

Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2020”

Moscow, June 17–20, 2020

VISUAL VS. AUDITORY PERCEPTION OF BULGARIAN STIMULI BY RUSSIAN NATIVE SPEAKERS

Stenger I. (ira.stenger@mx.uni-saarland.de),
Avgustinova T. (avgustinova@coli.uni-saarland.de)
Saarland University, Saarbrücken, Germany

This study contributes to a better understanding of receptive multilingualism by determining similarities and differences in successful processing of written and spoken cognate words in an unknown but (closely) related language. We investigate two Slavic languages with regard to their mutual intelligibility. The current focus is on the recognition of isolated Bulgarian words by Russian native speakers in a cognate guessing task, considering both written and audio stimuli. The experimentally obtained intercomprehension scores show a generally high degree of intelligibility of Bulgarian cognates to Russian subjects, as well as processing difficulties in case of visual vs. auditory perception. In search of an explanation, we examine the linguistic factors that can contribute to various degrees of written and spoken word intelligibility. The intercomprehension scores obtained in the online word translation experiments are correlated with (i) the identical and mismatched correspondences on the orthographic and phonetic level, (ii) the word length of the stimuli, and (iii) the frequency of Russian cognates. Additionally we validate two measuring methods: the Levenshtein distance and the word adaptation surprisal as potential predictors of the word intelligibility in reading and oral intercomprehension.

Keywords: online experiments, receptive multilingualism, linguistic factors, Bulgarian, Russian

DOI: 10.28995/2075-7182-2020-19-684-695

ПОНИМАНИЕ БОЛГАРСКИХ СЛОВ-СТИМУЛОВ В ПИСЬМЕННОЙ И УСТНОЙ ФОРМАХ НОСИТЕЛЯМИ РУССКОГО ЯЗЫКА

Штенгер И. (ira.stenger@mx.uni-saarland.de),

Августинова Т. (avgustinova@coli.uni-saarland.de)

Университет земли Саар, Саарбрюккен, Германия

В настоящей статье представлен материал исследования в области рецептивного мультилингвизма, в центре которого находится выявление сходств и различий в процессе успешного понимания слов-когнатов незнакомого, но (близко)родственного языка в письменной и устной формах. Мы рассматриваем два славянских языка с точки зрения их взаимопонятности. В данной работе мы анализируем понятность изолированных болгарских слов-стимулов для носителей русского языка при выполнении экспериментов по свободному переводу слов-когнатов в письменной и устной формах. Результаты проведенных экспериментов свидетельствуют о достаточно высоком уровне понятности болгарских слов для носителей русского языка, а также указывают на трудности в процессе понимания в зависимости от формы восприятия. С целью объяснения полученных результатов мы анализируем, какие из лингвистических факторов играют более важную роль в понимании болгарских слов при чтении и на слух. Полученные результаты онлайн-экспериментов коррелируются со следующими переменными: а) идентичные и неидентичные соответствия на орфографическом и фонетическом уровнях, б) длина слова-стимула, в) частотность русских когнатов. Дополнительно производится оценка двух методов измерения: расстояние Левенштейна и мера неожиданности слова, в качестве потенциальных параметров, объясняющих понятность слов при чтении и на слух.

Ключевые слова: онлайн-эксперименты, рецептивный мультилингвизм, лингвистические факторы, болгарский язык, русский язык

1. Background

Globalization and migration processes as well as the rapid development of new technologies in recent years pose new challenges in communication. In this regard, a special mode of language use known as *receptive multilingualism* [Braunmüller, Zeevaert 2001] or *intercomprehension* [Doyé 2005] is becoming more and more topical. It shows the robustness of human language competence by taking on various types of information—e.g. contextual, situational, socio-cognitive—and employing multisource compensatory and guessing strategies. This unconventional communication form relies both on *inherent* intelligibility, i.e. when speakers of one language can

understand unknown but related languages because of structural linguistic similarities, and on *acquired* intelligibility, i.e. when they have learnt the respective language [Gooskens 2019]. An acquired *lingua receptiva* can apply to less related or unrelated languages, too [Muikku-Werner 2014], and this is also known as *mediated* receptive multilingualism [Branets et al. 2019] which eases understanding between typologically distant languages through the medium of a language closely related to the target.

In receptive multilingualism, the reading activity is expected to be essentially different from the listening one. While in listening the time available for input processing is limited, in reading one can jump back at will [Möller, Zeevaert 2015]. In foreign language teaching and learning, for example, the listening ability is generally considered more difficult than reading comprehension. This can certainly be true of receptive multilingualism, too. Examining the mutual intelligibility of some West and South Slavic languages, [Golubović 2016] reports slightly lower intelligibility scores in the spoken word translation task than in the written one, as well as generally lower values for the spoken cloze test in comparison to the written¹ one.

Our key assumption is that the recognition of individual cognates² is a precondition for any further text understanding, be it via reading or listening. In (closely) related languages, the recognition of isolated words is certainly not to be equated with the reading or listening of actual texts. In addition to the identification of orthographic or phonetic similarities, further assumptions about the text—based on external characteristics, textual context, or morphological and syntactic analysis—can play a significant role in text exploration [Möller, Zeevaert 2010]. An additional fact to consider here is that a given context is only useful for inferring purposes if it has been understood, which is not always the case in an intercomprehension scenario. While the possibility for intercomprehension is strongly linked to the amount of common vocabulary in the respective languages, the respective cognates are neither identical nor transparent enough [Möller, Zeevaert 2015]. Therefore, it is justified to scrutinize cognate recognition in isolation, i.e. without context, and based on these findings to proceed with the exploration of entire texts.

For our online experiments, we use the INCOMSLAV platform³, concentrating on the spoken and written translation of cognates. We want to see whether the degree of similarity between the source Bulgarian words and their cognate Russian targets could predict the expected mutual intelligibility, and how this might be different in visual and auditory perception. After presenting the experimental setup, we introduce linguistic factors that may influence and explain the cross-lingual intelligibility of visual and auditory linguistic input. We then present the experimental results in both modalities and analyze the influence of the linguistic factors on human performance, before we draw some general conclusions.

¹ Only participants who indicated that they could read Cyrillics were presented with written tasks in Bulgarian [Golubović 2016: 75].

² I.e. historically (etymologically) related word pairs that still bear the same meaning in both languages.

³ The website includes a large number of different online experiments in 11 Slavic languages (as well as in German and English) carried out as challenges in a linguistic game (<http://intercomprehension.coli.uni-saarland.de>).

2. Cognate guessing and free translation

After completing a background questionnaire, participants were asked to translate randomized 120⁴ written and spoken⁵ Bulgarian (BG) stimuli into their native Russian (RU) language in two series of 60 stimuli each. The items were taken from parallel lists consisting of internationalisms, Pan-Slavic vocabulary, and cognates from Swadesh lists (for more details cf. [Fischer et al. 2015], [Stenger et al. 2019]). As we are interested in *inherent* intercomprehension, only people who speak RU natively and who do not know the stimulus language BG have been included in the analysis. In order to avoid any learning effects we consider here the results of the initial experiment in each modality.

The number of participants in the *written* word translation task is 40, aged between 18 and 71 years (i.e. average age 33) with 32 women and 8 men⁶. They saw the stimuli on their screen one at a time, and had 10 seconds to translate each word. The number of participants in the *spoken* word translation task is 29, aged between 16 and 48 years (i.e. average age 32)⁷ with 23 women, 5 men, and 1 not specified. They listened to the stimuli one by one (each word was played twice), and had to provide a written translation within 10 seconds. The time limit was chosen based on the experience from other intercomprehension experiments, and the results were automatically categorized as ‘correct’ or ‘wrong’ via pattern matching with pre-defined correct answers and acceptable alternatives. The responses were then manually checked for typographical errors in the final analysis.

3. Explaining variables in linguistic stimuli

3.1. Orthographic and sound correspondences

We distinguish between *identical* and *mismatched* correspondences and assume that identical correspondences positively affect intelligibility while that mismatched correspondences affect it negatively [Stenger 2019]. We have calculated identical and mismatched correspondences by means of the Levenshtein algorithm for 120 BG–RU cognate pairs on the orthographic and phonetic level. For example, the BG–RU cognate pair *автомобил* [əftomoˈbɨl]—*автомобиль* [əftəməˈbʲilʲ] ‘car’ has 9 identical orthographic correspondences and 1 mismatched orthographic correspondence and 4 identical and 5 mismatched sound correspondences.

⁴ 118 nouns and 2 numerals.

⁵ BG stimuli were read aloud by a female native speaker and recorded in a professional sound studio.

⁶ For more details of the written word intelligibility test see [Stenger 2019].

⁷ In total 30 participants took part at the experiments, one male participant noticed the knowledge of the BG and was excluded from the analysis.

3.2. Word length

Word length has been shown to influence the intelligibility of individual words. For example, [Kürschner et al. 2008] found that longer words are more easily recognized than shorter words in an intercomprehension listening scenario. We calculated word lengths in terms of the number of characters and sound segments. The longest word in BG is *автомобил* [əftomob'itʃ] 'car' and consists of 9 characters and 9 sound segments. The shortest BG word is *ещ* [ɛʃ] 'hedhehog', consisting of 2 characters and 2 sound segments.

3.3. Word frequency

Word frequency may also influence the correct understanding of cognates, since speakers are exposed more often to frequent words [Kürschner et al. 2008]. Word frequencies of RU cognates we use are based on frequency lists from the Russian national corpus [Lyashevskaya, Sharov 2009]. With regard to the stimuli, the most frequent RU cognate is *один* 'one' (2245.7 ipm⁸) and the least frequent one is *кельнер* 'waiter' (0.4 ipm). So, the initial hypothesis is that the more frequent the RU cognate, the easier it would be to understand the equivalent BG stimulus.

4. Predictors of mutual intelligibility

4.1. Levenshtein distance

Linguistic similarity is commonly measured on cognates (cf. [Gooskens 2019], [Stenger 2019], [Vanhove 2014]) by computing the *Levenshtein distance* (LD), i.e. taking into account—for two corresponding items—the minimum number of symbols that need to be inserted, deleted or substituted in order to transform a word in one language into the corresponding word in another language. Employing a modified Levenshtein algorithm [Levenshtein 1965], which disallows matching between a vowel and a consonant; we have calculated the orthographic and the phonetic⁹ distances between 120 BG–RU cognate pairs. This objective measure, we calculated automatically using the *incompy* tool of [Mosbach et al. 2019]. While in the basic form of the algorithm all string operations have the same cost, we use 0 for the cost of mapping a character/sound to itself, e.g. *a:a* or [a]:[a], and 1 for the cost of aligning it to a character/sound of the same kind (vowel vs. consonant), e.g. *a:o* or [a]:[o].¹⁰ To compensate

⁸ instances per million words

⁹ For the transcription in IPA (International Phonetic Alphabet) we used the following online resources: <https://de.glosbe.com/>; https://en.wiktionary.org/wiki/Wiktionary:Main_Page; [Dellert, Jäger 2017] <http://www.northeastalex.org/> (18.11.2019). The missing IPA transcriptions for a few BG words were added according to Ternes, [Vladimirova-Buhtz 2010].

¹⁰ For orthographic distance we assign only the pair *e:ë* a substitution cost of 0.5. For phonetic distance we assign only the pair [l]:[ʎ] a substitution cost of 0.5. The semi-vowel [ɹ] may correspond to a vowel and to a consonant.

for word length effects we normalize dividing the total sum of costs by the number of slots in the alignment. For example, the normalized LD of BG–RU cognate pair in the written condition *язык–язык* ‘tongue/language’ is $2/4 = 0.5$ or 50%, while in the spoken condition $[ɛ'zɪk] - [ɪ'zɪk]$ ‘tongue/language’ it is $3/5 = 0.6$ or 60%. The assumption is that the higher the normalized LD, the more difficult it is to understand another language in an intercomprehension scenario.

4.2. Word adaptation surprisal

We take an information-theoretic approach [Shannon 1948] using the concept of *Word Adaptation surprisal* (WAS) to approximate the difficulty humans encounter when mapping one orthographic or phonetic system. Essentially, it reveals how predictable a given correspondence is in the respective pair of languages. Technically, WAS is calculated in bits according to the character or sound transformation probabilities, summing up values for character adaptation surprisal (CAS) or sound adaptation surprisal (SAS), respectively. For example, CAS is defined as in (1).

- (1) $CAS(L1 = c1|L2 = c2) = -\log_2 P(L1 = c1|L2 = c2)$
 L1—native language, c1—character of the native language
 L2—stimulus language, c2—character of the stimulus language

Since WAS between two words is computed by summing up the CAS and the SAS values of the contained characters and sounds in the aligned word pair, it strongly depends on the number of available word pairs (for more details see [Mosbach et al. 2019], [Stenger 2019]). Finally, we normalize the WAS based on the set of 120 BG–RU cognates. For example, the normalized WAS of the BG–RU cognate pair *язык–язык* ‘tongue/language’ is 0.799 bits in the written condition, while in the spoken condition $[ɛ'zɪk] - [ɪ'zɪk]$ ‘tongue/language’ it is 1.488 bits. The assumption is that the higher the normalized WAS value, the more difficult it is to comprehend the unknown language on the word level.

5. Results of intercomprehension experiments

The mean percentage of correctly translated items constitutes the intercomprehension score in a given modality (Table 1). The results show no great difference between the visual and the auditory perception of BG words by RU native speakers: the RU participants understand a slightly larger number of written BG words (71.33%) than spoken ones (68.42%).

Table 1. The results of written and spoken word translation tasks

	Bulgarian word translation task	
Native language	written	spoken
Russian	71.33%	68.42%

The next question is whether RU subjects are more successful in cognate guessing when they perceive the BG stimuli as a visual or as an auditory input. **Fig. 1** presents quantitative data on the 62 successfully guessed spoken cognates (left side), the 56 successfully guessed written cognates (right side) and the two cognate pairs yielding identical scores (middle). The vertical axis expresses the difference as the percentage of correct translations.

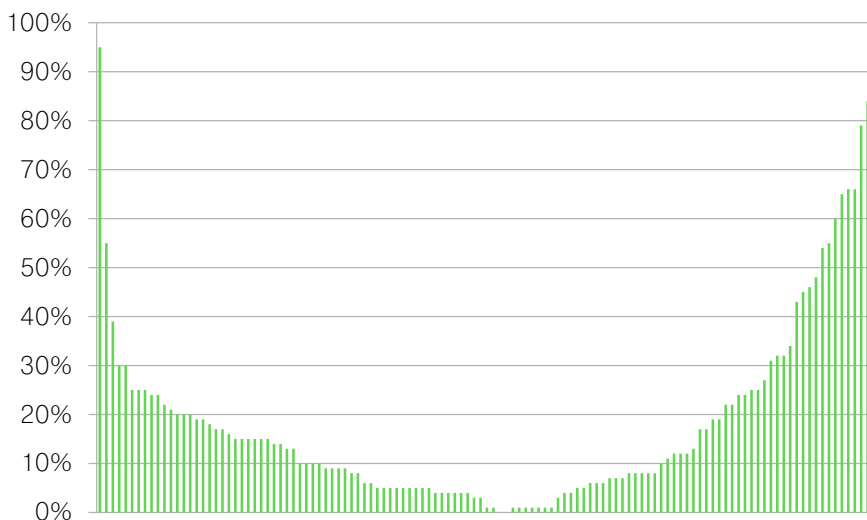


Fig. 1. Number of successfully guessed cognates while listening (left) vs. reading (right)

In **Table 2** we split the differences between the intercomprehension scores into three categories: high (more than 60%), medium (between 60% and 30%), and low (less than 30%). For 6 BG words in the high-percentage group, reading gave much better results than listening, e.g., the BG stimulus *еж* [ɛʃ] was correctly understood as *ёж* [jəʒ] ‘hedgehog’ to 90% in the written translation task and only to 6% in the spoken translation task. For 10 BG words at the middle group, reading caused fewer difficulties than listening, e.g., the BG stimulus *лѐн* [lɛn] was correctly translated in the written task as *лѐн* [lʲɛn] ‘flax’ to 60% and in the spoken task only to 5%. These examples show that for RU subjects the BG–RU mismatched orthographic correspondences *e:ë* is more transparent than the respective mismatched sound correspondences. There were a total of 40 BG stimuli in the low-percentage group, where RU subjects performed better in the written than in the spoken task. Only 1 BG stimulus *ръка* [rʲɔ'ka] was much better understood by RU subjects in auditory perception as *рука* [ru'ka] ‘hand, arm’ (100% intelligibly vs. 5%).

Table 2. Grouping cognate pairs according to differences in correct translations

% difference	visual perception	auditory perception
60–100	N = 6 BG <i>еж</i> [ɛʃ] RU <i>ёж</i> [jɵʂ] ‘hedgehog’	N = 1 BG <i>ръка</i> [rʁ'ka] RU <i>рука</i> [rʊ'ka] ‘hand, arm’
30–60	N = 10 BG <i>леп</i> [lɛn] RU <i>лён</i> [l'ɵn] ‘flax’	N = 4 BG <i>кръв</i> [krʁf] RU <i>кровь</i> [krɔfʲ] ‘blood’
1–30	N = 40 BG <i>дъб</i> [dʁp] RU <i>дуб</i> [dup] ‘oak’	N = 57 BG <i>език</i> [ɛ'zik] RU <i>язык</i> [ɹ'zik] ‘tongue’

It is remarkable that the most frequent incorrect answer in the written word translation task was the RU word *река* ‘river’ which can be considered an orthographic neighbor to the BG stimulus and to the RU cognate word¹¹. Furthermore, 4 spoken BG stimuli were in the middle group and 57 spoken BG words fell into the low-percentage group of differences with regard to intelligibility. To conclude, the slightly better performance in the written task as opposed to the spoken one (71.33% vs. 68.42%) is due to the higher number of correctly translated words in the first two groups.

6. Correlation of intercomprehension scores with explaining variables

We correlated the experimentally obtained intercomprehension scores with the explaining variables introduced in **Section 3**—the identical and mismatched correspondences (ic vs. mc), the word length (wl) and the word frequency (wf)—in order to determine their impact on word-level written and spoken intelligibility. An overview of the statistical results (Pearson’s *r* and *p*-value) is provided in Table 3.

Table 3. The correlations between intercomprehension scores and explaining variables

Translation task	Explaining variables			
	ic	mc	wl	wf
visual perception	$r = 0.3480$ $p < 0.0001$	$r = -0.3830$ $p = 1.61e-05$	$r = 0.3210$ $p < 0.0005$	$r = -0.0160$ $p = 0.09$
auditory perception	$r = 0.2359$ $p < 0.01$	$r = 0.0550$ $p = 0.55$	$r = 0.3620$ $p < 0.00005$	$r = 0.1610$ $p = 0.08$

¹¹ Previous research shows that the neighborhood density may play a significant role in spoken and written word recognition without context [Kürschner et al. 2008], [Stenger 2019]. For example, Bulgarian and Serbian written intelligibility to Russian native speakers shows that the higher the neighborhood density, the lower is the number of successful translations, although this is not the case for Ukrainian, Belarusian, and Macedonian stimuli when presented to Russian readers [Stenger 2019].

In visual perception, intercomprehension scores correlate significantly with the identical and mismatched orthographic correspondences, as well as with the word length. The correlation between the written intelligibility and word frequency is not positive as assumed, but insignificant.

In auditory perception, only two variables have an impact on word recognition: identical sound correspondences and the word length. The correlation between the intercomprehension scores and the mismatched sound correspondences is not negative as assumed, but yet insignificant. It might be the case that qualitative characteristics of the mismatched sound correspondences are more decisive in spoken word recognition than their number. We also found a positive but low and not significant correlation between the intercomprehension scores and the word frequency in the spoken modality. This was already shown by [Kürschner et al. 2008] and could be replicated in our experiment.

7. Correlation of intercomprehension scores with predictors of mutual intelligibility

7.1. Levenshtein distance

To investigate how word intelligibility depends on orthographic and phonetic similarities, the intercomprehension scores were correlated with the respective normalized LDs (Fig. 2).

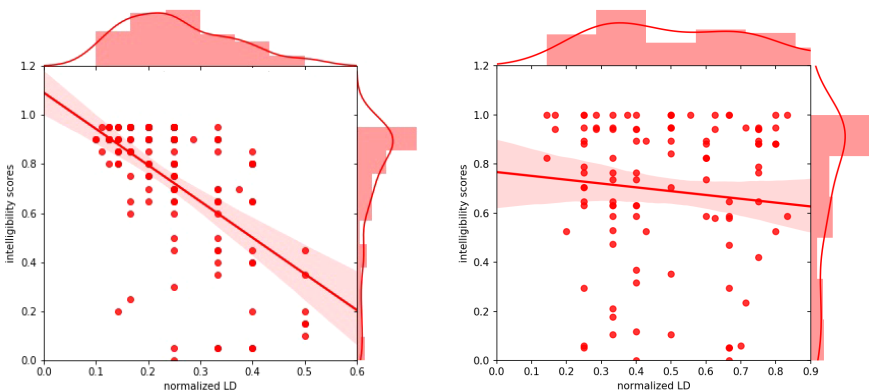


Fig. 2. Normalized LD as predictor of written (left) and spoken (right) intelligibility

There is a negative and significant correlation of -0.566 ($p = 1.47e-11$) in the written modality (Fig. 2 left), which means that on the word level the written intelligibility can be predicted well from the normalized LD. This is in contrast to the spoken modality (Fig. 2 right), where we can see that the spoken intelligibility is not well predictable from the normalized LD (Pearson's $r = -0.1185$, $p = 0.197$). This can

be explained by the fact that the normalized LD depends on the number of mismatched sound correspondences as well as on their weights and thus increases with more mismatched correspondences. The RU listeners seem to have a high tolerance towards the phonetic distance and are able to process sensitive sound differences quite well. To model the perception of spoken BG stimuli by RU listeners more successfully, we decided to integrate into the LD algorithm sensitive sound distances with a substitution cost of 0.5 between non-palatalized and palatalized consonants, as well as between the following similar correspondences: [ɛ]:[e], [i]:[ɪ], [ɔ]:[o], [u]:[ʊ], [ʃ]:[ʂ], [ʒ]:[ʒ̣], [ʦ]:[ʦ̣]. As a result, we found a significant negative correlation between the intercomprehension scores and the phonetic distance: (Pearson's $r = -0.2019$, $p < 0.05$).

7.2. Word adaptation surprisal

The correlations between the normalized WAS and the intercomprehension scores of both translation tasks (Fig. 3) show that written and spoken intelligibility cannot be predicted reliably from the normalized WAS on the word level. Both correlations are very low and not significant (written: Pearson's $r = -0.1350$, $p = 0.14$, spoken: Pearson's $r = -0.1492$, $p = 0.1$).

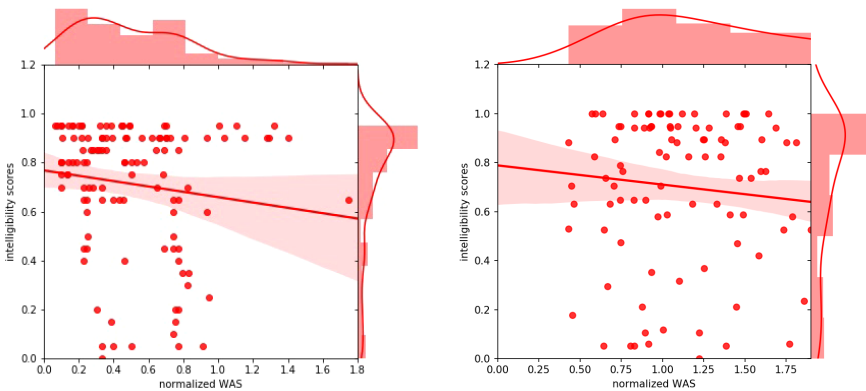


Fig. 3. Normalized WAS as predictor of written (left) and spoken (right) intelligibility

The low and insignificant correlations can be explained by the fact that identical orthographic and sound correspondences may still have a small adaptation surprisal value, which increases the whole WAS value. Hence, we decided to modify our normalized WAS calculation in such a way that all identical orthographic and sound correspondences were measured with 0 bits. The calculated CAS values for mismatched orthographic correspondences and the SAS values for mismatched sound correspondences remained unchanged in the modified calculation. After the modification of the WAS method, we found a negative and significant correlation between the modified normalized WAS and the intercomprehension scores in both conditions (written: Pearson's $r = -0.210$, $p < 0.05$, spoken: Pearson's $r = -0.181$, $p < 0.05$).

8. Conclusions

In a series of online experiments, we investigated how well RU natives could understand written and spoken BG words in an intercomprehension scenario. Overall, RU subjects are good at recognizing both written and spoken BG stimuli: 71.33% and 68.42% respectively. The obtained intercomprehension scores in the written task are only slightly higher than in the spoken one. We considered which linguistic factors can determine the mutual intelligibility and to what degree. First, we examined the individual variables that are known to affect intelligibility and found that identical correspondences and word length of the BG stimuli have a positive and significant impact on human performance in both perception modes. However, while the statistical analysis reveals that mismatched orthographic correspondences could be significant in negatively affecting written intelligibility, the number of mismatched sound correspondences cannot be an explanatory variable for the difficulties in spoken word recognition. The frequency of cognates is not a reliable predictor in listening as well as in reading and seems not to have any effect on word recognition. Second, we validated two predictors of written and spoken intelligibility: the normalized LD and normalized WAS. While the written intelligibility was well predictable from the orthographic distance, the explanatory potential of the phonetic distance was very small and not significant. The incorporation of sensitive sound distances into the LD algorithm showed a significant influence of the phonetic distance in spoken word comprehension, but still to a low degree. After the modification of the WAS method, we found that predictability of both orthographic and sound correspondences have slight effects on subject's ability to understand an unknown but related language. The knowledge gained from this study will serve as a useful point for further research in understanding of the specific linguistic factors that contribute to (un)successful intercomprehension.

9. Acknowledgements

We wish to thank Marius Mosbach for his support in the implementation of our data in the *incompy* tool. This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 232722074—SFB 1102.

References

1. Branets A., Bahtina D., Verschik A. (2019), Mediated receptive multilingualism: Estonian- Russian-Ukrainian case study, in *Linguistic Approaches to Bilingualism*, available at: <https://doi.org/10.1075/lab.17079.ver> (18.11.2019).
2. Braunmüller K., Zeevaert L. (2001), Semicommunication, receptive multilingualism and related phenomena. A bibliographical overview, [Semikommunikation, rezeptive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandsaufnahme], Working papers in multilingualism [Arbeiten zur Mehrsprachigkeit], Series B, No. 19, University Hamburg [Universität Hamburg].
3. Dellert J., Jäger G. NorthEuraLex (version 0.9). available at: <http://www.northeuralex.org/> (18.11.2019).

4. *Doyé P.* (2005), *Intercomprehension. Guide for the development of language education policies in Europe: from linguistic diversity to plurilingual education*, Reference study, Strasbourg, DG IV, Council of Europe.
5. *Fischer A., Jágrová K., Stenger I., Avgustinova T., Klakow D., Marti R.* (2015), *An Orthography Transformation Experiment with Czech-Polish and Bulgarian-Russian Parallel Word Sets*, in Sharp B., Lubaszewski W., Delmonte R. (eds.), *Natural Language Processing and Cognitive Science 2015 Proceedings*, Libreria Editrice Cafoscarina, Venezia, pp. 115–126.
6. *Golubović J.* (2016), *Mutual intelligibility in the Slavic language area*, PhD thesis, University of Groningen (Netherlands).
7. *Gooskens C.* (2019), *Receptive multilingualism*, in Montanari S., Quay S. (eds.), *Multidisciplinary perspectives on multilingualism: The fundamentals*, De Gruyter Mouton, Berlin, pp. 149–174, http://www.let.rug.nl/gooskens/pdf/publ_multidisciplinary_perspectives_2019.pdf.
8. *Kürschner S., van Bezooijen R., Gooskens C.* (2008), *Linguistic determinants of the intelligibility of Swedish words among Danes*, *International Journal of Humanities and Arts Computing* 2(1/2), pp. 83–100.
9. *Levenshtein V. I.* (1965), *Binary codes capable of correcting deletions, insertions, and reversals* [Dvoičnye kody s ispravleniem udalenij, vstavok i zamen simbolov], *Doklady of the Soviet Academy* [Doklady Akademii Nauk SSSR], 1965, Vol. 163, No. 4, pp. 845–848.
10. *Lyashevskaya O. N., Sharov S. A.* (2009), *Frequency dictionary of the contemporary Russian language* [Chastotnyy slovar' sovremennogo russkogo yazyka], Azbukovnik, Moskva.
11. *Muikku-Werner P.* (2014), *Co-text and receptive multilingualism Finnish students comprehending Estonian*, *Journal of Estonian and Finno-Ugric Linguistics* 5(3), pp. 99–103.
12. *Möller R., Zeevaert L.* (2010), „There I spontaneously think of a blackboard“—For word recognition in related Germanic languages [„Da denke ich spontan an Tafel“—Zur Worterkennung in verwandten germanischen Sprachen], *Journal for foreign language research* 21(2) [Zeitschrift für Fremdsprachenforschung 21(2)], pp. 217–248.
13. *Möller R., Zeevaert L.* (2015), *Investigating word recognition in intercomprehension: Methods and findings*, *Linguistics* 2015 53(2), pp. 313–352.
14. *Mosbach M., Stenger I., Avgustinova T., Klakow D.* (2019), *incom.py—A Toolbox for Calculating Linguistic Distances and Asymmetries between Related Languages*, in Angelova G., Mitkov R., Nikolova I., Temnikova I. (eds.), *Proceedings of Recent Advances in Natural Languages Processing, RANLP 2019*, pp. 811–819.
15. *Shannon C. E.* (1948), *A mathematical theory of communication*, *Bell System Technical Journal* 27 (379–423), pp. 623–656.
16. *Stenger I.* (2019), *On the role of orthography in Slavic intercomprehension with a special focus on the Cyrillic script* [Zur Rolle der Orthographie in der slavischen Interkomprehension mit besonderem Fokus auf die kyrillische Schrift], PhD thesis, Universaar, Saarbrücken.

17. *Stenger I., Avgustinova T., Belousov K., Baranov D., Erofeeva E.* (2019), Interaction of linguistic and socio-cognitive factors in receptive multilingualism, in Selegej V. et al. (eds.), *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference 'Dialogue' (2019)*, Issue 18, Supplementary volume, pp. 211–226.
18. *Ternes E., Vladimirova-Buhtz T.* (2010), Bulgarian, in *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, Cambridge University Press, Cambridge, pp. 55–57.
19. *Vanhove J.* (2014), Receptive multilingualism across the lifespan. Cognitive and linguistic factors in cognate guessing, PhD thesis, University of Fribourg (Switzerland).