forms in Table 1, a simplified notation for stems has been adopted, interpreting the vowel preceding the ending as a part of the stem (cf. below, 3.2.3).

^b Note that, as has been stated, constellations with a common ending are not considered in the analyses.

The present study, by contrast, seeks to include the entire domain of inflected word classes without analytical subdivision. This notably includes the token-frequent category of pronouns. While Hentschel (forthcoming) focused on language-internal factors — i.e., the differing morphological constellations as criteria for the occurrence of Ukrainian vs. Russian endings — the present analysis correlates general inflectional patterns with sociobiographical aspects. Specifically, it explores to what extent the speakers who provided the language material originated from Ukrainian-speaking or potentially Russian-speaking backgrounds, and were accordingly influenced.

The basis of the quantitative analysis is the *Oldenburg Corpus of Ukrainian Suržyk* (OCUS), which contains slightly more than 730,000 word tokens, cf. OCUS-Doc: Hentschel et al. (2026).²⁷ A total of 316 respondents contributed to the corpus. The present analysis considers only the material that appears in hybrid sentences or utterances — i.e., those containing morphological and/or lexical elements aligned partially with Ukrainian, partially with Russian. Approximately 535,000 word tokens fall into this category. Of these, about 265,000 are inflected word forms in which stems and endings can be differentiated.²⁸ Among these inflected forms, based on token frequency, 41% have Ukrainian stems, 21% Russian, 34%

²⁷ See also the *Introduction* to this volume by Hentschel, Palinska, and Zeller (forthcoming b).

²⁸ Especially in the case of personal pronouns, inflected forms frequently occur that are inseparable into stem (root) and ending. These non-segmentable forms are not taken into account. However, so-called zero endings are consistently considered when an inflected form consists “substantially” of only the stem, while in other positions of the paradigm it does exhibit a “substantial” or “non-zero” ending.

common, and 4% hybrid. For the corresponding endings, the distribution is 40% Ukrainian, 13% Russian, 47% common, and 0% (0.16%) hybrid. Since nearly half of the endings are common — and therefore, like the very rare hybrid endings, offer no insight into the influence of Russian or Ukrainian on Suržyk’s inflectional morphology — only just over half of the roughly 265,000 inflected word forms in the corpus are relevant for the following analyses. More precisely, about 140,000 word forms are analyzed.

The first partial analysis includes all four stem types — provided they combine with Ukrainian or Russian endings. Table 1 shows the percentage share of Ukrainian endings in each case; the difference from 100% represents the proportion of Russian endings. Table 1 is presented in two variants: Table 2 covers material from all 316 respondents, amounting to roughly 140,000 inflected word forms. Table 3 includes only 163 respondents (i.e., 153 fewer), amounting to approximately 105,000 inflected word forms.

The reason for this separate breakdown — where Table 3 is extensionally included within Table 2 — is that the OCUS consists, strictly speaking, of four subcorpora from three research projects (see Hentschel, Palinska, & Zeller, forthcoming and OCUS-Doc: Hentschel et al., 2026), collected between 2014 and 2021. In the first project (and subcorpus) from 2014, to which 85 respondents contributed in family conversations in central Ukraine, the collection of sociobiographical data was only rudimentary, due to the study’s “non-sociolinguistic” design, compared with the two subsequent surveys, which used extensive questionnaires. However, some participants (exactly 16) in family conversations in the third research project, which focused on the Black Sea oblasts of Odesa, Mykolaïv, and Xerson, were only marginally involved. In such cases, no sociobiographical data were collected at all, often because the participants were minors or very elderly individuals. For one respondent from this area — where fieldwork was conducted during the pandemic — the questionnaire was lost. As a result, sociolinguistic analysis is almost²⁹ impossible for these 102 respondents. A further 61 respondents are excluded for quantitative reasons: for the purposes of this study, a minimum threshold of 500 inflected word forms (again, from hybrid sentences) was established for inclusion in order to reduce random fluctuation. Thus, 163 “central” respondents remain.

In the following, specific mention will be made when the analysis includes all 316 respondents (aR). If this is not the case, the analyses refer to the 163 central respondents (cR).

Table 2 *Percentage of Ukrainian endings (and Russian endings^a) differentiated by stem type (all respondents — aR)*

	A	B	C	E	F
Values in % for respondents	all stems	Ukr. stems	Russ. stems	common stems	hybrid stems

²⁹ They can only be included to a very limited extent — e.g., with regard to regional origin and gender, cf. OCUS-Doc: Hentschel et al. (2026) and Hentschel (2024).

mean — AM		88.0	54.2	73.1	77.8
median		88.2	54.8	75.9	80.4
standard deviation — SD		7.6	20.0	14.3	19.2
range		50.0	95.7	80.6	100.0
minimum		50.0	4.3	15.1	0.0
maximum		100.0	100.0	95.7	100.0
N stems (all)	140,192	67,726	25,642	41,392	5,432
% stems (all)	100	48.3	18.3	29.5	3.9
N respondents	(316)	273	289	275	281

a — The value for the Russian endings corresponds to 100 minus the value given for Ukrainian endings.

To begin with the aR group in Table 2, it should be noted that not all 316 respondents could always be evaluated for the four stem types. Some of the peripherally involved respondents did not produce one or several of the stem types (Ukrainian, Russian, ...) with contrasting Ukrainian or Russian endings, but only with common ones (which account for nearly half of the ending tokens), or not at all. Therefore, the number of respondents for the four stem types is significantly lower than the total number of 316 participating respondents and fluctuates accordingly (see the row N respondents). This is different for the respondents (cR) in Table 3:

Table 3 *Percentage of Ukrainian endings (and Russian endings) differentiated by stem type (central respondents)*

	A	B	C	E	F
Values in % for respondents	all stems	Ukr. stems	Russ. Stems	common stems	hybrid stems
mean — AM		89.1	50.8	73.6	75.9
median		88.7	51.1	75.9	77.8
standard deviation — SD		4.8	17.2	12.1	16.0
range		25.1	89.0	69.6	75.0
minimum		74.4	4.3	22.6	25.0
maximum		99.5	93.3	92.2	100.0
N stems (all)	105,592	51,545	19,824	30,253	3,970
% stems (all)	100	48.8	18.8	28.7	3.8
N respondents	163	163	163	163	163

Apart from the number of respondents, the Table 2 and Table 3 show very similar patterns. This means that although only the smaller subgroup of respondents can often be considered for methodological reasons, analytically this has no impact on the findings.³⁰

The measures of central tendency — i.e., both the mean (arithmetic mean, AM) and the median³¹ — show in both tables that Ukrainian stems are, for the average speaker, very stably realized with Ukrainian endings at a rate of 90%. Russian endings thus account for only about 10%. This indicates that the variation for Ukrainian stems is clearly limited, confirming the finding by Hentschel (forthcoming) for selected inflectional morphological constellations across the entire range of inflected word forms.

Hentschel (forthcoming) also identifies a key reason why the value for Ukrainian stems is not even closer to 100%: in a few specific CNG constellations of adjectival and nominal inflection, there is a more or less noticeable tendency for Ukrainian stems to connect with Russian endings — for language-internal reasons.³² Broadly speaking, these general statistical proportions, according to the hypothesis formulated above (H2), indicate the expected quantitative relations for a Ukrainian-based Suržyk: Ukrainian stems are relatively “resistant” to Russian inflectional morphological influence, apart from the structural exceptions mentioned.

For Russian stems, the quantitative ratio is balanced: for the average speaker, roughly half of them have Ukrainian endings, the other half Russian endings. Russian stems are indeed already noticeably adapted to Ukrainian inflectional morphology. Hentschel (forthcoming) also points here to the influence of language-internal factors. However, these factors cannot fully explain the strong variation that lies behind the general numerical balance between Ukrainian and Russian endings.

The common stems display average values more similar to Ukrainian stems: for the average speaker, they occur with Ukrainian endings in approximately 75% of cases (and otherwise with Russian endings) — see the pale-orange shaded cells in the tables. If the large number of common stems and the small number of hybrid stems function predominantly like Ukrainian stems in mixed speech, this too points to a Ukrainian-based Suržyk: these stems, which are linguistically nonspecific — that is, not classifiable as either Ukrainian or Russian — are clearly perceived by most respondents as Ukrainian. The very heterogeneous

³⁰ Note that the difference between the number (N) of stems in Table 2 and Table 3 amounts to roughly 35,000 inflected forms, i.e., about 25% of all relevant inflected stems illustrated in Table 2. Hentschel, Palinska, and Zeller (forthcoming), in the *Introduction* to this special issue, note that the four subcorpora of OCUS — namely Center-Family, Center-Interview, Black Sea-Family, and Black Sea-Interview — are of roughly equal size. Thus, the reduction in the number of inflected forms from 140,000 in Table 2 to 105,000 in Table 3 is largely due to the fact that (almost) no sociobiographical information is available for the Center-Family subcorpus.

³¹ The median is, alongside the arithmetic mean (AM), another measure of central tendency. It is especially preferred to the AM when outliers distort the picture. However, this is not the case here, which is reflected in the very similar values for the AM and the median.

³² Hentschel (forthcoming) offers explanations in terms of Natural Morphology (Mayerthaler, 1981; Wurzel, 1984) or, alternatively, frequency-based explanations in the sense of Haspelmath (2008, 2021).

hybrid stems are excluded from the following discussion. They account for only about 4% of the analyzed word forms.

There is thus strong evidence to suggest that the corpus predominantly reflects an originally Ukrainian-based, or at least strongly Ukrainian-influenced, Suržyk. This does not necessarily mean that there are no speakers of a Russian-based Suržyk among the respondents. They may simply be too few to have a significant impact on the overall quantitative distribution. This should be kept in mind while also examining the respondents' language-biographical background (see below).

In another respect, however, common stems (and hybrid ones) resemble Russian stems — namely in terms of their measures of dispersion: in all three of these stem groups, the differences between the measures of spread between speakers are much greater than among the Ukrainian stems. Two statistical measures are particularly relevant here: First, the standard deviation (SD), which among the central respondents (cR) is two to three times higher than for Ukrainian stems, and among all respondents (aR), three to four times higher. (The latter is due to the inclusion of peripheral respondents, whose smaller data contributions are more subject to randomness — see above.) Second, the range (difference between maximum and minimum). For Ukrainian stems, the range for the data in Table 3 (cR) is only about 25 percentage points, while for the other stem types, it is nearly three times greater — see the green shaded cells in the tables. In Table 2, the range is even higher.

The range correlates with the minimum value of Ukrainian endings among individual respondents (again in %): for Ukrainian stems, this minimum is around 75% (for cR), i.e., quite high; for common (and hybrid) stems, it drops to 25%, and for Russian stems, it tends toward zero. The particularly low minimum values for the last three stem types also indicate that the Suržyk of some respondents is much more strongly influenced by Russian than that of most others. However, the high minimum value of Ukrainian endings on Ukrainian stems casts doubt on the idea that this involves a Russian-based Suržyk — i.e., a Suržyk spoken by individuals transitioning from Russian to Ukrainian.

The bar chart in Figure 1 illustrates the tendencies in the combinations of the three stem types (excluding hybrid stems from here on) with Ukrainian endings — and, indirectly, with Russian endings — among the respondents (cR) in more detail.

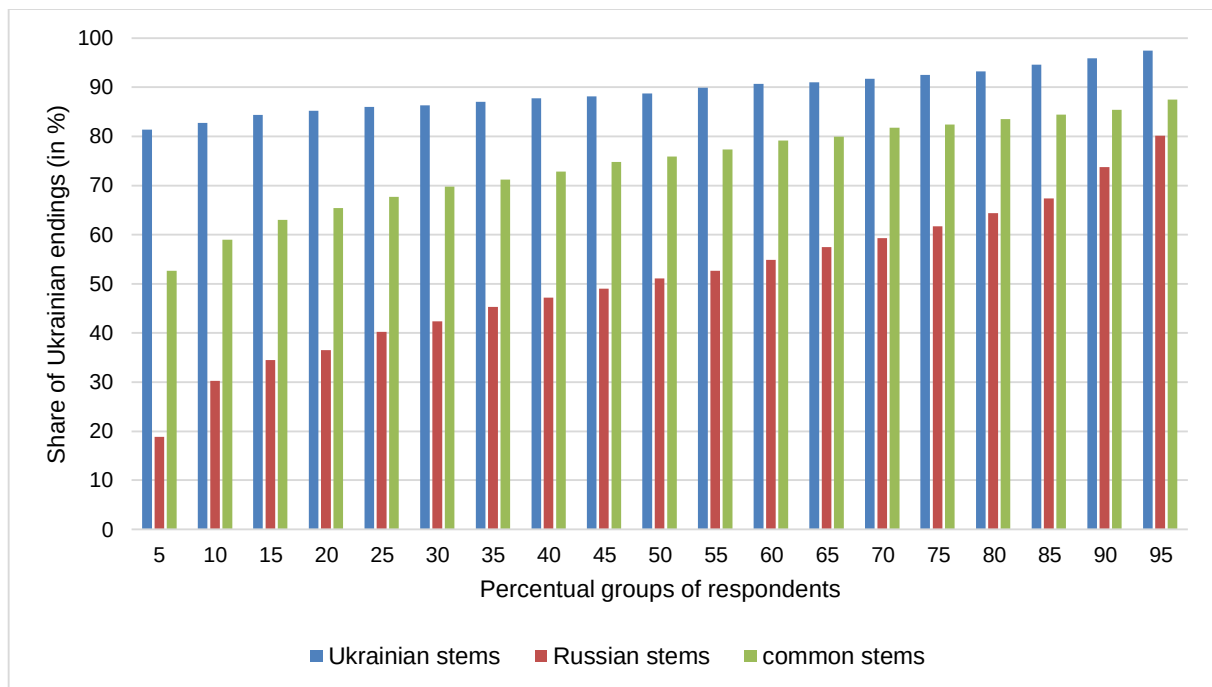


Fig. 1 *Proportion of Ukrainian endings (%) by stem type (cR)*

The bars in figure 1 show the percentage of Ukrainian endings on the three different stem types where such endings occur. The difference to 100% then corresponds again to the proportion of Russian endings. The horizontal axis is divided into 20 (theoretical) equally sized groups of respondents based on the share of Ukrainian endings for Russian stems,³³ each representing 5% of the sample (8–9 out of 163 respondents).

The most compact pattern, closely matching that of Table 1, is seen in the Ukrainian stems. Even the group with the lowest values (the first 5 % of the respondents on the horizontal axis) combines Ukrainian stems with Ukrainian endings in over 80% of cases. The increase is gradual, rising only about 15 percentage points at the far right of the scale, and does so quite linearly.

In stark contrast, Russian stems begin with a value of just under 20% Ukrainian endings on the left side of the scale. On the right, this rises to about 80% — an increase of around 60 percentage points. The bars suggest a shallow, logit curve, with slightly steeper increases on both ends.

The values for common stems lie consistently between those for Ukrainian and Russian stems, typically closer to the Ukrainian values. The rise is correspondingly milder than for Russian stems, but more pronounced than for Ukrainian ones: from just over 50% to just under 90% — more than a 35 percentage point increase.

³³ Note that here the differences are clearer than with Ukrainian and common stems with Ukrainian endings.

The bars on the right-hand side of the diagram suggest a growing adaptation of Russian stems to Ukrainian inflectional morphology, while the left-hand side shows a differentiation in the assignment of endings according to stem origin — i.e., etymological solidarity — not only in the case of Ukrainian stems and endings but also for Russian stems and endings. The latter points to the influence of sociobiographical factors, such as higher proficiency in both languages. However, the fact that even these speakers combine Ukrainian stems with Ukrainian endings at a consistently high rate (over 80%) strongly suggests they are not transitioning from a Russian basis toward Ukrainian. All in all, it must be underlined that none of the respondents combine both Russian stems and Ukrainian stems mostly with Russian endings.

A hierarchical cluster analysis was conducted to identify gradually differentiated types or subtypes of Suržyk speakers — especially to investigate whether a Russian-based Suržyk might exist in at least one, smaller subgroup. Respondents are assigned to the four clusters on the basis of the frequency with which they attach Ukrainian (or Russian) endings to two stem types: Ukrainian and Russian ones. Including the inherently “compromised” common stems in the cluster calculation adds no analytical value; in the table, they are only included for comparison. Table 4 presents the results.³⁴

Table 4 *Clusters based on the percentage of Ukrainian endings by stem type (cR)*

Cluster	A	B	C	D	All
N resp.	18	71	63	11	163
Table 4a Russian stems					
AM	80.4	58.6	39.8	15.7	50.8
SD	6.6	6.6	6.6	7.0	17.2
range	22.8	24.5	25.6	25.1	89.0
minimum	70.5	48.2	24.4	4.3	4.3
maximum	93.3	72.7	50.0	29.4	93.3
Table 4b Ukrainian stems					
AM	91.4	90.0	87.8	86.4	89.1
SD	5.2	5.0	4.0	5.6	4.8
range	15.5	24.0	16.8	15.9	25.1
minimum	84.0	74.4	79.9	78.5	74.4
maximum	99.5	98.4	96.7	94.4	99.5
Table 4c common stems					

³⁴ Since speakers of Suržyk are often (to put it mildly) described as “educationally disadvantaged” by representatives of Ukrainian cultural and academic elites — a view already refuted by Hentschel and Zeller (2017) — a brief overview of the educational background of the 163 respondents is in order at this point of the discussion: One-fifth have completed or partially completed a so-called middle school. Just under one-third have completed or are attending a further vocationally oriented school. About half have completed or are attending a university. Between the four clusters, only for the two larger ones, B and C, is there a slight trend toward better education in Cluster C; for example, 48% in Cluster B fall into the “university” category, compared to 57% in C. This difference can be disregarded, as no connection with the focus of this study — the inflectional morphological adaptation — can be established in the analyses to follow.

AM	80.0	78.9	70.8	43.4	73.6
SD	7.1388	6.5375	8.8284	12.6320	12.1
range	30.2	29.5	33.2	37.8	69.6
minimum	60.5	62.7	53.7	22.6	22.6
maximum	90.7	92.2	86.9	60.4	92.2

The analysis was set to a four-cluster solution.³⁵ It resulted in a relatively small number of respondents in the extreme groups A and D (18 and 11, respectively), and a large number in the middle clusters B and C (71 and 63, respectively).

The starkest contrasts emerge with the Russian stems, as the values from Table 3 and Figure 1 already suggest. The proportion of Russian stems with Ukrainian endings decreases consistently — by slightly more than 20 percentage points per step — from 80% in Cluster A to 15% in Cluster D. Maximum and minimum values follow the same trend.

The values for Ukrainian stems are quite different: the fourfold division is hardly discernable, neither in means nor in minimum and maximum values. While Ukrainian stems were included in the cluster analysis in exactly the same way as the Russian ones, their influence is subtle due to their high internal homogeneity. This manifests in the fact that the strong differentiation between the clusters in the case of Russian stems is not mirrored by complementary groupings for Ukrainian stems. The minima of one cluster always fall just below the maxima of the adjacent one to the right — indicating that for respondents with highly similar values on Russian stems, their value for Ukrainian stems can determine cluster assignment.

Common stems behave similarly to Ukrainian ones. Clusters A through C show very little difference — if any at all. Cluster D stands out: while in the other three clusters, common stems are treated much like Ukrainian ones (Ukrainian endings appear in around 75% of cases), this drops to under 50% in Cluster D.

Thus, the clusters can primarily be interpreted as groups of respondents according to the extent to which they adapt Russian stems to Ukrainian inflection, as shown in Table 5.

Table 5 *Clustered characterization of respondents based on Ukrainian ending shares (%) by stem type (cR)*

Cluster	Adaptation of Russian stems to Ukrainian inflection	“Solidarity” between Ukr. stems — Ukr. endings	Common stems rather as
A	very strong	very strong	Ukrainian
B	not very strong		Ukrainian
C	rather weak		Ukrainian
D	very weak		intermediate Ukr.-Russ.

³⁵ In a preliminary phase, analyses were also conducted with three and six predefined clusters. With three clusters, clusters B and C were merged into one large group. Clusters A and D, with 18 and 11 respondents respectively, remained unchanged. With six predefined clusters, the two central ones were retained, while the two peripheral ones were each subdivided further. In other words, regarding the peripheral groups, neither the three- nor the six-cluster model offers an advantage for generalizing conclusions.

The fact that none of the clusters significantly deviates from the strong and stable alignment between Ukrainian stems and Ukrainian endings argues against the presence of a group of respondents using a Russian-based Suržyk. This holds true even for Cluster D. The specific feature of this group is that the etymological solidarity between Russian stems and Russian endings is also very strong. In part, this is reflected — exclusively here — among common stems as well.

The vast majority of the Suržyk speakers in the sample — about four-fifths — belong to the middle clusters B and C, where a notable decline in the adaptation of Russian stems to Ukrainian inflection is already observable. The two means lie around 50%. However, this does not imply that these speakers randomly alternate between Ukrainian and Russian endings. Rather, it is largely attributable to the structural differences already mentioned between CNG and (less so) TNG/P constellations: 5 out of 25 constellations tested show a share of Ukrainian endings with Russian stems of more than 90%, another 5 of more than 80%, whereas 8 constellations show a share of maximally one third (see Hentschel forthcoming for details).

4.2. The previous results in relation to the respondents' sociobiographical characteristics³⁶

Table 6 illustrates the correlations between the types (clusters) of speakers with gradually differing tendencies towards inflectional morphological adaptation of Russian stems to Ukrainian inflectional patterns on the one hand (cf. Table 4) and various (mostly) sociobiographical parameters on the other.

Table 6 *Suržyk speaker types in relation to selected other, mostly sociobiographical data (cR)*

cluster / type of speaker:		very strong N=18 A	rather strong N=71 B	rather weak N=63 C	very weak N=11 D	Spearman's rho
criterion		AM	AM	AM	AM	
share of Ukrainian endings in % with Russ. stems (out of sum of Ukr. & Russ. stems)		80.4	58.6	39.8	15.7	target variable
share in % of Ukr. stems (out of sum of Ukr. & Russ. stems)		81.9	71.3	61.0	39.5	.578
degree of exposure to Russ.: between 0 and 31		6.2	9.6	12.9	18.5	-.449
competence in Russian: between: none (1) excellent (10)		6.3	7.3	8.2	8.6	-.317
frequency of use of Russian: from never (1) to all the time (5)	by the respondent	2.0	2.8	3.4	3.6	-.438
	in the family	1.8	2.6	3.0	3.4	-.323
	in other contexts	2.3	3.4	3.8	4.1	-.332
frequency of use of Ukrainian: from never (1) to all the time (5)	by the respondent	4.0	3.7	3.2	3.5	.212
	in the family	4.1	3.6	3.0	2.6	.312

³⁶ A comprehensive description of the respondents according to sociobiographical criteria, as well as details on the conditions of data collection, can be found in OCUS-Doc: Hentschel et al. (2026).

	in other contexts	3.9	3.5	3.2	3.0	.222
frequency of use of Suržyk: from never (1) to all the time (5)	by the respondent	4.2	4.5	4.3	4.1	.079
	in the family	4.1	4.6	4.4	4.2	.147
	in other contexts	4.1	4.4	4.1	4.1	.205

The (dark or light) orange-highlighted row represents the target variable: the degree of inflectional morphological adaptation of Russian stems to Ukrainian, treated as the dependent variable. The mean values (AM) of the percentage of Russian stems with Ukrainian endings (from the sum of Ukrainian and Russian stems) for the four groups are repeated here from Table 4 for clarity.

In the rows below, the correlation between the respective parameter and the target variable is examined. The far-right columns display the rank correlation coefficient according to Spearman — Spearman's rho³⁷ — as a measure of the quantitative correlation. For the reader less familiar with statistics, the following should be noted: if the value for a variable were 1, the rankings of speakers for each of the two criteria (bivariate) would be identical. Clearly, this is not the case. A value of 0 would indicate a complete lack of correlation: total randomness. Of note here are the values that range from 0.25 and above, and especially from 0.5 and above. According to Bortz and Schuster (2010, pp. 217–218), values above 0.25 indicate a moderate correlation, and values above 0.5 a strong one.³⁸

(a) The first test variable is the “percentage of Ukrainian stems” from the total number of Ukrainian and Russian stems, which is, of course, not a sociobiographical one. The values (blue-highlighted) here show the same tendency as those of the dependent variable, though the differences between the four groups are less pronounced. The correlation coefficient reveals a strong relationship: $r = .578$,³⁹ the highest value shown in the table. This indicates that speakers that use more Russian lexemes, i.e., stems, also tend to use more Russian endings on Russian stems: the abovementioned tendency toward etymological solidarity between stem and ending. Furthermore, the declining values (from left to right) indicate as well that speakers with a decreasing tendency to adapt Russian stems to Ukrainian inflection show lower shares of Ukrainian stems as well, although the decrease here is moderate compared with the decrease in the values of the target variable.

³⁷ This correlation model, based on the ranks of values (not on the values themselves), was chosen because several of the included variables do not follow a normal distribution.

³⁸ The grouping of coefficients by strength levels varies among authors in statistical literature. Some consider a correlation to be “strong” only from a value of 0.6 upwards, others from 0.7. The scaling of strength ranges from three to six levels. However, it is always emphasized that such gradations must be determined based on subject matter and disciplinary criteria, and that the levels merely serve as interpretive aids. In the social sciences, the thresholds chosen here appears to be standard (cf. Cohen, 1988, pp. 79). As will become clear from what follows, much stronger correlations are not to be expected.

³⁹ Statistical significance values are omitted here. Unless otherwise noted, the results are either significant at the 1% level or highly significant at the 0.1% level — far from indicators of randomness. The only non-significant value in Table 6, which approaches 0 and therefore indicates no correlation, is set in italics.

(b) The next row (dark or light red-highlighted) represents a complex sociobiographical index indicating the respondents' "degree of exposure to Russian", whose composition will be outlined below. Suffice it for the moment to say that this index yields the second-highest correlation value with the dependent variable — the degree of inflectional morphological adaptation of Russian stems to Ukrainian: $r = -.449$. The negative sign indicates an inverse relationship: the higher the respondents' exposure to Russian, the stronger their resistance to adapting Russian stems morphologically to Ukrainian, as one would expect. This is particularly clear in the mean (AM) for Cluster D, the smallest cluster, which differs noticeably from Clusters C, B, and A.

If a higher degree of exposure to Russian among respondents leads to a lower frequency of Ukrainian inflectional endings and a greater frequency of Russian ones, the same might be expected for Ukrainian stems. However, this is not the case, as already suggested by the mean values in Table 4. The correlation coefficient calculated here (not presented in Table 6) shows that there is not even a minimal correlation: $r = -.039$.

The next four independent variables are included within the mentioned index (and therefore highlighted in the same color as the latter), and their composition will be explained now:

(c) The ability to distinguish between Ukrainian and Russian inflectional patterns should correlate with good competence in both languages. Respondents were asked to self-assess their competence on a 10-point scale from "none" to "excellent." The most important quantitative findings are that only 26 of 163 respondents rate their competence in Russian as higher than their competence in Ukrainian; 66 rate both competences as equal, and 71 consider their competence in Ukrainian to be stronger. If we define "good knowledge" as level 7 or higher, then about two-thirds of respondents claim such a level for both languages. The correlation between "competence in Russian" and the target variable has the value $r = -.317$; it is a moderate correlation — a "weakly moderate" one, so to speak. Values for Clusters D and C are close to each other. By contrast, competence in Ukrainian (not presented in Table 6) shows only a very weak, negligible correlation of $r = -.108$ with the target variable. There are no clear differences among the four clusters here, which aligns with the similar, high values for Ukrainian stems with Ukrainian endings across all clusters. Interestingly, the sign is also negative, as with Russian competence. At the very least, it suggests that stronger Ukrainian language skills do not promote the adaptation of Russian stems to Ukrainian morphology.

(d) The next three independent⁴⁰ variables concern the "frequency of use of Russian": (i) by the respondent personally across all communicative contexts; (ii) within their family; and (iii) in their external environment. The first phenomenon to be noted is that the mean value AM is highest for Cluster D, context

⁴⁰ Naturally, the variables referred to here as "independent" are not entirely independent from each other. Such dependencies will be addressed only to the extent necessary. They are independent only in the sense that they are causally independent from the frequency of Russian endings, which is potentially dependent on them.

outside the family (iii) with 4.1 on a 5-point scale, about 0.5 points higher than the average for the respondents themselves (i), which in turn is slightly higher than the mean for the family (ii). In contrast, correlation values for the tendency to adapt Russian stems to Ukrainian inflection align more closely with the self-reported frequency of Russian usage by the respondent: $r = -.438$ — a “stronger” moderate correlation. The frequency of Russian usage in the family and general environment shows “weaker” moderate correlations: (ii) $r = -.323$, (iii) $r = -.332$.

The respondents’ self-estimated frequency of their Russian usage, thus, seems to play a crucial role, which is illustrated in Table 7 in comparison with their usage of Suržyk and Ukrainian:

Table 7 *Frequency of usage of the three codes and degree of inflectional adaptation of Russian stems to Ukrainian*

degree of inflect. adaptation of Russ. stems to Ukr. (grouped)	Russian (in % of respondents)					respondents	
	never	seldom	from time to time	often	all the time	%	N
very strong	27.8	44.4	27.8			100	18
rather strong	10	28.6	38.6	18.6	4.3	100	70
rather weak	4.9	14.8	26.2	39.3	14.8	100	61
very weak		36.4	9.1	18.2	36.4	100	11
	Suržyk						
very strong			22.2	38.9	38.9	100	18
rather strong		2.8	8.5	21.1	67.6	100	70
rather weak	1.6	1.6	14.8	29.5	52.5	100	61
very weak			27.3	36.4	36.4	100	11
	Ukrainian						
very strong	5.6	5.6	5.6	50	33.3	100	18
rather strong	2.9	8.6	20	51.4	17.1	100	70
rather weak	1.6	27.9	27.9	32.8	9.8	100	61
very weak	9.1	9.1	36.4	18.2	27.3	100	11

The figures in Table 7 allow for the following observations in addition to the correlations outlined in the comments on Table 6:

- (a) The usage of Russian: Respondents with a very strong tendency to adapt Russian stems to Ukrainian inflection use Russian at best from time to time. Following the degree of the tendency to inflectional adaptation, the usage of Russian becomes gradually more frequent.

- (b) The usage of Suržyk: By far the most frequent usage of Suržyk can be observed with the two groups with a rather strong or rather weak inflectional adaption, the largest groups. Between 80 and 90% of the respondents here use the mixed code often or, most of them, even all the time. To be precise, the “peripheral” groups do use Suržyk often as well: more than 70% at least often.
- (c) The usage of Ukrainian: Ukrainian exhibits the highest values in the two groups with a rather strong and, especially, very strong tendency for inflectional adaption of Russian stems to Ukrainian: with about 70 to 80% using it at least often. The two groups with a weaker tendency to inflectional adaption show values of about 40 to 50% of a frequent or steady usage of Ukrainian.

All in all, the inflectional adaption of Russian stems is to some degree dependent on the frequency of the respondents’ use of Ukrainian and Russian. However, one has to bear in mind that this holds only for those inflectional constellations that exhibit a variation between Ukrainian and Russian endings, not for those that more or less always allow for Ukrainian or (rarely) Russian endings (cf. Hentschel forthcoming).

I return to Table 6. It also includes data on the frequency of Ukrainian and Suržyk use across the four respondent clusters, again broken down into contexts involving the respondent themselves, their families, and their broader environment. Three observations should be highlighted here: (a) Only the degree of Ukrainian present in the family shows a moderate correlation with the morphological adaptation of Russian stems to Ukrainian, suggesting it promotes adaptation slightly. (b) The values for Suržyk usage show only weak or no correlations — least of all for the respondents themselves. (c) The twelve mean values for Suržyk usage across respondent, family, and environment in each of the four clusters range from a minimum of 4 (frequent), with a clear tendency toward 5 (constant), especially in the middle Clusters B and C. By contrast, for Russian and Ukrainian, only one or two of the corresponding twelve values reach 4 or higher, and almost half are below 3 (occasional). The three projects that collected the material all fundamentally aimed to capture the linguistic form of Suržyk between Ukrainian and Russian. Language data were collected from respondents who, by self-report and/or assessment by our Ukrainian collaborators,⁴¹ clearly exhibit mixed Ukrainian-Russian speech that those collaborators would label as Suržyk. The numerical findings confirm that this goal was successfully met and that the respondents’ self-assessments are realistic.

Here are further clarifications of the complex index “exposure to Russian”: As noted, the index calculation includes stated Russian competence, as well as respondents’ overall Russian usage, in their families, and in their general environment. Additionally, the following features were incorporated: (i) the nationality of the respondent himself and the respondent’s parents and partner (husband/wife); (ii) the presence of Russian during early language socialization before school, as well as the possible presence of Ukrainian and Suržyk; (iii) the use of Russian as the language of instruction (metalinguistic medium) in elementary and middle school; (iv) whether respondents lived outside Ukraine (formerly the Ukrainian SSR) in a Russian-

⁴¹ Special thanks are owed to the Ukrainian colleagues O. O. Taranenko, Kyiv, and T. Kuznietsova, (at the time of data collection) Odesa.

speaking area for more than six months;; (v) whether the Russian, if stated as the primary language used, more closely resembles Ukrainian-accented Russian or standard Russian; (vi) in addition to the scalar self-estimated competence in Russian, whether the respondent is fluent in spoken and/or written Russian; (vii) use of Russian media in terms of watching Russian-language TV and reading Russian texts; (viii) whether Russian, if used frequently, is the only code, one of two, or one of three used⁴² (cf. OCUS-Doc: Hentschel et al., 2026, for details). All these criteria, along with the four from Table 6, were weighted relative to each other and to response options. No single feature correlates more strongly with the degree of morphological adaptation of Russian stems to Ukrainian patterns than the index as a whole.

One might argue over the selection and weighting of questions.⁴³ Nevertheless, as has been and will be shown, plausible correlations emerge between index values and various morphological findings in Suržyk. Figure 2 illustrates the distribution of respondents based on the index.

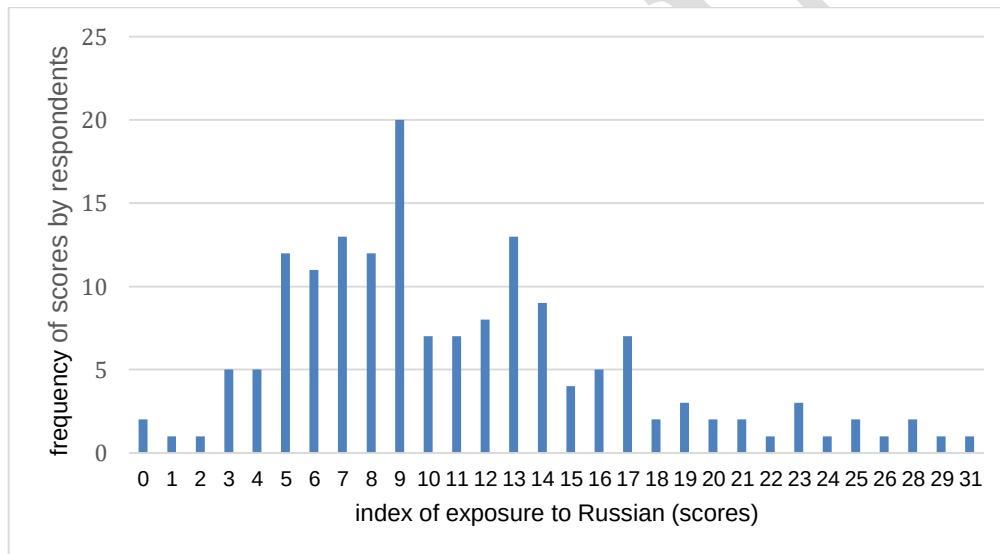


Fig. 2 *Degree of respondents' exposure to Russian*

Figure 2 shows the number of respondents for each point score. The minimum score of 0 is held by two respondents — possibly reflecting a deliberate, politically motivated choice. The maximum achieved score is 31; the theoretical maximum was 44. Even to the naked eye it can be seen that most respondents score between 5 and 17. A hierarchical cluster analysis divides respondents into three clusters. Table 8 illustrates

⁴² This final question was intended as a kind of control for the earlier question regarding the frequency of use of the three codes by the respondents themselves.

⁴³ The list of criteria for weighting the questions and their response options can be found at: https://uol.de/f/3/inst/slavistik/projekte/Ank_UKR_surzhik_v200124.pdf?v=1606812460 (in Ukrainian) and https://uol.de/f/3/inst/slavistik/projekte/Ank_RUS_surzhik_v200124.pdf?v=1606812460 (in Russian).

this by cross-tabulating these three groups with the four clusters based on degree of morphological adaptation.⁴⁴

Table 8 *Grouping of respondents by degree of exposure to Russian*

all values in percent (except for N)		exposure to Russian (grouped)			All %	all N
		weak up to 11	medium: 12–18	strong: min. 19		
degree of inflect. adaptation of Russ. stems to Ukr. (grouped)	very strong	94.4	5.6	--	100	18
	rather strong	66.2	29.6	4.2	100	71
	rather weak	47.6	34.9	17.5	100	63
	very weak	18.2	36.4	45.5	100	11
all		58.9	29.4	11.7	100	163
all N		96	48	19	163	

What immediately stands out is the inverse relationship between values in the columns on the left and the right — i.e., weak vs. strong exposure: a decreasing vs. increasing tendency toward morphological adaptation. At a medium level of exposure, only the strongly adapting group stands out. The group of respondents with a very strong tendency toward morphological adaptation (N=18) consists almost entirely of respondents with weak exposure to Russian. Of course, caution is needed when interpreting percentages with small *n* (as here, 18). Still, the values seen here — not only for respondents showing very strong adaptation but also for those showing rather strong adaptation (N=71) and with strong exposure to Russian (N=19) — fit consistently into the trends within the rows and columns. The highly morphologically adaptive respondents show — cautiously speaking, due to small numbers — a *Suržyk* primarily based on lexical transfer of stems, with virtually no Russian influence on inflectional morphology. All other values confirm that exposure to Russian plays a substantial role in morphological adaptation, though not an absolute one, as indicated by the coefficient $r = -.449$ (cf. Table 6 above).

Figure 1 had already shown that common stems resemble Ukrainian stems more than Russian ones. The four clusters based on morphological adaptation were only faintly visible there. Broadly speaking, one can say that common stems are generally perceived as Ukrainian, not Russian. Nonetheless, exposure to Russian does correlate with the values for adding Ukrainian endings to common stems, albeit less strongly: $r = -.325$ versus $r = -.449$ for Russian stems.⁴⁵ This is due to five (of 25) constructions studied by

⁴⁴ If the cluster analysis is preset to produce more groups, a four-way division results in a small group of four respondents — specifically the four shown at the far left in Figure 3. A five-way division also isolates these four as one group, and separates four respondents at the far right into another. What is interesting here is the following: the four respondents with extremely low scores are among those who show strong adaptation of Russian stems, while the four on the right are distributed among three groups according to the degree of adaptation.

⁴⁵ To readers unfamiliar with statistics, the difference may not seem that clear. However, it is important to note that we are dealing here with correlations of respondent rank positions. That is, the interval between two adjacent ranks is always 1. The differences in the actual values (here, percentage values that determine the ranks) between two consecutive ranks can vary significantly. The same or a similar rank order of

Hentschel (forthcoming) in which Ukrainian stems take on Russian endings to a notable extent. In exactly these cases, the proportion of Ukrainian endings with common stems resembles that of Russian stems, not Ukrainian ones.

5 Inflectional morphology vs. other structural domains

5.1 Phonetics and phonology — or phonics

Earlier, the hypothesis was proposed that the sound patterns of a potentially Russian-based Suržyk would tend to be oriented toward Russian, whereas an undeniably existing Ukrainian-based Suržyk would display more Ukrainian sound patterns. However, Zeller (2018, 2022)⁴⁶ already shows that this hypothesis is at least partially problematic. A given phonic contrast that is distinctive for the two standard languages often corresponds to more complex distinctive relations at the dialectal level. This means first and foremost that, as in the case of the lexicon and inflectional morphology, some Ukrainian dialects exhibit features aligned with Russian standard norms rather than Ukrainian ones. Zeller (forthcoming) identifies seven relevant phonic contrasts. Some of these show little variation and a strong fixation on one of the two realizations. Two examples are highlighted here:

- a) “Hard” voiceless alveolar affricate, Ukr. [tʃ] vs. “soft” (palatalized) Russ. [tʃʲ]/[tɕ]: Based on approx. 25,000 tokens, Zeller observes an extremely strong tendency toward the evidently Russian realization: in 85% of cases. A certain, albeit limited, role is played here — just as with many of the inflectional morphological variables examined by Hentschel (forthcoming) — by the morphological stem type. For Russian stems, Russ. [tʃʲ]/[tɕ] occurs in over 90% of cases; for Ukrainian stems, this value is “only” less than 80%. However, dialectal and colloquial developments within Ukrainian itself must be considered. Historically, the soft variant is older and still widespread in Ukrainian dialects (Žovtobryx et al., 1979, 311–315; Matvijias et al., 1984–2001, map 123). This continues to influence the pronunciation habits of some speakers even in Ukrainian standard speech in (audio) media. As early as Zales’kyj (1974), it was suggested that dialectal background and influence from the Russian standard may interact in this context (see further detail in Zeller forthcoming).
- b) Voiced velar fricative, Ukr. [ɦ], vs. plosive, Russ. [g]: Here, very similar patterns are found, but with reversed polarity. Zeller (forthcoming) finds the Ukrainian fricative [ɦ] in nearly 100% of approx. 22,000 tokens, without any distinction between Ukrainian and Russian stems. It is worth noting that [ɦ] is also widespread in southern Russian dialects. However, among the respondents

persons with respect to two variables may exist even if the values for one variable are on average lower than those for the other. This is roughly the case when comparing Russian and common stems — but (practically) never for Ukrainian stems, where minimal differences result in more divergent rank orders.

⁴⁶ In these earlier papers, Zeller’s data set consisted of one part (one half) of the data used for, among other analyses, Zeller (forthcoming).

who contributed to the corpus, there is no evidence of southern Russian dialect influence, so this outcome clearly represents the establishment of a Ukrainian feature.

These two variables alone show that, even in the case of clear phonic contrasts between Ukrainian and Russian, relatively stable characteristics have become established in Suržyk, and spontaneous or even chaotic mixing certainly does not dominate. In the other phonic contrasts examined by Zeller (forthcoming), Ukrainian realizations predominate overwhelmingly. This is true almost without exception (in over 90% of tokens) for Ukrainian stems; only with Russian stems can more or less clear tendencies toward Russian realizations be observed, i.e., some degree of variation.

For the present focus on hybridization from both directions, these findings mean the following: The fact that — aside from the Ukrainian-Russian phonic contrast discussed in a) — Ukrainian stems are almost exclusively realized with Ukrainian phonics supports, analogously to the findings on inflectional morphology, the view that the respondents who contributed to the examined corpus must in principle be considered speakers of a (at least historically) Ukrainian-based Suržyk. At least for some of them, phonic Russian realizations have been established to a certain degree for morphologically Russian stems, with variation depending on the phenomenon in question. Conversely, in the case discussed under a) — where even beyond Suržyk, in colloquial registers of Ukrainian and even in standard idiolects of prominent public figures, a “soft Russian” articulation [ʃʲ]/[tɕ] is found instead of the “hard Ukrainian” [ʃ] — a small number of Suržyk speakers still tend toward the latter, at least in the case of morphologically Ukrainian stems. Thus, even in sound patterns, there are gradual shifts toward Russian starting from a Ukrainian-based Suržyk.

In cases where the situation is not clear-cut — that is, where certain phenomena treated by Zeller (forthcoming) show somewhat more variation for Russian stem types — there appears to be, with one exception, both a (weaker) correlation with the respondents’ exposure to Russian and a clearer correlation with their tendency to adapt Russian stems⁴⁷ to Ukrainian inflection, as Table 9 shows.

Table 9 *Correlations between Ukrainian realizations of phonic phenomena and sociobiographical exposure to Russian as well as the tendency to adapt Russian stems to Ukrainian inflection*

	okanje in Russ. Stems	hard consonant before /e/		hard apical affricate in Russ. Stems
		Russ. stems	common stems	
degree of exposure to Russian	-.343**	-.272**	-.384**	0.023
share of Ukr. endings with Russ. stems	.376**	.504**	.587**	0.001

The occurrence of Ukrainian *Okanje* and the Ukrainian hard realization of consonants correlates (negatively) with moderate strength to respondents’ exposure to Russian: higher exposure somewhat

⁴⁷ As a reminder: Ukrainian stems already predominantly display Ukrainian endings.

decreases the frequency of Ukrainian pronunciation. The correlation is even clearer between these same phonic phenomena and the strength with which Russian stems are morphologically adapted to Ukrainian inflection. The realization of apical affricates — hard Ukrainian [tʃ] or soft “Russian” [tʃʲ]/[tɕ] — does not fit this pattern. The particular character of this phenomenon was discussed above. Its occurrence is conditioned differently than other signs of Russian influence on Ukrainian Suržyk, and it is also widespread in other registers. Table 10 illustrates this as well.

Table 10 *Gradations of inflectional adaptation of Russian stems to Ukrainian and phonic realizations*

	Phenomenon (figures in %)						
adaptation of Russ. stems to Ukr. inflection	okanje		hard consonant before /e/			hard apical affricate in Russ. stems	N resp.
	Russ. stems	<i>Ukr. Stems</i>	Russ. Stems	common Stems	<i>Ukr. Stems</i>		
very strong	84.8	<i>99.8</i>	31.0	92.5	<i>88.9</i>	4.4	18
rather strong	77.0	<i>99.0</i>	19.2	86.7	<i>95.4</i>	13.1	71
rather weak	66.4	<i>96.9</i>	14.6	76.1	<i>95.4</i>	6.7	63
very weak	37.4	<i>89.4</i>	9.1	43.3	<i>90.4</i>	6.7	11
Column	A	<i>B</i>	C	D	<i>E</i>	F	G

Two additional points should be noted (please initially disregard the italicized columns B and E):

(a) Surprisingly, the lowest share of Ukrainian hard articulation of the apical affricate is found among respondents who show the greatest morphological adaptation to Ukrainian. However, the differences are small, and the number of respondent (18) is also small. Thus, this counterintuitive observation can be neglected. Otherwise, the frequency of Ukrainian realizations of phonic variables corresponds to the trend toward morphological adaptation of Russian stems.

(b) Regarding the group with very weak morphological adaptation, the phonic values raise the question of whether these might point to a Russian-based variety. Even in those phenomena where all other respondents still show a clear dominance of Ukrainian articulation (e.g., *Okanje*, hard consonants before /e/), this group reveals at least a slight dominance of the Russian variants (*Akanje*, soft consonants before /e/). However, the values for these same phenomena in Ukrainian stems in columns B and E show that these weakly adapting respondents hardly differ from others. This strongly indicates that even among those who barely adapt Russian stems morphologically and phonically, the Suržyk variant they speak is at least originally — i.e., in their parents’ or grandparents’ generation — Ukrainian-based.

It would be far-fetched to assume that a potentially newly emerged, Russian-based *Neo-Suržyk* would already produce 9 out of 10 phonically “correct” realizations in Ukrainian stems.

5.2 Lexicon

Based on the same corpus used here, Hentschel (2024) analyzed 107 pairs of Ukrainian-Russian translation equivalents (“hyperlexemes”) and determined that 37 of them display the Ukrainian variant in over 80% of the tokens (in this study: class U!), while 29 show the Russian variant (class R!). That is, in 66 of these pairs — clearly more than half — there is a distinct tendency toward, if not the full establishment, of either

the Ukrainian or the Russian variant in Suržyk. This leaves 41 pairs that — when considering the entire area where data were collected — still show greater variation (class U~R).⁴⁸ These findings are summarized in Table 11.

Table 11 *Scope of preference classes of hyperlexemes*⁴⁹

class:	U!	U~R	R!	All
N tokens	41,777	22,668	17,663	82,108
N hyperlexeme	37	41	29	107
mean: token / hyperlexeme	1,129	553	609	767

This raises the question of the interrelations between the findings based on inflectional morphology — that is, within a relatively closed subsystem of linguistic structures — and the lexical findings, i.e., the findings in the most open linguistic “subsystem”. Since dependencies between respondents’ inflectional morphological preferences and their level of exposure to Russian have been identified, their lexical preferences have also been analyzed.

These findings are illustrated in Table 12. The results are presented in separate cross-tabulations for the three hyperlexeme classes (U!, U~R, R!). For each class, the four respondent groups with varying degrees of adaptation of Russian stems to Ukrainian inflectional patterns are differentiated.

Table 12 *Lexical and inflectional-morphological preferences*

		Share of respondents grouped in “inflectional adaption groups” as to their tendencies for using either Ukrainian or Russian lexemes					respondents N = 163	
Hyperlex. class	U!	< 20%	< 40%	< 60%	< 80%	> 80%	total %	total N
Adapta- tion of Russ. stems to Ukr. infl.	very strong	--	--	--	5.6	94.4	100	18
	rather strong	--	--	--	9.9	90.1	100	71
	rather weak	--	1.6	22.2	31.7	44.4	100	63
	very weak	--	27.3	45.5	27.3	--	100	11
Hyperlex. class	U~R	< 20%	< 40%	< 60%	< 80%	> 80%	total %	
Adapta- tion of	very strong	--	11.1	27.8	44.4	16.7	100	18
	rather strong	25.3	26.8	29.6	12.7	5.6	100	71

⁴⁸ In fact, Hentschel (2024) did not work with a three-level scale, but with a seven-level one. The two levels with the clearest tendency toward Ukrainian or Russian, respectively, have been grouped here as U! and R!. The three intermediate levels, which show greater variation between Ukrainian and Russian lexemes (translation equivalents), are included here in class U~R.

⁴⁹ The last row of Table 11 shows that the average token frequency in class U! is also twice as high as in the other two classes. However, these high values are due to two or three particularly frequent hyperlexemes, especially *buty* = *byt* ‘to be’. Hentschel (2024) shows that there is no evidence of a general tendency for a high-frequency hyperlexeme — or the corresponding Ukrainian lexeme — to be protected from being replaced by a Russian translation equivalent in Suržyk.

Russ. stems to Ukr. infl.	rather weak	47.6	23.8	14.3	14.3	--	100	63
	very weak	45.5	45.5	9.1	--	--	100	11
Hyperlex. class	R!	< 20%	< 40%	< 60%	< 80%	> 80%	total %	
Adaptation of Russ. stems to Ukr. infl.	very strong	88.9	11.1	--	--	--	100	18
	rather strong	88.8	7.0	4.2	--	--	100	71
	rather weak	96.8	3.2		--	--	100	63
	very weak	100.0	--	--	--	--	100	11

For hyperlexeme class U!, the first row shows a value of 94.4 % for the 18 strongly adapting respondents in the column “> 80%”. This means that of the 18 respondents, 94.4% (i.e. 17) realize hyperlexemes of class U! using the Ukrainian variant in more than 80% of instances of usage.⁵⁰ Only one respondent, or 5.6% of this group, realizes U!-hyperlexemes with the Ukrainian variant (lexeme) with a value between 60% and 80%. Very similar values are observable among respondents on the next lower level of inflectional adaptation. The picture becomes more diffuse in the next two rows of class U!

This diffusion intensifies in the U~R class. However, a clear trend remains (the higher values, indicated by darker shading, increasingly shift to the left): Particularly notable are the values for respondents who hardly adapt Russian stems to Ukrainian inflection. Nine out of ten of these respondents show a clear tendency to use the Russian variant of the hyperlexemes — i.e., they are also strongly oriented toward Russian lexically.

This becomes even more pronounced in the R! class. Regardless of the respondents’ tendency to adapt Russian stems to Ukrainian inflection, at least about 90% of them realize these hyperlexemes using the Russian variants. Thus, the latter, i.e., these Russian lexemes, have to be seen as already deeply established loans in Suržyk, if not in (at least spoken) Ukrainian in general.⁵¹

Finally, the calculation of the correlation (Spearman’s rho) between the tendency of respondents to adapt Russian stems to Ukrainian inflectional patterns on the one hand and their tendency to use Ukrainian lexemes and not the Russian translational equivalents yields a high value: $r = .567^{**}$ and thus shows the strength of this correlation. So one can formulate the dependency relation as follows: A strong preference for Ukrainian lexemes over their Russian translatory equivalents entails a strong inclination to adapt Russian lexical imports to Ukrainian inflectional patterns.

⁵⁰ Of course, percentage representations based on small N — such as 18 here, or 11 for the very weakly adapting respondents — are somewhat problematic. Nevertheless, they offer the advantage of better comparability with the larger groups, i.e., the strongly adapting group with 71 respondents and the moderately adapting one with 63.

⁵¹ This must be left to further research, which, unfortunately, seems to be impossible in the nearer future due to the war.

6 Sociobiographical additions to the central question of this study regarding bidirectional hybridization

As already mentioned above, the eleven respondents who scarcely combine Russian stems with Ukrainian endings are the most likely candidates to be considered as speakers of a Russian- rather than a Ukrainian-based Suržyk. Table 11 also clearly shows that these respondents display a very strong lexical preference for Russian elements.

However, a counterargument has already emerged in the section on inflectional morphology: even these respondents show no notable tendency to combine Ukrainian stems with Russian endings — which would be expected from original speakers of Russian. The sociobiographical information provided by these respondents leads to the following observations:

Nationality: Only one identifies as Russian in terms of nationality for themselves and both parents; another states their father's nationality as Russian. One of the respondents has a spouse with Russian nationality.

First language: Five respondents report Suržyk as the code of early childhood socialization, one of whom also mentions Russian. Two respondents each name only Ukrainian or only Russian, one names both languages, and another lists Suržyk and Russian.

Mother tongue: Only one respondent names Russian as their mother tongue, but does so alongside all three codes.⁵² *Languages / codes currently used frequently:*

Six respondents name Russian as a frequently used code, nine name Ukrainian — but always in conjunction with at least Suržyk; in six cases, all three codes are mentioned. These figures show clearly that no quantitatively relevant shift from Russian to Ukrainian can be observed even among these few respondents exhibiting relatively strong influence of Russian.

The eleven out of 163 respondents who show the strongest orientation toward Russian both in inflectional morphology (only for Russian stems) and lexically do not correspond to the stereotypical image of a speaker of a potentially Russian-based Suržyk — i.e., a Russian speaker turning to Ukrainian and mixing the languages in the process. Nor do these data support the assumption of a Russian-based Suržyk.

What can be observed, however, are sociobiographical characteristics of these eleven Suržyk speakers who are more strongly influenced by Russian, which make a gradual shift in favor of Russian components in Suržyk plausible, for example: (i) Eight respondents come from oblasts where Hentschel and Taranenko (2021, 295) observed the strongest presence of Russian in the survey area (Xarkiv, Odesa, Mykolaïv, Xerson). (ii) Six attended primary and secondary schools where the language of instruction was exclusively or predominantly Russian. (iii) Seven had completed higher education. In this context, it should be noted that nine of the eleven respondents are 40 years or older, meaning many were likely educated at institutions where Russian remained the dominant language of instruction, particularly in the oblasts they came from.

⁵² For the interpretation of respondents' statements about their "mother tongue" in Ukraine, see Hentschel and Palinska (2022a).

(iv) Only one respondent reports that Russian is rarely used in their environment; seven report frequent use, and three say it is used constantly.

7 Conclusions

In light of the foregoing, can we speak of a “hybridization from both sides” in today’s Ukraine (prior to the Russian invasion), understood as the coexistence of an old, generationally transmitted Ukrainian-based Suržyk and a new Russian-based Suržyk that emerged after 1990? The answer is obviously no.

No individuals could be identified from the sociobiographical profiles of the respondents for whom a Russian-based Suržyk, hypothetically and prototypically outlined by Flier (2008), applies: clearly Russian-speaking individuals who switched to Ukrainian. As noted above, sociobiographical metadata are available for 214 of the 316 respondents who contributed to OCUS. Of these, 203 answered all questions concerning which of the three codes they used frequently (if not very frequently) during childhood and in adulthood. Among these 203 respondents, only 16 reported that exclusively Russian (12) or Russian and Suržyk (4) were spoken at home during childhood, but not Ukrainian. Only one of them did not name Russian as one of the codes frequently used in adulthood, whereas 15 added Ukrainian to their regularly used linguistic repertoire. Conversely, 68 respondents did not name Russian, but Ukrainian (50) or Ukrainian and Suržyk (18), as the codes most frequently used during childhood. Only three of them do not report Ukrainian as part of their adult linguistic repertoire, whereas 39 additionally include Russian, in 35 cases alongside Ukrainian.

The majority of respondents (111) reported Suržyk as their sole code in childhood. Twelve of them report the same for adulthood, while all but three include it alongside Ukrainian and/or Russian as a frequently used code. Almost half (49) report that, as adults, they frequently use both Ukrainian and Russian in addition to Suržyk; 32 report using only the former and 15 only the latter.

The dominant pattern in the respondents’ linguistic biographies is thus the expansion of their linguistic repertoires rather than the abandonment of childhood codes. In most cases, this expansion involves the addition of the “cultural” codes Ukrainian and Russian (the latter less frequently). Notably, almost all respondents who did not report Suržyk for childhood do report it as a code acquired later in life: 60 out of 62.⁵³ In these cases, the Suržyk in question was most likely Ukrainian-based, as a Russian-based Ukrainian-

⁵³ Note that this is, to some extent, an artefact of the design of the analysis. A condition for participation was that respondents used Suržyk regularly, not merely “sometimes,” at the time of data collection. It may often have been the case that parents spoke with their children in a more “cultural” code in order to help them secure an advantageous position in society. However, when a low variety is present in the surrounding environment, children may acquire it from others, whether peers or adults, already in childhood or later. A comparable situation can be illustrated by my own experience: in a village in northern Germany in the 1950s, my parents and grandparents did not speak Low German to me (only Standard

Russian mixed speech, before and shortly after 1990, was at best an individual phenomenon with single speakers in specific social circumstances, e.g., marriage of a Russian speaking person into a Ukrainian-speaking family.

Indeed, the assumption that Russian-speaking Ukrainians (i.e., Ukrainian citizens) shift to Ukrainian while developing a Russian-based Neo-Suržyk is overly simplistic. Such a view fails to take into account how — or from which code — they or their parents (or grandparents) previously shifted to Russian. In many cases, they are unlikely to have been Russian-speaking for generations, but rather to have transitioned from a Ukrainian-based Suržyk to Russian. To some extent — though varying widely across individuals — they may have retained this Suržyk as a family code. Without doubt, this code is likely to have become “more Russian” over time. The extent to which a full “turnover” of the base language has taken place in such cases remains unclear. It may be precisely such developments that account for the small group (approx. 10) of respondents in this study who exhibit a very strong influence of Russian, yet without giving rise to a clearly identifiable Russian-based Suržyk. However, this cannot be determined on the basis of the available sociobiographical data.

In Ukraine, speakers oriented toward the Russian language clearly participate reluctantly in surveys on language use and attitudes, and are even more hesitant to have their (possibly mixed) speech documented and analyzed for academic purposes. This experience is echoed by Paweł Levchuk (personal communication), based on his extensive data collection on the languages of Ukrainian war refugees in Poland (see Levchuk, 2024). It is theoretically possible, then, that the data used in this and the other analytical studies in this volume do not sufficiently capture the potential group of originally Russian-speaking citizens who have turned toward Ukrainian. What does this imply for answering the central research question of the project on which this and many other contributions in this volume are based?

This introduces a theoretical issue: What is meant by the term *Suržyk*? Already the initial project whose data entered the OCUS corpus (see *Introduction*: Hentschel, Palinska, and Zeller, forthcoming b) operated on the hypothesis that a code which (1) is based on a widespread practice of mixing two (or more) codes over at least several decades, (2) is used in communication between the “mixers” themselves (not only with speakers of the code to be approximated), and (3) is passed on to the younger generation through early language socialization, should exhibit phenomena of structural or habitual stabilization — that is, a reduction of variation across many variables at different structural levels. This aligns with Trudgill’s (1986) findings on new mixed dialects in the English-speaking world, where stabilization sets in by the third or fourth generation. Mixtures between structurally similar and genetically related languages show similar mechanisms to dialect or variety mixing *within* a single language (see Thomason & Kaufman, 1988; Muysken, 2000 on “congruent lexicalization”). For the Ukrainian-based Suržyk, this has been demonstrated

German), but used it partially among themselves. My active competence developed primarily through conversations with neighbors, which was facilitated by my familiarity with the “sound” of Low German.

— based on extensive material — in several analytical studies in this volume, refuting the assumption made by many proponents of so-called “patriotic linguistics” (see Hentschel, 2017, pp. 26–29) that the co-occurrence of Ukrainian and Russian elements in Suržyk is absolutely spontaneous, unpredictable, i.e. chaotic.⁵⁴ A comparable scenario for the emergence of a hypothetical Russian-based Suržyk, with similar stabilizations, i.e. reductions of variation as in the “old” Ukrainian-based Suržyk is highly unlikely over a mere 35 years of independent Ukraine.

In conversations with Ukrainians, one nevertheless does hear that some families — particularly in highly educated circles — do make such a shift, perhaps initially out of economic necessity (e.g., in public service, including the education sector), and increasingly since 2014 or especially 2022 due to political rejection of the aggressor’s language. Numerous reports available online point to a marked shift from Russian to Ukrainian following the Russian invasion, whereas prior to 2014 or 2022, such a shift was at best tentative. In other words, there may be numerable “new speakers” of Ukrainian.⁵⁵ To what extent such recent survey results on a language shift from Russian to Ukrainian can be considered reliable under wartime conditions and whether the use of Russian ceases entirely *among* those who now identify with Ukrainian, and whether we are witnessing a lasting change in language behavior, remains unclear — if not written in the stars. Regarding today’s linguistic situation in Ukraine, one can only hope that these uncertainties are not ruled by the “stars” atop the Kremlin. At least according to studies like (Hentschel & Brüggemann, 2015), during the first two decades of independence, there was no indication of widespread language conflict among the population (in contrast to political and cultural spheres), nor of a mass moral rejection of Russian. Instead, there was a discernible desire to gradually improve the status and support of the Russian language in Ukraine (at least in areas with a considerable number of Russian-speaking people), particularly, but by far not exclusively among Russian speakers.

To the best of my knowledge, no systematic empirical studies exist on the structural-linguistic features of new Ukrainian speakers. The assumption that recent switchers from Russian to Ukrainian speak an approximative form of Ukrainian in a similar way as earlier switchers from dialectal Ukrainian to Russian during the Tsarist or Soviet era spoke an approximative and highly variable Russian is highly unlikely. The social, educational, and media conditions in independent Ukraine are completely different to those in the 19th and 20th centuries, until 1990. There is currently no solid empirical evidence for a Russian-based “Neo-Suržyk” akin to the Ukrainian-based one. Moreover, strong theory-based counterarguments exist. Certainly, new speakers of Ukrainian will show interferences from Russian in their approximative

⁵⁴ Moser (2016, p. 51) even claims to be able to determine that “in reality” the Suržyk of person A has nothing in common with the Suržyk of person B, except that neither represents an instance of “clean speech”. In view of the results presented by the authors in this volume and elsewhere, the question arises whether it is not deficits in theory and method and/or purist-ideological dispositions that prevent these authors from “seeing the structure of the forest for all the trees.”

⁵⁵ Cf. Woolhiser (2013) on new speakers of Belarusian in Belarus’ and the recent paper of Kudriavtseva (2025).

Ukrainian speech. But should one name this “Suržyk”? I do not think so, because no stabilization can be expected here.

The Ukrainian-based Suržyk, however, in spite of a considerable degree of stabilization, still shows a certain amount of gradual individual variation based on differences in the frequency of using either Ukrainian or Russian. A small group of “maximally Ukrainian-oriented” speakers do speak Suržyk frequently and, somewhat less often, Ukrainian. However, they only use Russian maximally “from time to time”. The very small group of “highly Russian-oriented speakers”, without doubt, exhibits a strong impact of Russian. Every third respondent here speaks Russian “all the time”, whereas in the other groups, on average, only every tenth respondent does so. Be that as it may, that even among the latter group this is a Ukrainian-based Suržyk, though with a certain dominance of Russian in some substructures of their mixed code, is evident from the fact that variation between Ukrainian and Russian features in inflectional morphology and phonology appears primarily in *Russian* word stems — that is, Russian influence is concentrated there.⁵⁶ Ukrainian word stems remain largely untouched in *all* respondent groupings.

Russian stems additionally show clear adaptation to Ukrainian patterns across multiple phonic and inflectional variables. Nevertheless, there are instances where Ukrainian stems too display strong or even complete Russian influence. Two key factors contribute to this: (1) a substantial structural overlap between local Ukrainian dialects and *Russian* standard language (e.g., infinitive endings or the soft alveolar affricate), and (2) in inflectional morphology, cases where Russian variants are preferable from a functional or structural perspective. In cases of convergence between Ukrainian dialects and Russian standard forms, the question of which factor predominates can only be assessed case by case — and even then, only tentatively (see Hentschel & Palinska, 2022b).

Wherever inflectional variation could be determined — mainly in Russian stems — a third influencing factor emerges: the degree of individual respondents’ exposure to Russian, which is, of course, to be expected.

The Russian influence is, unsurprisingly, much stronger in the lexicon. There is far more frequent reduction in variation in favor of the Russian variant than in phonological or inflectional domains. For many, but not a majority of “competing” Ukrainian and Russian lexemes, the latter are in stable, almost exclusive use. Still, stable usage of Ukrainian variants of these “hyperlexemes” is by no means rare. Given that in some cases the choice of inflectional or phonic variants correlates with stem origin (Ukrainian or Russian), the many lexical transfers from Russian into the (Ukrainian-based) Suržyk can be seen as a “Trojan horse” through which, under specific conditions (dialect–standard convergence, structural weaknesses of

⁵⁶ Matras (2015, pp. 48–66ff.) refers to such donor- and recipient-language-driven distribution patterns between morphemes within a word form as *morphological compartmentalization* in mixed codes. Similarly, Auer (1997) mentions *co-occurrence restrictions* involving elements of sound patterns and at morphological levels in donor/recipient varieties.

Ukrainian endings), Russian features may establish themselves at some structural points in the closed grammatical systems as well.

The traditional Suržyk can therefore — based on structural considerations — be fully understood as a variety of Ukrainian that exhibits significant lexical influence from Russian and displacement of Ukrainian lexemes by Russian ones. Under specific conditions, this lexical influence can even lead to limited establishment of Russian inflectional or phonic features.

References

- Auer, P. (1997). Co-occurrence restrictions between variables. A case for social dialectology, phonological theory and variation studies. In F. Hinskens, R. van Hout, & L. Wetzels (Eds.), *Variation, change and phonological theory* (pp. 71–102). John Benjamins.
- Auer, P. (1999). From code switching via language mixing to fused lects: Towards a dynamic typology of bilingual speech. *International Journal of Bilingualism*, 3(4), 309–332.
- Bilaniuk, L. (2004). A typology of suržyk: Mixed Ukrainian-Russian language. *International Journal of Bilingualism*, 8(4), 409–425.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th ed.). Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Coseriu, E. (1967). Lexikalische Solidaritäten. *Poetica*, 1(3), 293–303.
- Cyxun, H. (1998). ‘Trasjanka’ jak ab”ekt linhvistyčnaha dasledavannja. In L. I. Sjameška, & M. R. Pryhodzič (Eds.), *Belaruskaja mova ŭ druhoj palove XX stahoddzja. Materyjaly mižnarodnaj navukovaj kanferencyi* (Minsk, 22–24 kastryčnika 1997 h.) (pp. 83–89). BDU.
- 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>
- Dzjuba, I. (2012). Rusyfikacija v ukraïns’komu konteksti. In V. A. Smolij (Ed.), *Encyklopedija istorii Ukraïny* (Vol. 9, pp. 372–380). Naukova dumka.
- 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].
- Florinskij, T. D. (1900). *Malorusskij jazyk i “ukrains’ko-rus’kyj” literaturnyj separatizm*. Tipografija A. S. Suvorina.
- Grečko, D., Palins’ka, O. (2019). Slovoobrazovanie v belorussko-russkoj i ukrainsko-russkoj smešannoju reči. *Linguistica Copernicana*, 16, 311–335.
- Grenoble, L. A. (2010). Contact and the development of Slavic languages. In R. Hickey (Ed.), *The handbook of language contact* (pp. 581–597). Wiley-Blackwell.
- Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics*, 19(1), 1–33.
- Haspelmath, M. (2021). Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics*, 57(3), 605–633.
- Hentschel, G. (2013). Zwischen Variabilität und Regularität, “Chaos” und Usus: Zu Lautung und Lexik der weißrussisch-russischen gemischten Rede. In G. Hentschel (Ed.), *Variation und Stabilität in*

Kontaktvarietäten: Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien (pp. 63–99). BIS Verlag. [= *Studia Slavica Oldenburgensia* 21].

Hentschel, G. (2017). Eleven Questions and Answers about Belarusian-Russian Mixed Speech (“Trasyanka”). *Russian Linguistics*, 41(1), 17–42. <https://doi.org/10.1007/s11185-016-9175-8>

Hentschel, G. (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). Inflectional morphology in Ukrainian-Russian code-mixing — On endings and stems of main lexical word classes in Suržyk. *Russian Linguistics*, 50.

Hentschel, G., & Brüggemann, M. (2015). Gibt es in der Ukraine einen ukrainisch-russischen Sprachenkonflikt? Über das Miteinander, Gegeneinander und Durcheinander von Ukrainisch und Russisch. In E. Bakalova, T. Endrich, Kh. Shlyakhtovska, G. Spodarets (Eds.). *Ukraine. Krisen. Perspektiven* (pp. 95–118.). (= *IMPULSE - Studien zu Geschichte und Gesellschaft*, Bd. 6). Wissenschaftlicher Verlag.

Hentschel, G., & Kittel, B. (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, G., & Palinska, O. (2022a). 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., & Palinska, O. (2022b). 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., & Zeller, J. P. (forthcoming). Introduction: background, aims, source material, and brief outlook on this special issue on Ukrainian Suržyk. [Special issue “Suržyk” between Ukrainian and Russian – Hybridization from Two Sides?]. *Russian Linguistics*, 50.

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., & Taranenکو, O. (2021). Bilingualism or tricodicism: 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. (2014). Belarusians’ pronunciation: Belarusian or Russian? Evidence from Belarusian-Russian mixed speech. *Russian Linguistics*, 38(2), 229–255.

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

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

Horbyk, R., & Palko, O. (2017). Righting the Writing. The Power Dynamic of Soviet Ukraine Language Policies and Reforms in the 1920s-1930s. *Studi Slavistici*, XIV, 67–89.

Kappeler, A. (2019). *Kleine Geschichte der Ukraine*. 5., überarbeitete und aktualisierte Auflage. C. H. Beck Verlag.

Karunyk, K. (2017). The Ukrainian spelling reforms, half-reforms, non-reforms and anti-reforms as manifestation of the Soviet language policy. *Studi Slavistici*, XIV, 91–110.

Kloss, H. (1987). Abstandssprache und Ausbausprache. In U. Amman, N. Dittmar, & K. J. Mattheier (Eds.), *Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. De Gruyter.

- Kravciv, B., & Kubijovyč, V. (1973). Rusyfikacija. In V. Kubijovyč (Ed.), *Encyklopedija ukraïnoznavstva. Slovnykova častyna* (Vol. 7., pp. 2643–2652). Molode žyttja. <http://litopys.org.ua/encycl/euii202.htm>.
- Kudriavtseva, N. (2025) New speakers of Ukrainian: Ideologies of linguistic conversion. *Language in Society, First View*, 1–20. <https://doi.org/10.1017/S0047404525101711>
- Levchuk, P. (2024). *Wielojęzyczność migrantów wojennych z Ukrainy w Polsce*. Wydawnictwo Instytutu Slawistyki Polskiej Akademii Nauk.
- Liskovec, I. V. (2006). *Russkij i belorusskij jazyki v Minske: problemy dvujazyčija i otnošenija k jazyku* [Unpublished doctoral dissertation]. European University Sankt-Petersburg.
- Ljapunov, B. M. (1919). *Edinstvo russkogo jazyka v ego narečijax*. Tipografija Južno-Russkogo prosvetitel'skogo obščestva.
- Masenko, L. (2019). *Suržyk: miž movoju i jazykom* (2nd ed.). Vydavnyčyj dim “Kyjevo-Mohyljans’ka akademija”.
- Masenko, L. et al. (Eds.). (2005). *Ukraïns’ka mova u XX st.: istorija henocydu. Dokumenty i materialy*. Vydavnyčyj dim “Kyjevo-Mohyljans’ka akademija”.
- Matras, Y. (2009). *Language contact*. Cambridge University Press.
- Matras, Y. (2015). Why is the borrowing of inflectional morphology dispreferred? In F. Gardani, P. Arkadiev, & N. Amiridze (Eds.), *Borrowed morphology* (pp. 48–80). De Gruyter Mouton.
- Matras, Y., & Bakker, P. (2003). The study of mixed languages. In Y. Matras & P. Bakker (Eds.), *The mixed language debate: Theoretical and empirical advances* (pp. 1–20). Mouton de Gruyter.
- Matvijas, I., Zakrevs’ka, Ja., & Zales’kyj, A. (Eds.). (1984–2001). *Atlas ukraïns’koï movy*. Naukova dumka.
- Mayerthaler, W. (1981). *Morphologische Natürlichkeit*. Akademische Verlagsgesellschaft Athenaion.
- Menzel, T., & Hentschel, G. (2017). *Flexionsmorphologische Irregularität im innerslavischen Sprachkontakt. Sprachinhärente Präferenzen oder politisch-soziale Dominanz: Russisch vs. Weißrussisch / Ukrainisch — Polnisch vs. Kaschubisch / Lemkisch*. Peter Lang.
- Meyerhoff, M. (2002). Communities of practice. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 526–548). Blackwell.
- Moser, M. (2016). “Suržyk” czy “suržyky”? *Ukraïns’ka mova*, 1, 27–54.
- Muysken, P. (2000). *Bilingual speech. A typology of code-mixing*. Cambridge University Press.
- Myers-Scotton, C. (2002). *Contact linguistics*. Oxford University Press.
- Myxal’čuk, O. (2023). Terytorial’ni dialekty v konteksti linhvozaxsnyx praktyk jevropejs’koho arealu. *Naukovyj visnyk Užhorods’koho universytetu: Filolohija*, 2(50), 327–334.
- OCUS-Doc: Hentschel, G., Zeller, J. P., & Palinska, O. (2026). *Documentation of the “Oldenburg Corpus of Ukrainian Suržyk”*. <https://uol.de/slavistik/forschung/sprachwissenschaft/bereich-mischvarieteten/okus-dok>
- Palinska, O., & Hentschel, G. (2022). Regional’nye osobennosti ispol’zovanija ukrainsko-russkoj smešannoï reči (suržika) i vlijanie dialektov: pristavki i predlogi VID / OT. *LingVaria*, 34(2), 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.
- Rusačenko, N. P., & Koxančyk, S. R. (2019). Lemkivs’kyj hovir: problema statusu. *Filolohični studii. Zbirnyk naukovyx prac’*, 12, 73–78.
- Skopnenko, O. I. (2012). Reformy v ukraïns’kij i bilorus’kij literaturnyx movax 1933 roku: liniï rozvytku kriz’ symvoličnu funkciju. *Komparatyvni doslidžennja slov’jans’kix mov i literatur*, 17, 101–110.

- Taranenko, O. (2007). Ukrainian and Russian in contact: attraction and estrangement. *International Journal of the Sociology of Language*, 183, 119–140.
- Thomason, S. G. (2015). When is the diffusion of inflectional morphology not dispreferred? In F. Gardani, P. Arkadiev, & N. Amiridze (Eds.), *Borrowed Morphology* (pp. 27–46). De Gruyter — Mouton.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- Trudgill, P. (1986). *Dialects in contact*. Basil Blackwell.
- Weinreich, U. (1953). *Languages in contact*. Mouton.
- Weinreich, U. (1954). Is a structural dialectology possible? *Word*, 10(2–3), 388–400.
- Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies*. Cambridge University Press.
- Woolhiser, C. (2013). New speakers of Belarusian: Metalinguistic discourse, social identity, and language use. In Ch. Gribble (Ed.), *American contributions to the 15th International congress of Slavists, Minsk, August 2013* (pp. 63–115). Slavica Publishers.
- Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit*. [= *Studia grammatica* XXI]. Akademie Verlag Berlin.
- Zales'kyj, A.M. (1974). Jak treba vymovljaty šypljači zvuky? *Ridne slovo*, 9, 38–45.
- Zeller, J. P. (2015). *Phonische Variation in der weißrussischen 'Trasjanka'*. *Sprachwandel und Sprachwechsel im weißrussisch-russischen Sprachkontakt*. BIS Verlag [=Studia Slavica Oldenburgensia, 27].
- Zeller, J. P. (2018). Zum ukrainisch-russischen Sprachkontakt: Phonische Variation im ukrainischen 'Suržyk' im Vergleich zur weißrussischen 'Trasjanka'. In S. Kempgen, M. Wingender, & L. Udolph (Eds.) *Deutsche Beiträge zum 16. Internationalen Slavistenkongress Belgrad 2018* (pp. 365–374). Harrassowitz Verlag.
- Zeller, J. P. (2022). 'Okannia' and 'akannia' in Ukrainian-Russian mixed speech ('Suržyk'). In *Ukrajins'ka Mova*, 2(82), 38–59.
- Zeller, J. P. (forthcoming). The impact of Ukrainian and Russian on sound patterns in Suržyk. *Russian Linguistics*, 50.
- Žovtobrjux, M. A., Rusanivs'kyj, V. M., & Skljarenko, V. H. (1979). *Istorija ukrajins'koji movy. Fonetyka*. Naukova dumka.