A Novel Typology of Mutually Intelligible Words: The Case of Slavic Languages

Klyshinsky E., Badryzlova Yu.
HSE University
Staraya Basmannaya Str. 21/4 Bld. 1, Moscow, Russia
yuliya.badryzlova@gmail.com, eklyshinsky@hse.ru

Abstract

In this paper, we demonstrate that using the notion of cognates in the task of evaluating mutual intelligibility (MI) in closely related languages can be confusing. We suggest a new term – percipiants – which handles MI of words irrespective of whether they have a common origin. We propose four classes of percipiants,which are differentiated by the degree and type of the closeness of meanings within the word pair. Furthermore, we claim that MI of individual words across a set of languages may be established computationally via normalized Levenshtein Distance (LD). We test our hypotheses by analyzing data from a psycholinguistic experiment in which the respondents had to predict missing words in a text in their native language. In the experimental group, the respondents had access to the text in their native language and in a language from the same language group (Slavic); in the control condition, the respondents were performing the same test in the presence of their native language only. The analysis demonstrates that (a) psycholinguistically, MI may be defined as the difference between the average correctness of answers in the experimental and the control groups; (b) normalized LD may serve as an adequate predictor of the experimentally measured MI; (c) the said MI corroborates the four classes of percipiants; (d) contextual factors weaken the predictive force of LD, requiring further investigation.

Keywords: mutually intelligible words, cognates, percipiants

1. Introduction

The task of assessing mutual intelligibility (MI) of texts written in different genetically related languages has attracted heightened interest in foreign language instruction (e.g. Squires et al. 2020), as well as in studies of bilingualism (e.g. Miwa et al. 2014) and multilingualism (e.g. Stenger et al. 2020). Intelligibility of a foreign language is understood as a situation in which speakers of one language are capable of understanding written or spoken texts in another language without prior training or preparation. Typically, texts in closely related languages are intuitively comprehensible even to speakers with no professional linguistic background, while the intelligibility of distant languages is known to depend on the presence and scope of common borrowings from a third language (or languages).

It is commonly accepted that MI is promoted primarily via cognates. Consequently, a substantial body of research has focused on the impact of cognates bearing orthographic (Gooskens and Heeringa, 2004; Van Bezooijen and Gooskens, 2005) as well as phonetic (Gooskens and Heeringa, 2004; Stenger et al., 2020) similarity. Yet, empirically-based intuition prompts that cognates are not the only factor behind MI. On the one hand, some cognates may be poorly recognized by the recipient: this occurs when a common cognate base attaches language-specific affixes or undergoes substitutions of letters or sounds; when a word is written in a different alphabet, or when the orthographic characters have language-specific diacritics. On the other hand, meaning may be reliably inferred not only for cognates. Namely, MI is possible for words which are not true cognates, yet they possess phonetic or graphic similarity and a shared meaning – though such cases are quite rare, cf. *much* (English) vs. *mucho* (Spanish). (Here, we are setting aside the discussion of the distinction between cognates and borrowings, which is beyond the scope of this paper.)

Tracing etymological roots is typically very challenging, which, in turn, complicates the assessment of MI and impedes its practical application (e.g., in language teaching). Consequently, we consider it reasonable for practical purposes to distinguish between the shared ancestry of words and their MI.

We introduce a new term — *percipiant* — to denote words that are mutually intelligible across languages irrespective of whether they share a common origin; this allows us to shift the focus from similarity of origin toward similarity of comprehension. In order to gain insight into mutual cross-linguistic intelligibility, we analyze the results of an experiment in which native speakers of Russian were to fill in missing words in a Russian text while also being exposed to a parallel translation into one of eight Slavic languages. The MI of a word was computed as the increase in correct answers in the presence vs. absence of parallel texts.

Our results show that percipiants can be divided

Proceedings of Workshop on Dialects in NLP — A Resource Perspective (DialRes) , pages 337–344
16 May 2026.

into four groups based on their similarity, as manifested in the different statistical patterns of respondents' behavior in recovering the missing words. We also demonstrate a correlation between Levenshtein distance (Levenshtein, 1966) and MI; however, we conclude that Levenshtein distance alone does not fully account for the variation in this behavior; contextual factors exert a strong influence as well, thus calling for a closer analysis.

2. Previous work *vs.* need for a novel approach

Research on MI in genetically related languages has largely focused on analysis of cognates. Crystal (2008, p. 83) defines cognates as follows: "A language or a linguistic form which is historically derived from the same source as another language/form, e.g. Spanish/Italian/French/Portuguese are 'cognate languages' (or 'cognates'); *père*/*padre*, etc. ('father') are 'cognate words' or cognates" – thereby stressing the common ancestry. The definition by Squires et al. (2020) states that cognates are "...pairs of vocabulary words that share meaning with similar phonology and/or orthography in more than one language, such as *rose–rosa* (English–Spanish) or *carrot-carotte* (English–French)". According to Miwa et al. (2014), "The words *coffee* /khɔfi/, *koffie* /kɔfi/, and コーヒー /koohii/ ... are examples of cognates" – thus stressing the role of phonetic resemblance for the MI of cognates.

Roembke et al. (2024) conducted a meta-analysis of a large body of studies on MI. They found that "Operationalizations of cognate status varied greatly across studies, though three methods were most common: (1) similarity was quantified by using some kind of norming procedure to obtain similarity ratings (i.e., people either classified word pairs as cognate/non-cognate or rated them on a similarity scale…), (2) (normalized) Levenshtein distance,… or (3) Van Orden's graphemic similarity algorithm (Van Orden, 1987)". Method 1 mentioned in the meta-analysis indicates that it is not uncommon in research practice to establish cognates via the perceived intelligibility judged by human subjects, whereas common historical origins are not taken into account.

Levenshtein distance (LD) is one of the most commonly used metrics for measuring the orthographic similarity of words. MI is often analyzed via its correlation with LD (e.g. Van Bezooijen and Gooskens 2005). Thus, LD was shown to adequately quantify orthography-based similarity across entire languages (Golubović and Gooskens, 2015; Gooskens and Heeringa, 2004). At the same time, attempts to apply apply LD to individual words demonstrated a heavily diffuse pattern, with normalized LD values ranging between 0.0 and 0.5 (Saturno, 2023; Stenger et al., 2020).

Roembke et al. (2024) cite works where words are recognized as cognates at LD > 0.83 and as non-cognates at LD < 0.14. This approach, again, treats cognates as words being similar rather than having common historical roots. A notable omission of this approach is that it fails to account for the fact that inflectional factors (e.g., prefixation and suffixation) may dramatically impact graphic similarity between words – which is particularly relevant for morphologically rich languages, such as Germanic and Slavic languages.

In addition to LD, word adaptation surprisal (Stenger et al., 2020) and conditional entropy (Gooskens et al., 2007) were proposed to measure similarity of word pairs. These methods have the advantage of being asymmetrical, i.e. speakers of different languages are likely to exhibit varying intelligibility of words within the same pair of texts.

Thus, interpretation of cognates as similarly perceived words persists across a large number of studies, thereby violating the fundamental definition of cognates and leading to grave confusions. (Having said that, one should not deny the value of similarity-based methods for automated identification of candidates for true cognates.)

In order to divorce the two problems – finding common etymology and assessing MI in word pairs – the present paper introduces the notion of *percipiants* and their four classes as follows:

A) *True percipiants* are pairs of words which are perceived as being similar, but not necessarily sharing a common origin, e.g. povernulsja[1] :: povernuvsa (RU[2] повернулся :: UA повернувся turn-PST.M.SG-REFL '(he) turned (himself) around'[3]). Percipiants may be close to full, partial, or false cognates; however, we do not attempt to verify their common origin.

B) *Percipiant synonyms* are two (or more) words where a word has a synonym or a near-synonym within the same language; the (near-)synonym, in turn, is a true percipiant to the foreign counterpart of the pair, e.g. dorogu :: put :: putja (RU дорогу :: RS пут road-ACC.SG :: BG пътя road-DEF. The Russian synonym of dorogu is puti (пути road-

[1] Throughout this paper, both Latinate and Cyrilic scripts are transliterated using the normalized, pseudo-phonetic Latinization scheme (see Section 3) in order to ensure consistency of graphic representation.

[2] The following language codes are used throughout this paper: BG – Bulgarian; BY – Belarusian; CZ – Czech; PL – Polish; RS – Serbian; RU – Russian; SI – Slovene; SK – Slovak; UA – Ukrainian.

[3] In cases when a gloss and translation are valid for two or more languages (like here), the former are presented only once.

ACC.SG)).

C) Consequently, *non-percipiants* are pairs of words which are not perceived as being similar while having a common scope of meaning. This may occur due to the following:
   - they are true non-cognates, e.g. truboju :: surmoju :: polnitsou (RU трубою :: UA сурмою trumpet-INSTR.SG '(with) a trumpet' :: poľnicou SK (hunting).horn-INSTR.SG '(with) a horn').
   - they are possible cognates, yet their mutual recognizability is rather low e.g. ostanovilis :: zaustavili (RU остановились PFV-become.steady-CAUS-PST-PL-REFL 'they stopped' :: RS заустави PFV-PREV-set/place-CAUS-PST.PL '(they) stopped (something)').

D) Finally, *false percipiants* are pairs of words in which the foreign counterpart may be similar to the native word in its appearance, while having a different meaning, cf. spinoju :: grba (RU спиною back-INSTR.SG 'with (one's) back' :: BG гърба back-DEF 'the back'), where the Bulgarian word resembles the Russian gorba (горба hump-GEN.SG '(of one's) hump'). Similarly, povernulsja :: zavjarnuwsja (RU повернулся :: BY PFV-turn-SUF-PST.M.SG-REFL '(he) turned (himself) around'), where the Belarusian word resembles the Russian zavernulsja (завернулся PFV-wrap-SUF-PST.M.SG-REFL '(he) wrapped himself (up)'.

Bearing these four classes in mind, now let us examine the impact of (normalized) graphic proximity on mutual intelligibility of percipiants.

3. Data and methods

The data utilized in this experiment is the psycholinguistic dataset introduced in Klyshinsky and Lukashevich (2025). In the experimental group, the responses were collected from native speakers of Russian who were exposed to a Russian text concurrently with its translation counterpart in one of the eight Slavic languages: Belarusian, Bulgarian, Czech, Polish, Serbian, Slovak, Slovenian, and Ukrainian (for detail, see Klyshinsky 2021). Individual words were omitted in the Russian text; the respondents were tasked with restoring the missing words. In the control group, the same task was offered, yet with no parallel translation. In order to to standardize the representation of text for Russian-speaking respondents, the stimuli from non-Cyrillic graphic systems were presented both in the original (Latin) and the normalized Cyrillic forms, i.e. transliterations into the Cyrillic alphabet which were performed with IPA transcription (International Phonetic Association, 1999) as an intermediary.

In total, the dataset contains responses on 186 unique lemmas for 232 gaps across six texts. A total of 1,975 respondents in the experimental group and 283 respondents in the control group were surveyed. All word pairs (consisting of an omitted Russian word and its foreign counterpart) are manually tagged with one of the four classes of percipiants presented in Section 2; the number of cases per class is shown in Figure 4 (Section 4). Furthermore, the collected responses (both in the experimental and the control groups) were manually inspected and scored as follows:

- 1 point: correct answer, when a respondent chose the same word as in the original text, its synonym or near-synonym; divergencies from the original grammatical parameters were allowed.
- 0.5 points: partially correct answer, when a respondent chose a word derived from the original word yet of a different part of speech; or a word of the same part of speech which is unrelated to the original, yet without disrupting the narrative logic. The underlying assumption is that the user either understood the word while misunderstanding the syntactic structure; or they grasped the syntactic structure yet failed to find a correct lexical match.
- 0 points: wrong answer, when a respondent left the blank unfilled, or chose an entirely inadequate word – either from a wrong part of speech or disruptive to the flow of the narration.

For instance, in the phrase . . . rassuzhdal on po doroge. . . (RU . . . рассуждал он по дороге. . . '... he was reflecting as he was going down the road. . . '), the last word, road-DAT.SG is the target. Therefore, the responses (po) doroge and (po) puti (по дороге, по пути along=road-DAT.SG) are scored as correct; partially correct answers, for example, are (po) hodu (по ходу along=course-DAT.SG 'in the course (of something)'), (po) naitiju (по наитию by=intuition-DAT.SG 'intuitively'), (po) nocham (по ночам at=night-DAT.PL 'at nights'); finally, the following choices are marked as wrong: (po)-svoemu (по-своему ADVZ-one's.own-DAT 'in one's own manner'), (po-) obychnomu (по-обычному ADVZ-usual-DAT 'in the usual way'), (po-) rimski (по-римски ADVZ-Roman-NOM.SG.M 'in the Roman manner'), (po) sebe (по себе for=self-DAT 'to oneself'). It is reported that the annotation by percipient types and the scoring were performed by two native Russian linguists; the decisions of the first annotator were cross-checked by the second annotator; disagreements (~5% of the cases) were resolved by consensus.

Since the present study seeks to find a correlation between Levenshtein distance (LD) and the

correctness of respondents' answers, following Heeringa (2004) we resorted to phonetic representation of words rather than to their graphemic form. Therefore, we applied the same scheme of transcription as in the original dataset; the subsequent computations of LD on languages with Latin alphabets were performed on their normalized Cyrillic representations.The texts in languages with the Cyrillic script (other than Russian) were also normalized to ensure consistency, i.e. the characters that are not present in the Russian alphabet were substituted with their Russian equivalents (e.g., *i* → и, *j* → й).

Furthermore, in following Heeringa (2004), while computing LD we prioritized insertions and deletions over replacements, keeping equal weights across these three operations, which forces the algorithm to favor greater spans of similar characters. We used a weighted version of LD, where a transposition is assigned a total cost of 1 (i.e., 0.5 per character moved) rather than a cost of 2 (which would have been the cost for two separate substitutions). Substitutions between similar sounds (e.g., [o]-[u], [a]-[o], [ʃ]-[ʒ], [g]-[k]) were assigned a cost of 0.6.

Besides, we normalized the LD values by the length of the longest word within a pair; this occasionally resulted in the normalized LD > 1.0 for some words, that is, in the cases when transliteration increases a word's character count, for instance, when the Czech grapheme *ř* is transliterated as the respective longer Cyrillic digraph *rz*.

It is essential to note that, unlike previous studies on mutual intelligibility (Golubović and Gooskens, 2015; Gooskens and Heeringa, 2004), we analyze words as they appear in context; however, the present study does not aim to provide an in-depth analysis of the contextual factors that might impact intelligibility. This research question, therefore, remains to be explored in subsequent investigations; for present purposes, we will limit our discussion to brief observations and comments.

4. Results

As we showed above (Section 2), many studies use LD in order to measure the proximity of entire languages by averaging the values across massive collections of vocabulary. Such an approach has been shown to quite adequately capture cross-linguistic trends; thus, Gooskens & Heeringa (2004) demonstrate that mutual intelligibility (MI) is declining with the increase in LD. However, attempts to measure the MI for a pair of words (e.g. see Saturno 2023; Van Bezooijen and Gooskens 2005) yielded less impressive results. The present study also began with an attempt to straightforwardly apply normalized LD to evaluate MI of word pairs, with the MI value measured as the average score of responses from the experimental group across the prompts from all the eight foreign languages. However — quite expectedly — this attempt led to a not-so satisfactory outcome (see Figure 1).

In Figure 1, the majority of true percipiants have lower LDs as compared to the other percipient classes, with a fuzzy threshold between them running at around LD=0.5. However, some of the words marked by the experts as non-percipients occasionally have a low distance (LD < 0.2). It can also be seen from the figure that the overall trend of decreasing average scores along with increasing LDs manifests itself mainly at the lower boundary, i.e. some words will always be reliably restored irrespective of their LDs. A manual analysis of these instances indicated that, in the majority of cases, the missing word is recoverable solely from the monolingual context, irrespective of the similarity within the word pair. This happens, for example, when the missing word is a part of a frequent collocation or an idiom, e.g. in (1), where the underlined part of the sentence is a fixed collocation. Language-pairwise analysis of the data reveals that intelligibility may vary depending on the language of the non-Russian counterpart.

Interestingly, this behavior is occasionally observed even when LD is next to 0. Thus, the average score of kvadrata (see 2a) in the control group is 0.52; meanwhile, its average score in the presence of a parallel text ranges between 0.89 for UA, through 0.85 for BY and 0.76 for BG, to 0.71 for RS – although the LD to the target in all the cases is 0 (or just below): kvadrata :: kvadrata (RU, UA, BY квадрата square-GEN.SG; BG квадрата square-DEF.M.SG[GEN] 'of a square'). We presume that such variation is due to the high LD values of not the target per se, but of its surrounding context, cf. Examples 2a and 2b vs. 2a and 2c below.

Meanwhile, the responses given by the informants indicate that not all word pairs marked by the experts as true percipiants are perceived as such. Thus, the lowest average scores in the experimental group were obtained for the pairs raba :: *robski* (RU раба slave-GEN.SG :: BG робски slave-ADJ.SG 'of the slave') with the average score of 0.27 and LD=0.583; truboju :: trubkou (RU трубою :: CZ trubkou trumpet-INS.F.SG '(with) a trumpet') with the score of 0.51 and LD=0.286; and osuzhdennyh :: osudzhenike (RU осужденных condemn-PST.PTCP.PL-GEN.ACC.PL :: RS осуђенике condemn-PST.PTCP.PL-LOC.PL '(of/about) the convicted (people)') with the score of 0.3 and LD=0.5. For example, cf. (3), where the average score in the control group is 0.352 and in the experimental group it is 0.514. The poor recognizability of the target in the experimental condition is due to a number of factors, such as very low LDs

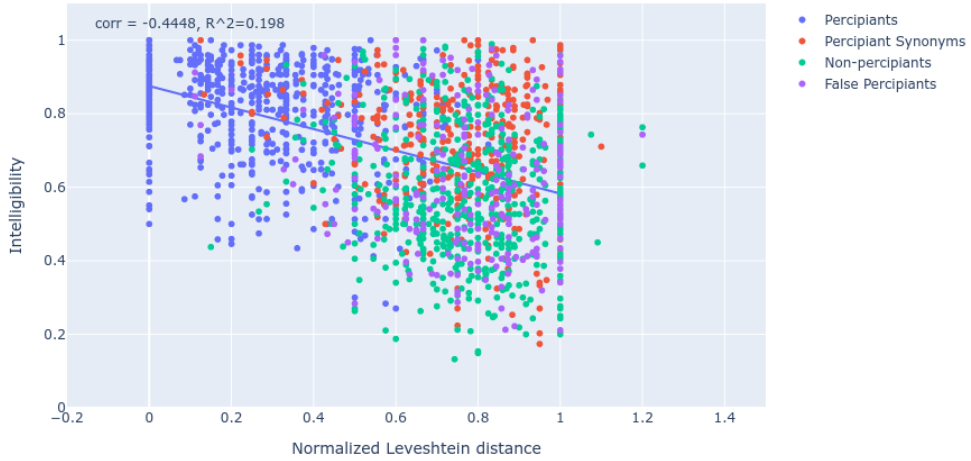

Figure 1: Distribution of mutual intelligibility among percipiants by type and LD.

(Ex. 1)

От	этой	мысли	у	Виниция	потемнело	в	глазах
Ot	etoj	mysli	u	Vinitsija	potemnelo	v	glazah
from	this.-F.GEN	thought.-GEN	at	Vinicius.-GEN	PFX-dark-PST-N.SG	at	eye-PREP.PL

'This thought made Vinicius's eyes grow dark.'

of the left and the right contexts; the profound syntactic differences which lead to differences in the linear word order (so that words from the original's left context occasionally move into the right context and vice versa) and to different syntactic means being used in a similar function (i.e. participial phrase vs. relative clause), etc. The opposite is also true, when pairs which was annotated as non-percipiants yielded high average scores, e.g. (4), where the non-percipiant pair poljubil vs. pokokhav (RU полюбил :: UA покохав PFV-love-PST.M.SG '(he) fell in love') has LD=0.688 and the average score of 0.95. Here, the high recognizability is promoted by the extra-sentential context and narrative motif which describes the protagonist's emotions and reflections caused by love. This context allows reliably restoring the missing word even when no parallel text is to help.

In contrast, consider (5) as an example of an 'ideal' percipiant: vecher :: vecher (RU вечер :: CZ večer evening-NOM.M.SG), with LD=0; the average score in the control group was 0.58, but in the experimental group it plummeted to 1.0. The reason is that the Russian sentence has a zero link verb, consequently, the gap allows for almost unrestrained combinability; in this case, the similarly-looking prompt from the parallel text provides critical assistance.

To sum up, the assessment based on word-stem similarity which is provided by experts, may not adequately align with the perception by respondents. That is, mutual intelligibility (MI) is affected by other, mainly contextual, factors except LD, as shown above.

Since the first stage of the analysis demonstrated that MI cannot be reliably quantified as a relationship between LD and the average score in the experimental group, we moved on to a more complex design. Here, MI of a word pair was computed as the difference between the average scores of the experimental and the control group (see Figure 2). The intuition behind this heuristic was that the predictability of the word in the native-language context is to increase when a parallel text in another mutually intelligible language is present. Note that MI may take both positive and negative values, i.e. presence of the parallel text may both facilitate and inhibit recognition of the target. It should be noted that if the average score for a word in the control group is close to 1, then the prompt from the parallel text will bring about little or zero increase; thus, the MI of the word pair will be close to 0.0. We realize that such cases do not adequately reflect MI and may introduce distortions into the overall result.

Therefore, they require a separate analysis and more scrupulous design of contexts in future research. Figure 2 demonstrates an increase in correlation and R^2, as compared to Figure 1 – from -0.44 to -0.53, and from 0.198 to 0.277, respectively. Most importantly, we see that true percipiants exhibit higher MI than the other classes of percipiants, alongside lower LD values. The mean MI for true percipiants is 0.208 (SD = 0.166), for percipiant synonyms it is 0.08 (SD = 0.143), for non-percipiants -0.04 (SD = 0.127), and for false percipiants -0.003 (SD = 0.138). Thus, we observe that true percipiants are, on average, significantly more intelligible than the other three classes.

Figure 3 presents the distributions of average

		1 – Avg. Score; 2 – LD, left context, 3 – LD, right context	1	2	3
(Ex. 2a)	RU	. . . на всех трех сторонах квадрата, кроме той, где были наши собеседники. . . . na vseh storonah kvadrata, krome toj, gde byli nashi sobesedniki.	0.52	NA	NA
(Ex. 2b)	UA	. . . на всіх трьох сторонах квадрата, крім тієї, де були наші бесідники. . . . na vsih troh storonah kvadrata, krim tiej, de buli nashi besedniki.	0.89	0.13	0.28
(Ex. 2c)	RS	. . . на све три стране квадрата, осим оне на којој су се налазили наши саговорници. . . . na sve tri strane kvadrata, osim one na kojoj su se nalazili nashi sagovornitsi.	0.71	0.5	0.76

'...on all three sides of the square, except for the one where our interlocutors were.'

(Ex. 3) RU . . . и мимо Пилата последним проскакал солдат с пылающей на солнце трубою за спиной.
. . . i mimo Pilata poslednim proskakal soldat s pylajusczcej na solntse truboju za spinoj.

CZ . . . poslední projel těsně kolem Piláta voják s trubkou na zádech, která plála v paprscích slunce.
. . . posledni projel tesne kolem Pilata vojak s trubkou na zadech, ktera plala v paprscich slunce.

'...and finally, past Pilate galloped a soldier, with a trumpet ablaze in the sunlight, slung across his back.'

(Ex. 4) RU Лишь теперь ему стало ясно, как сильно он ее полюбил.
Lish teper emu stalo jasno, kak silno on ee polubil.

UA Щойно лише йому стало ясно, як її покохав.
Szczojno lishe jomu stalo jasno, jak jj pokohav.

'Only now did he realize how much he loved her.'

(Ex. 5) RU . . . и увидел, что на Патриарших вечер.
. . . i uvidel, chto na Patriarshih vecher.

CZ . . . a viděl, že se mezitím na Patriarchovy rybníky snesl večer.
. . . a videl, zhe se mezitim na Patriarchovy rybniky snesl vecher.

'...and he saw that evening had fallen on the Patriarch's Ponds.'

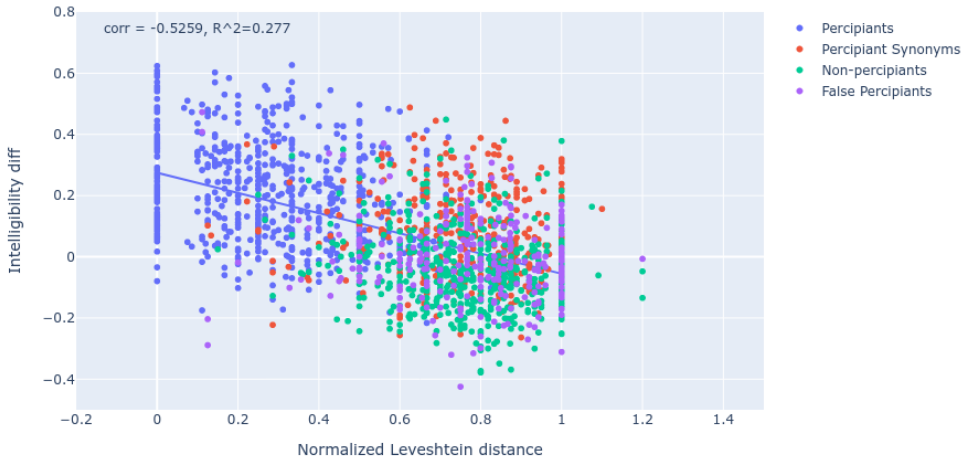

Figure 2: Distribution of the difference in intelligibility between the experimental and the control group, by the type and LD.

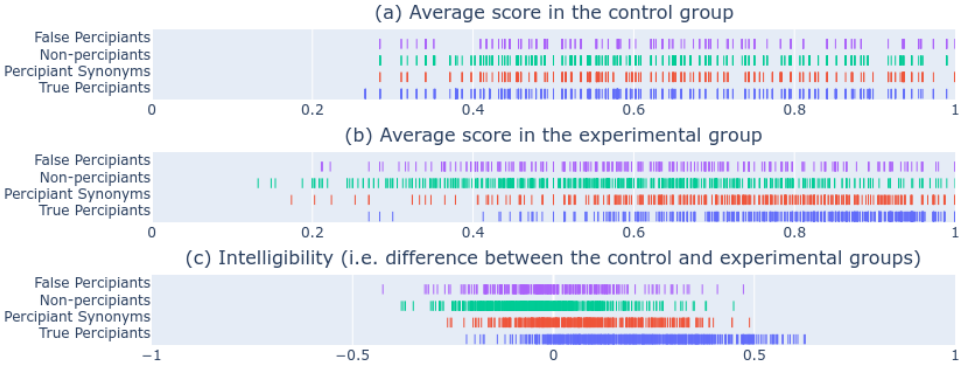

Figure 3: Distribution of the intelligibility of percipiants across the types.

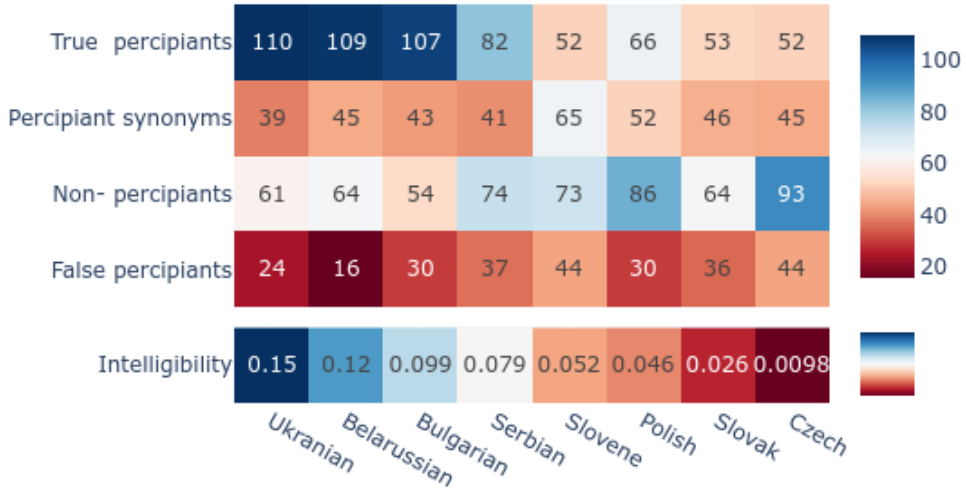

Figure 4: Distribution of the intelligibility of percipiants across the types.

scores (3a and 3b) and intelligibility (3c) across the percipiant classes. The distributions in the control group do not display statistically significant difference (p-values1 between 0.06 and 0.85), except true percipiants vs. percipiant synonyms (p-value<0.001). As for the distribution in the experimental group, it demonstrates a statistically significant difference (p-value<0.001) for each pair of classes. The MI gain relative to the control group is higher for true percipiants. Non-percipiants and false percipiants show the lowest increase, with their mean situated around zero. Percipiant synonyms demonstrate an intermediate behavior. Thus, we conclude that the introduction of the novel MI metric allows differentiating among the classes of percipiants.

Pearson's correlation between MI and the normalized number of true percipiants per language pair (see Figure 4) yielded a strong positive coefficient of 0.9257. We also built a linear regression model to predict MI from the distribution of the four classes, which produced the following coefficients: true percipiants 0.288; percipiant synonyms 0.078; non-percipiants -0.168; false percipiants -0.198; intercept 0.039, R^2 = 0.876. This suggest the following interpretation: true percipiants are the strongest predictors of MI; the predictive power of percipiant synonyms is about four times lower; non-percipiants and false percipiants are about twice as inhibitory as true percipiants are predictive.

5. Conclusion

We introduced the notion of percipiants which are defined as a pair of words which are mutually intelligible within a pair of languages. As distinguished from cognates, percipiants are intentionally agnostic about etymology and common origins. We also defined four classes of percipiants which are differ in the degree of their graphemic and semantic relatedness to their counterparts in genetically related languages. Two experiments were run. The first one analyzed the correctness of predicting missing words under the condition of a parallel text. Similarly to previous work, our analysis attested the general tendency to declining answer correctness with the increase in LD. The second experiment quantified MI as a difference between the average scores from the experimental and the control groups. Such metric enables accurately capturing the net gain in the intelligibility of text with vs. without support from the parallel text. Remarkably, loss in intelligi-

bility is also possible – therefore, parallel text may occasionally become an impediment. The implications of these observations are twofold. Firstly, they reveal a far more pronounced tendency for MI to decrease with higher LD. Secondly, they confirm the validity of dividing percipiants into four classes.

Note, however, that the proposed approach is not free from limitations. On the one hand, the task of restoring an omitted word does not allow isolating the influence exerted by contextual parameters from the perception of individual words. On the other hand, polysemous words may influence the choice of a percipiant pair if stripped of context. MI values with very low or negative gain over the control condition may be explained by genuinely weak MI within the pair, as well as by informational sufficiency of the monolingual context for word prediction. Further experimental and analytical work in this direction is required for a more in-depth understanding of MI and the role of context.

Research ethics: The Institutional Review Board of the National Research University Higher School of Economics at its meeting on August 21, 2023 approved this project as conforming entirely to ethical standards.

References

David Crystal. 2008. *A dictionary of linguistics and phonetics*. Blackwell Publishing.

Jelena Golubović and Charlotte Gooskens. 2015. Mutual intelligibility between West and South Slavic languages. *Russian Linguistics*, 39(3):351–373.

Charlotte Gooskens and Wilbert Heeringa. 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language variation and change*, 16(3):189–207. Publisher: Cambridge University Press.

Charlotte Gooskens, John Nerbonne, and Nathan Vaillette. 2007. Conditional entropy measures intelligibility among related languages. In *LOT Occasional Series*, volume 7, pages 51–66. Publisher: LOT, Netherlands Graduate School of Linguistics.

Wilbert Jan Heeringa. 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. Ph.D. thesis, University of Groningen, Groningen.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Eduard Klyshinsky. 2021. Statisticheskiĭ metod kolichestvennoĭ otsenki poniatnosti inostrannykh slavianskikh iazykov dlia russkoiazychnogo chitatelia [A statistical method for quantitative evaluation of the intelligibility of foreign Slavic languages for Russian-speaking readers]. *Vestnik Novosibirskogo gosudarstvennogo universiteta*, 19(1):61–79.

Vladimir Levenshtein. 1966. Binary codes capable or correcting deletions, insertions, and reversals. *Soviet Physics–Doklady*, 10(8):707–710.

Koji Miwa, Ton Dijkstra, Patrick Bolger, and R. Harald Baayen. 2014. Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: language and cognition*, 17(3):445–463. Publisher: Cambridge University Press.

Tanja C. Roembke, Iring Koch, and Andrea M. Philipp. 2024. What makes a cognate? Implications for research on bilingualism. *Bilingualism: Language and Cognition*, 27(5):1023–1028. Publisher: Cambridge University Press.

Jacopo Saturno. 2023. The impact of interlingual correspondences on cognate recognition in Slavic intercomprehension. *Russian Linguistics*, 47(2):209–229.

Lindsey R. Squires, Sara J. Ohlfest, Kristen E. Santoro, and Jennifer L. Roberts. 2020. Factors influencing cognate performance for young multilingual children's vocabulary: a research synthesis. *American Journal of Speech-Language Pathology*, 29(4):2170–2188.

Irina Stenger, Klára Jágrová, and Tania Avgustinova. 2020. The INCOMSLAV platform: experimental website with integrated methods for measuring linguistic distances and asymmetries in receptive multilingualism. In *Proceedings of the LREC 2020 Workshop on" Citizen Linguistics in Language Resource Development"*, pages 40–48, Marseille, France.

Renée Van Bezooijen and Charlotte Gooskens. 2005. How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? *Linguistics in the Netherlands*, 22:13–24.

Guy C. Van Orden. 1987. A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, 15(3):181–198.

6. Language Resource References

Klyshinsky and Lukashevich. 2025. *Intelligibility of Slavic Languages for Russian Native Speakers*. Science Data Bank, 1.0.