



THE ROLE OF ORTHOGRAPHIC AND PHONETIC DISTANCES IN MUTUAL INTELLIGIBILITY BETWEEN MONTENEGRIN AND BULGARIAN

Abstract: *It is generally believed that speakers of closely related languages are able to understand each other to some degree without resorting to a lingua franca as a communicative mode. The extent to which languages are mutually intelligible depends on a variety of linguistic and extralinguistic factors, including the genetic proximity of the language varieties concerned. This study looks at mutual intelligibility between two closely related Slavic languages, namely Montenegrin and Bulgarian, and examines the role of orthographic and phonetic distances in the process of word recognition. In a small-scale word translation experiment, native speakers of Montenegrin and Bulgarian were tested on their ability to decode the meaning of 98 written and spoken nouns in the related language without prior instruction. The results reveal a strong correlation between orthographic/phonetic distances and correctly recognized words, which suggests that linguistic distances may act as relatively significant predictors of intelligibility between the two languages in question.*

Key Words: *intelligibility, closely related language, orthographic distance, phonetic distance, neighbourhood density*

Introduction

Languages sharing a close genetic relationship are considered to be mutually intelligible to a lesser or greater extent. Yet, the degree of mutual intelligibility of closely related language varieties is contingent on a number of linguistic factors, including the lexicon, phonetics/phonology and morphosyntax, as well as on extralinguistic determinants such as language attitudes, amount of contact and orthography (Gooskens, 2018). As previous research on mutual intelligibility of closely related languages has shown, the closer languages are in terms of linguistic distances, the greater the chance that speakers of those languages will be able to understand each other and engage in a successful interaction. Such a form of multilingual constellation, in which interlocutors speak their own native L1s when communicating with each other, is known as receptive multilingualism (ten Thije & Zeevaert, 2007). This mode of intercultural communication

¹ Doctoral student at the Faculty of Philology, University of Montenegro.

is well established in the Scandinavian language area (see van Bezooijen and Gooskens, 2007; Schüppert, 2011), but it is less clear whether it is also possible to interact in this fashion in other European regions, including in Southeast Europe.

This study aims to look at the mutual intelligibility of two closely related Slavic languages – Montenegrin and Bulgarian, and explore the relative contribution of orthographic and phonetic distances to how well their native speakers can potentially understand each other. By testing the intelligibility of isolated words at the written and spoken levels, the paper will attempt to provide empirical evidence as to whether intelligibility of these two languages can be predicted by linguistic distances between words of the same historical origin. Results of the written intelligibility test will be compared to those obtained in the spoken intelligibility task, which will then be correlated with orthographic and phonetic distances.

2. Background

Mutual intelligibility of closely related languages is a relatively recent field of linguistic study, which can be traced back to Haugen's (1966) seminal work on what he labelled semi-communication. In this type of multilingual communication, speakers interact with each other using their native languages, though mutual understanding is often incomplete and fraught with problems. Semi-communication has recently come to be referred to as receptive multilingualism (ten Thije & Zeevaert, 2007), by which both hearer and speaker perspectives are given importance. To what extent closely related languages are mutually intelligible and whether such a form of interaction is possible in practice has been investigated by numerous studies, some of which will be referenced in this paper.

From the very beginning, researchers working on intelligibility between Indo-European languages largely focused on the Scandinavian language area (e.g. Zeevaert, 2004; Gooskens, 2006, 2007; van Bezooijen & Gooskens, 2007; Beijering et al., 2008, Schüppert, 2011, Schüppert et al., 2015), as speakers of Danish, Norwegian and Swedish have communicated with each other using their own L1s for centuries. The level of intelligibility of other closely related Germanic languages such as Dutch, German and Afrikaans has also attracted some scholarly attention (e.g. van Bezooijen & Gooskens, 2007). More recently, intelligibility studies have been carried out in the Slavic, Germanic and Romance language areas as part of the Micrela project at the University of Groningen (for an overview of results, see Gooskens & van Heuven, 2017).

On the whole, various linguistic and extra-linguistic factors have been shown to affect mutual intelligibility, including linguistic distances, which is of particular relevance for this paper. Examining the role of linguistic distances in the mutual intelligibility among spoken Danish, Swedish and Norwegian, Gooskens (2007) found that phonetic distances play a crucial role in the comprehension between speakers of these three languages, while lexical distances, which were measured as the percentage of non-cognates, do not appear to play a significant role in the intelligibility between these three languages. Along those lines, Beijering et al. (2008) reported a strong correlation between phonetic distances and intelligibility by testing the intelligibility of 17 Scandinavian language varieties for Danish speakers, while lexical distances were found to have less effect on mutual understanding. In a research project that is of particular relevance for the present study, Kürschner et al. (2008) tested the intelligibility of 384 isolated Swedish words and correlated the scores with different linguistic factors. Their findings point to a strong negative correlation between phonetic distances and intelligibility, while a number of other linguistic factors such as word length, neighbourhood density, orthography and word frequency also had some bearing on intelligibility scores.

As regards intelligibility within the Slavic language area, some research has recently been carried out in this regard. Golubović (2016) examined cross-language intelligibility between Slavic languages spoken in the European Union, namely Bulgarian, Croatian, Czech, Polish, Slovak and Slovene, resulting in 30 language combinations. This study was carried out within a large-scale web-based project named Micrela², which explored the degree of mutual intelligibility of 16 closely related languages within the Germanic, Slavic and Romance language groups in Europe. The intelligibility between these languages was measured using a word translation task, a cloze test and a picture-to-text matching task. She reported a high level of mutual intelligibility between Czech and Slovak (92.7% and 95.0% respectively), which is in line with some previous studies (Nábělková, 2007), as well as for Croatian and Slovene (43.7% and 79.4%). Of particular interest to the present study, Golubović (2016) found a relatively high degree of intelligibility of Croatian and Bulgarian, which will be discussed further below in more detail.

To the best of my knowledge, mutual intelligibility between Montenegrin and Bulgarian has not been tested so far. These two languages belong to the South Slavic language group and form part of the Slavic dialect continuum, a language area stretching from the Black Sea

² <http://www.micrela.nl/>

to the Eastern Alps (Ivić, 2001). Montenegrin is one of four languages that emerged from the former Serbo-Croatian, which used to be the official language in the former Yugoslavia until its dissolution in the early 1990s (Greenberg, 2004). Following the break-up, all countries that were once part of Yugoslavia named their official languages after their respective nations (Bosnian, Croatian and Serbian). Montenegro did the same in 2007 after it had voted independence, declaring Montenegrin as its official language (Glušica, 2011). All the above-mentioned languages share the same linguistic system and are mutually intelligible to a very high degree, which is why many linguists (e.g. Kordić, 2010) consider them as the standard varieties of the pluricentric Serbo-Croatian language.³ In line with the view maintained by Trudgill (1992), languages are as much political and cultural constructs as they are linguistic concepts, which is why Montenegrin will be regarded here as a separate language despite the fact that it shares the same structural properties with other descendants of Serbo-Croatian.

3.1. Research questions

As stated above, the present paper will look at the role of orthographic and phonetic distances in mutual intelligibility between spoken and written Montenegrin and Bulgarian. Specifically, the following research questions will be addressed:

1. To what extent are Montenegrin and Bulgarian mutually intelligible at the written and spoken word level?

³ Despite being mutually intelligible to a very high degree, the successor languages to Serbo-Croatian are now codified separately in each of the four independent countries (Bosnia and Herzegovina, Croatia, Montenegro and Serbia). The recognition of their mutual intelligibility was further reinforced by the Declaration on the Common Language (2017), stating that people in the four ex-Yugoslav countries speak a common pluricentric language and underlying that the use of different names for the four standard varieties – Bosnian, Croatian, Montenegrin and Serbian – does not imply that those are different languages. The Declaration was initiated by a group of prominent linguists, intellectuals and civil society activists from Bosnia and Herzegovina, Croatia, Montenegro and Serbia and has since received significant attention from both scholars and the general public, but also from politicians. While acknowledging the right of speakers to refer to the language as they wish, the Declaration maintains that the four languages together form a pluricentric language in the same way as English, German, Arabic, French, Spanish, Portuguese and many others. For more discussion on the sociolinguistic situation in the former Yugoslavia, see Greenberg, 2004 and Bugarski, 2018).

2. What is the relative contribution of orthographic and phonetic distances to mutual intelligibility between Montenegrin and Bulgarian?

3.2. Method and procedure

Mutual intelligibility between closely related languages can be tested at the word, sentence or text level and using a variety of methods, such as comprehension questions, translation tasks and reaction times, to name a few (for an overview of different methods and approaches, see Gooskens, 2013). This paper will employ a word translation task to test intelligibility between Montenegrin and Bulgarian, relying on the implicit assumption that word recognition plays a key role in text and speech comprehension. Given that the aim of the study is to investigate the contribution of linguistic distances to mutual intelligibility between these two languages, the word translation task is expected to reduce the effect of other factors on intelligibility scores to a minimum. As Gooskens (2013) suggests, factors such as context or word position in a sentence may bear upon intelligibility and consequently obscure the contribution of linguistic distances to word recognition.

The present experiment replicates the word translation task that was previously used by Golubović (2016), who tested how well Croatian speakers understand randomly chosen 25 out of the 100 most frequent Bulgarian nouns and vice versa. In the written word translation task, Golubović found an intelligibility level of 64.07% for Croatian speakers decoding Bulgarian written words, while Bulgarian speakers correctly identified 64.55% of Croatian words, which indicates that the intelligibility of written language forms is symmetric. Similar results were obtained for in the spoken word translation task, with Croatian speakers performing slightly better when confronted with Bulgarian words (67.35%) than vice versa (64.22%).

The present experiment tests the intelligibility of 98 most frequent singular nouns from the British National Corpus, which were translated into the two languages for the purposes of testing. This approach was adopted so as to ensure that no language was given a special status in terms of word choice and word frequency. The first 50 stimuli were used to test mutual intelligibility at the written word level, whereas the remaining 48 nouns⁴ from the above-mentioned frequency list were employed with a view to assessing the intelligibility of spoken

⁴ The initial plan was to test the intelligibility of 50 words in the spoken test as well. However, two words (one in each language group) failed to be retrieved from the compressed folder, which is why those pairs were excluded from the experiment given that additional recording could not be arranged in a timely manner.

word forms. Two different sets of words were used for practical reasons, as some of the participants took part in both written and spoken part. This approach was chosen in order to prevent a priming effect in the subjects who took part in the written test, as otherwise they would have been presented with the same words in the spoken intelligibility test.

The written task involved a total of 20 participants (10 in each language group), while 16 subjects took part in the spoken comprehension test (8 in each group). The majority of participants were recruited online via social media or by email, while several subjects were tested on the spot. All the participants were aged between 20 and 44 and either studied for or had a university degree in various subject areas. In addition, participation in the experiment was on a voluntary basis, so no financial compensation was provided for taking part in the test.

In both tasks, the participants were asked to provide a translation equivalent in L1s based on the words they read or heard. As regards the written experiment, the target words were presented in the Cyrillic script in both languages in order to control for the effect of orthography on word recognition by Bulgarian subjects, considering that Cyrillic is the only official script in Bulgaria. Montenegrin, on the other hand, is a biscriptal language, whereby Latin and Cyrillic enjoy an equal status, which is why the choice of a script is completely arbitrary. The subjects were asked to translate the given words without the help of the Internet, dictionaries or other tools.

The stimuli that were recorded for the spoken word recognition task were read out by a Montenegrin and a Bulgarian speaker respectively. The Montenegrin words were read out by a radio presenter from Montenegro, whereas the Bulgarian audio files were made by a standard speaker of Bulgarian within the above-mentioned Micrela project. All the recorded stimuli were set at approximately 70 db, so that none of the groups would get an advantage by way of volume intensity. Since the subjects lived at different locations, the monitoring of the spoken word recognition task could not be arranged for practical and financial reasons. Instead, the files were emailed to the subjects in a compressed folder along with an answer sheet and detailed instructions about the task. The participants were allowed to play each sound file twice and were required to provide their answer within ten seconds. Upon completion of the test, the subjects sent their answer sheets back and provided basic information about education, age, language background and optional feedback about the test.

While selecting participants, the most important criterion alongside level of education and age was that none of the participants had been exposed to the test language to a considerable extent. An additional requirement on the Montenegrin subjects was that they had

received no formal instruction in Russian, given that Bulgarian and Russian share several cognates that appear in the word list, which could have given an initial advantage to Montenegrin volunteers speaking Russian. Thus, for example, the Bulgarian word *prichina*, meaning 'reason', is identical in spelling with the Russian word *prichina*, whereas the Montenegrin word for 'reason' is *razlog*. Another motivation for adopting such a criterion in selecting participants was the fact that Bulgarian and Russian share several orthographic symbols which do not exist in Montenegrin, such as й, щ, ю, я and ъ. This requirement is another variation on the word translation task employed by Golubović (2016), who tested the intelligibility of Bulgarian written words only if a participant could read Cyrillic, which is not used in any of the official Slavic languages in the EU save for Bulgarian. As she admits, this might have introduced bias to her results, as those participants who could read Cyrillic may have been able to do so because they had learned a Cyrillic-based Slavic language such as Russian, for which it is necessary to be familiar with the Cyrillic script.

3.3. Intelligibility scores and linguistic distances

As pointed out in the research questions, the main goal of the study was to look at mutual intelligibility of Montenegrin and Bulgarian and examine the relationship between orthographic and phonetic distances and intelligibility scores. In both parts of the experiment, intelligibility was expressed as the percentage of correctly translated words. Answers were checked manually in consultation with a Montenegrin and Bulgarian native speaker respectively and each correct response was given one point. In case of polysemous words, such as the Bulgarian word *m"zh* (which could be translated as either husband or man), all possible translations were accepted.

Orthographic and phonetic distances were calculated by means of the Levenshtein algorithm, which measures the minimum cost of operations needed to change one word into another through insertion, deletion and substitution of characters or phonemes (Heeringa, 2004). Both types of linguistic distances were expressed as percentages on a scale from 0 to 100, whereby a zero was assigned to two Montenegrin and Bulgarian words that are identical in writing or pronunciation. For instance, the distance between the Bulgarian word *r"ka* (arm) and Montenegrin *ruka* amounts to 25%, as the two 4-letter words differ in one character – the symbol representing the mid back unrounded vowel in Bulgarian ([ɤ]), transliterated as ["]) is substituted by the letter [u] in Montenegrin. The cost of operations (i.e. 1) was divided by the number of alignments (4), so the distance equalled to $1/4 = 0.25$ or 25 percent (see Figure 1). Distances were only calculated for cognates, as words

which do not share the same historical origin are as a rule unintelligible without previous instruction. For this reason, the correlation analysis only looked at the relationship between intelligibility and linguistic distances among cognate words.

	1	2	3	4
Montenegrin	r	U	K	a
Bulgarian	r	"	K	a
	0	1	0	0

1/4 = 25% difference

Figure 1 – Illustration of the Levenshtein algorithm

Furthermore, linguistic distances were measured in two directions – from Montenegrin to Bulgarian and vice versa, particularly because some of the stimuli might have a close synonym for the corresponding word in the participants’ native language, by which linguistic distance is significantly reduced. Thus, for instance, the Bulgarian word *rabotja* (work) corresponds to the Montenegrin non-cognate *posao*. However, the Montenegrin participants could easily deduce the meaning of this word because of the existence of the cognate synonym *rabota* in Montenegrin. Consequently, the orthographic distance for this particular pair was calculated from Bulgarian to Montenegrin only, as Bulgarian participants were only presented with the non-cognate word (*posao*). Likewise, the Montenegrin word *pogled* (view) was correctly decoded by all Bulgarian subjects, given that the same word exists in the Bulgarian language. In this case, the distance from Montenegrin to Bulgarian equalled 0. By contrast, the Bulgarian test word *gledka* is quite distant from the Montenegrin *pogled*, which is why the distance from Bulgarian to Montenegrin with respect to this particular word was 66, as there is no synonym in Montenegrin resembling the Bulgarian counterpart.

4.1. Results and discussion: written intelligibility test

At face value, the degree of mutual intelligibility between Montenegrin and Bulgarian was found to be relatively high for both written and spoken language. With regard to written word comprehension, the mean percentage of correctly translated words among the Montenegrin subjects was 74.2% (SD = 4.56), which is slightly better than the average score achieved by the Bulgarian participants (M = 70%, SD = 5.49). The difference between the two groups was only significant at $t(18) = 1.86, p = 0.08$ (two-tailed), which might suggest that

a larger testing population is needed in order to ascertain whether this minor asymmetry is indeed present when it comes to written word recognition. These intelligibility scores are significantly higher than the ones reported by Golubović (2016) for the Croatian-Bulgarian pair (see above).

Having taken a closer look at the answers, it was evident that both groups had difficulties with translating non-cognates, which confirms the hypothesis that historically unrelated words are generally unintelligible without prior contact or instruction. Even when non-cognates were translated correctly, those were isolated cases and the correct answers might have come as a result of the subjects' previous exposure to the test language or recognition of a loanword from a language they are familiar with (see Gooskens, 2018). Moreover, false friends in both languages proved to be particularly misleading, as such words usually yielded wrong translations. For instance, the Bulgarian word *chast* (part) was incorrectly translated by the Montenegrin participants as *čast*, meaning 'honour', instead of *dio*, which is not cognate with the Bulgarian word for 'part'. Furthermore, problems were also caused by words that do share the same origin, but whose meanings shifted throughout history so that now they refer to different concepts. To illustrate, the Montenegrin word *riječ* (word) was wrongly translated by the Bulgarian participants as *rech*, meaning 'speech' in modern Bulgarian, while the correct term for this word would actually be the non-cognate *duma*.

4.1.1. Orthographic distances and intelligibility

As pointed out above, one of the research questions in the present study was to determine whether intelligibility can be predicted by orthographic distances. Since it is only reasonable to measure orthographic distances between cognates, words which do not share the same etymology and false friends were excluded from this part of the study. In particular, the correlation analysis looked at the relationship between orthographic distances for each word and the percentage of correctly translated words by each group. Because neither of the datasets demonstrated a normal distribution (Shapiro-Wilk test, $p < 0.001$), association between the two variables was assessed using Spearman's rho – the non-parametric alternative to Pearson's correlation coefficient.

The statistical analysis of the intelligibility between Bulgarian to Montenegrin speakers revealed a significant negative relationship between orthographic distances and correctly translated nouns ($\rho = -.62$, $p < 0.001$, two-tailed). In other words, the closer words are in terms of orthography, the higher intelligibility.

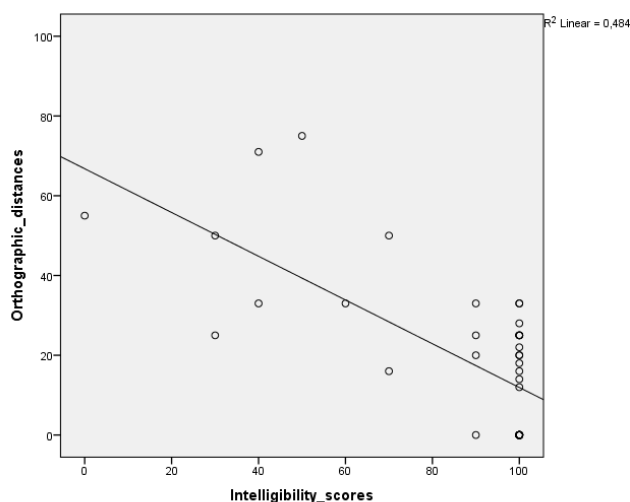


Figure 2 – Scatterplot showing the (negative) correlation between orthographic distances and the percentage of correctly translated words by the Montenegrin participants

Upon a closer look at the examples, it was noticed that Bulgarian words containing orthographic symbols which do not exist in Montenegrin caused some problems for the Montenegrin participants, especially when found in shorter words. By way of illustration, the Bulgarian stimuli word *нощ* (transliteration *nosh*) meaning ‘night’ was frequently translated as *нож* (knife) instead of *ноћ*, which could be attributed to the unfamiliarity of Montenegrin speakers with the letter *щ*, as well as to the high orthographic neighbourhood density of the target word. A word’s (orthographic) neighbourhood density refers to the number of words that can be generated by substituting a single letter in a target word (see Coltheart et al, 1977). Shorter words are generally known to have more neighbours (i.e. competitors) than longer words and thus are more difficult to recognize. In the case cited above, the target Montenegrin word *ноћ* has at least 4 competitors which differ from it by just one letter (*нож* (knife), *нос* (nose), *ној* (ostrich), *моћ* (power)), which made it likely to be confused with some of those words. The association between orthographic distances and the intelligibility of Montenegrin words for Bulgarian speakers was even higher ($\rho = -.75$, $p < .001$, two-tailed). This indicates that Bulgarian speakers also rely on orthography to a great extent when decoding the meaning of Montenegrin words.

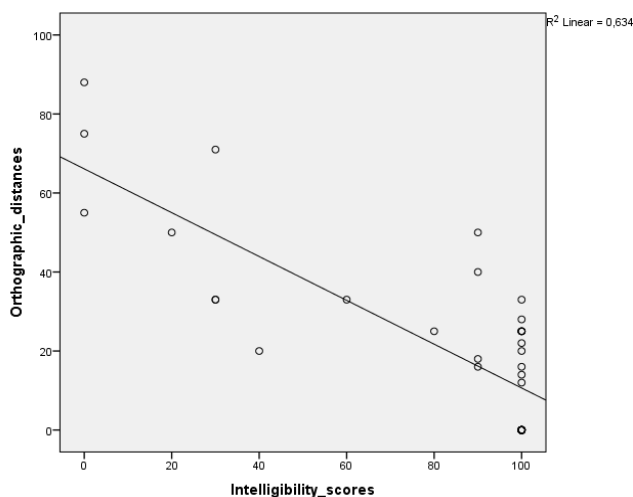


Figure 3 – Scatterplot showing the (negative) correlation between orthographic distances and the percentage of correctly translated Montenegrin words by the Bulgarian participants

Just like in the Montenegrin group, the Bulgarian participants experienced difficulties when trying to recognize shorter words because of the effect of their orthographic neighbourhood density. Thus, for instance, the word *dijete* meaning child was often confused with the word *dieta* (diet), while the appropriate Bulgarian term is *dete*. Also, similarly to the problem experienced by Montenegrin participants, the word *noć* (night) was frequently mistranslated as *nozh* (knife), which can also be accounted for by the high neighbourhood density of the target word *nosht* and the absence of the letter *ć* from the Bulgarian spelling system.

4.2. Results and discussion: spoken intelligibility test

Mutual intelligibility between spoken Montenegrin and Bulgarian was tested based on a list of 48 singular nouns. As in the written task, word recognition at the spoken level was found to be relatively high in both groups. The Montenegrin volunteers on average translated 76 percent of the Bulgarian test nouns correctly (SD=2.6), whereas the Bulgarian participants got as much as 80 percent of the answers right (SD=3.2). However, unlike the written intelligibility test, the difference between the two groups was significant ($t(14) = 3.09$, $p < 0.05$ (two-tailed)). Yet, more speakers would need to be tested in order to obtain a clearer picture about a potential discrepancy between the two groups, especially when with respect to written and spoken word intelligibility. Again, these spoken intelligibility scores are significantly

higher than those reported by Golubović (2016) for Croatian and Bulgarian (see above).

Similarly to the written test, non-cognate words were mostly unintelligible to the participants in both groups. In contrast to the written part, the spoken test contained few false friends (if any), so this phenomenon did not cause particular difficulties. The only exception might be the Bulgarian word *pazar*, which stands for 'market' or 'bazaar' and originally entered many of the Balkan languages via Turkish from Persian (*bāzār*). Due to its polysemous nature, this noun frequently came to be understood as takings, turnover or trade (*pazar*) and not as market, bazaar (*pijaca*), although in the past the word *pazar* was used in the sense of 'market' or 'square' in Montenegrin as well. However, the latter meaning has now become almost completely lost and thus the present-day speaker is unlikely to associate *pazar* with such archaic usage. Keeping this in mind, the word pair *pazar-pijaca* was excluded from the correlation analysis.

One of the reasons why Bulgarians scored slightly better than Montenegrins on the spoken intelligibility test might be the fact that recognition of some Montenegrin words was facilitated by the existence of phonetically very close words in Bulgarian, while their Bulgarian counterparts used in the test either had obsolete cognates in Montenegrin or were phonetically more distant from Montenegrin words. To illustrate this, the Montenegrin word *djevojka* (girl) was correctly decoded by all the Bulgarian participants, given that an almost phonetically identical cognate synonym exists in Bulgarian (*devoika*), aside from the more frequent term *momiche*. On the other hand, none of the Montenegrin subjects were able to recognize the Bulgarian stimuli *momiche*, as its cognate counterpart *moma* is only found in traditional Montenegrin and Serbian lyric poetry and is almost never used in contemporary Montenegrin.⁵ Interestingly enough, the Montenegrin participants unsuccessfully combined information from their L1 lexicon with the stimulus word *momiche*, which lead them to confuse the item with the word 'momče' (lad) due to negative transfer. By the same token, in the example cited earlier, all the Bulgarian participants correctly recognized the word *pogled* (view) since the same noun is used in Bulgarian, whereas the Montenegrin participants faced difficulties when presented with the phonetically distant word *gledka* (view), which no one got right.

⁵ Since the word *momiche* only had an obsolete cognate synonym in Montenegrin, this noun was treated as a non-cognate and hence was excluded from the correlation analysis.

Another problem that might have impeded recognition of some of the Bulgarian words on the part of Montenegrin speakers was potentially created by phonemes that do not exist in Montenegrin, such as the mid-back unrounded vowel, as well as by the prosodic properties of the stimuli words. This was especially the case with shorter stimuli and those with a high phonetic neighbourhood density (Luce & Pisoni, 1998). For instance, the Bulgarian word *s"vet*, meaning 'advice' or 'council', was often mistranslated as *svijet* (meaning 'world'), which could be put down to the speakers failure to identify the first vowel (/ʌ/ or the IPA symbol /ɤ/). Another reason for the confusion might be attributed to the fact that the participants wrongly associated the stimuli word with the ekavian⁶ version of the Montenegrin word *svijet* – i.e. *svet*, which is phonetically similar to the Bulgarian *s"vet*. Additionally, the position of the stress in this word is also likely to have contributed to negative transfer. By contrast, the word *savjet* ('advice' or 'council') was mostly translated correctly by the Bulgarian participants, as the open central unrounded vowel /a:/ appeared to have facilitated recognition.

4.2.1. Phonetic distances

As mentioned above, phonetic distances were measured using the Levenshtein algorithm. All the words were transcribed using SAMPA (Wells, 1997) in consultation with native speakers of Montenegrin and Bulgarian respectively. Apart from the insertion, deletion and substitution of phonemes, which were all given 1 point, stress position was also taken into account while calculating distances and was assigned half a point, as this prosodic feature is also considered to have some bearing on the intelligibility of words.

Just like in the written intelligibility test, the Spearman rho correlation coefficient was used in order to examine the relationship between mutual intelligibility and phonetic distances between Montenegrin and Bulgarian words in the light of non-parametric datasets. As regards the recognition of Bulgarian words by Montenegrin participants, there is a rather strong negative correlation between these two variables ($\rho = -.59$, $p < .001$). In simple terms, the less phonetic distance, the higher intelligibility.

⁶ Ekavian dialects are mainly spoken to the east of the Serbo-Croatian language area, predominantly in Serbia.

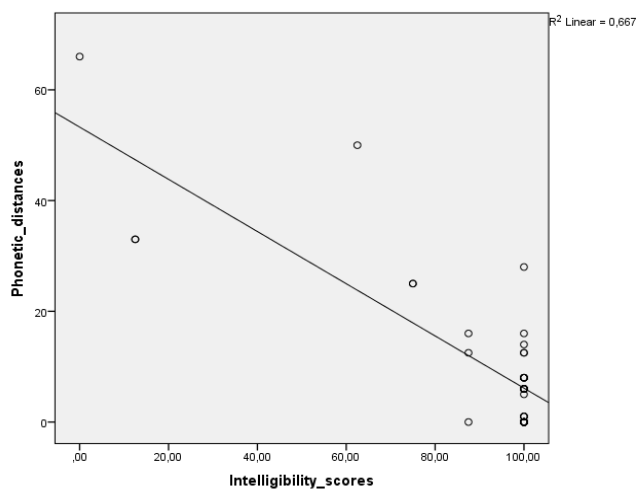


Figure 4 – Scatterplot showing the (negative) correlation between phonetic distances and the percentage of correctly translated Bulgarian words by the Montenegrin participants

The statistical analysis confirms that phonetic distances can predict to a great extent whether a Bulgarian word will be intelligible to Montenegrin speakers. Yet, in some cases, intelligibility could not be accounted for by the mere distance, particularly when it comes to shorter words that have a high phonetic neighbourhood density. For example, the Bulgarian stimuli word *syd* (court) was in most instances translated as *sat* (meaning ‘clock’, ‘hour’ or ‘watch’) instead of *sud*, even though the phonetic distance between the pair is 33 out of 100.

The relationship between phonetic distances and intelligibility of Montenegrin words for Bulgarian speakers seems to be rather moderate, but the association is still significant ($\rho = -.45$, $p < .01$).

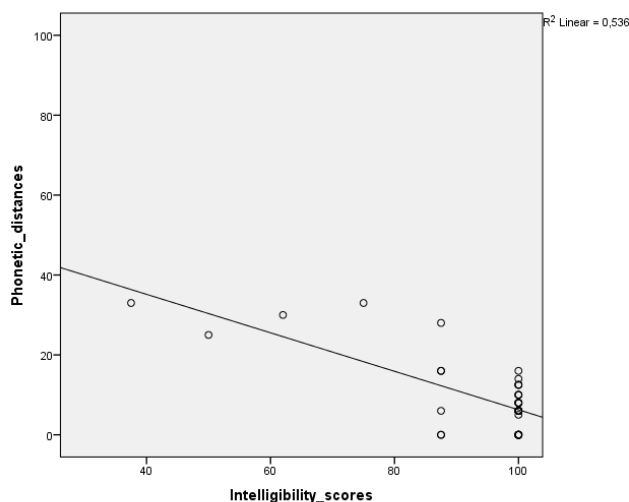


Figure 5 –Scatterplot showing the (negative) correlation between phonetic distances and the percentage of correctly translated Montenegrin words by THE Bulgarian participants

This somewhat weaker correlation might be due to the fact that some of the more distant Montenegrin words did not pose as much difficulty to the Bulgarian participants as distant Bulgarian counterparts did to the Montenegrin subjects, one of the factors being word length. Such is the case with the Montenegrin word *savjet*, which was translated correctly by 75 percent of the Bulgarian participants, in spite of the distance of 33. This speaks in favour of the hypothesis that longer words are better recognized than shorter ones, as the number of neighbours competing with the target word decreases with word length, which enhances intelligibility.

5. Conclusion

This paper reported on the results of a small-scale experiment on mutual intelligibility between Montenegrin and Bulgarian and examined the contribution of linguistic distances to the intelligibility between the two languages. Intelligibility was tested at the word level and empirical evidence was obtained for both the written and spoken language. With respect to the first research question, it can be concluded that mutual intelligibility between the two South Slavic languages in question is quite high in both written and spoken form. The level of mutual intelligibility was found to be higher than that of Croatian and Bulgarian reported by Golubović (2016), which might be explained by the fact that Montenegrin and Bulgarian are slightly closer to one another on the South Slavic dialect continuum than it is the case with

Croatian and Bulgarian. An interesting asymmetry arose when written and spoken tests were compared, as the Montenegrin group performed more successfully on the written part, whereas Bulgarians scored better on the spoken task. This discrepancy could be partly explained by the availability of synonyms in the participants' L1s that could be retrieved in trying to decode word meanings. Still, a larger sample size would be necessary to corroborate these findings.

As regards orthographic and phonetic distances, these were found to be moderate to strong predictors of mutual intelligibility between Montenegrin and Bulgarian. The association between linguistic distances and word recognition appears to be clearer at the written level, though this could have been affected by the different sets of words used in the two experiments. In terms of orthography, the unfamiliarity of speakers with some spelling symbols may have hindered word intelligibility, particularly in respect of shorter words and those having a high neighbourhood density. By the same token, the recognition of spoken words could also have been influenced by word length and a high neighbourhood density, but also by vowel quality, prosodic properties and unfamiliarity of speakers with certain phonemes, especially when confronted with shorter words.

To conclude, Montenegrin and Bulgarian appear to be intelligible to a high degree at the word level in both written and spoken form and the relative contribution of orthographic and phonetic distances seems to be relatively important. Yet, despite its limitations in terms of population size and methodology, this research has given rise to a number of questions in need of further investigation, including potential intelligibility asymmetries and the role of cognate synonyms in mutual understanding. Also, future research on this topic should aim to look at how other factors such as lexical distances might predict intelligibility between these two languages, preferably at the sentence or text level.

References

- Beijering, Karin, Charlotte Gooskens, and Wilbert Heeringa. "Predicting intelligibility and perceived linguistic distances by means of the Levenshtein algorithm." *Linguistics in the Netherlands* (2008): 13-24.
- Bugarski, Ranko. *Govorite li zajednički?*. Beograd: Biblioteka XX vek, 2018.
- Glušica, Rajka. "About the Re-standardization of the Montenegrin Language." *Studi Slavistici* 8.1 (2011): 271-278.
- Golubović, Jelena. *Mutual intelligibility in the Slavic language area*. Groningen: University of Groningen dissertation, 2016.
- Gooskens, Charlotte. "Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility." *Linguistics in the Netherlands*. Ed. Jeroen

- van de Weijer, and Bettelou Los. Amsterdam: John Benjamins, 2006. 101-113.
- Gooskens, Charlotte. "The contribution of linguistic factors to the intelligibility of closely related languages." *Journal of Multilingual and Multicultural Development* 28. 6 (2007): 445-467.
- Gooskens, Charlotte. "Methods for measuring intelligibility of closely related language varieties." *The Oxford Handbook of Sociolinguistics*. Eds. Robert Bayley, Richard Cameron, and Ceil Lucas. Oxford: Oxford University Press, 2013. 195-213.
- Gooskens, Charlotte. "Dialect intelligibility." *The Handbook of Dialectology*. Eds. Charles Boberg, John Nerbonne, and Dominic Watt. Hoboken, NJ: John Wiley & Sons, Inc, 2018. 204-218.
- Gooskens, Charlotte, and Vincent van Heuven. "Measuring cross-linguistic intelligibility in the Germanic, Romance and Slavic language groups." *Speech Communication* 89 (2017): 25-36.
- Greenberg, Robert. *Language and Identity in the Balkans*. Oxford: Oxford University Press, 2004.
- Haugen, Einar. "Semicommunication: The language gap in Scandinavia." *Sociological Inquiry* 36 (1966): 280-297.
- Heeringa, Wilbert. *Measuring dialect pronunciation differences using Levenshtein distances*. Groningen: University of Groningen dissertation, 2004.
- Ivić, Pavle. *Dijalektologija srpskohrvatskog jezika: Uvod u štokavsko narečje*. Sremski Karlovci/Novi Sad: Izdavačka knjižarnica Zorana Stojanovića, 2001.
- Coltheart, Max, Eileen Davelaar, Jon Torfi Jonasson, and Derek Besner. "Access to the internal lexicon." *Attention and Performance VI: Proceedings of the Sixth International Symposium on Attention and Performance*. Ed. Stanislav Dornic. Hillsdale, NJ: Erlbaum, 1977. 535-555.
- Kordić, Snježana. *Jezik i nacionalizam*. Zagreb: Durieux, 2010.
- Kürschner, Sebastian, Charlotte Gooskens, and Renée van Bezooijen. "Linguistic determinants of the intelligibility of Swedish words among Danes." *International Journal of Humanities and Arts Computing* 2.1-2 (2008): 83-100.
- Luce, Paul A., and Pisoni, David B. "Recognizing spoken words: The Neighborhood Activation Model." *Ear and Hearing* 19 (1998): 1-36.
- Nábělková, Mira. "Closely-related languages in contact: Czech, Slovak, 'Czechoslovak'." *International Journal of the Sociology of Language* 183 (2013): 53-73.
- Schüppert, Anja. *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*. Groningen: University of Groningen dissertation, 2011.
- Schüppert, Anja, Nanna Haug Hilton and Charlotte Gooskens. "Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and spoken word recognition." *Linguistics*, 53.2 (2015): 375-403.
- ten Thije, Jan D., and Ludger Zeevaert, eds. *Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts*. Amsterdam: Benjamins, 2007.

- Trudgill, Peter. "Ausbau sociolinguistics and the perception of language status in contemporary Europe." *International Journal of Applied Linguistics* 2 (1992): 167-177.
- van Bezooijen, Renée, and Charlotte Gooskens. "Interlingual text comprehension: linguistic and extralinguistic determinants." *Receptive multilingualism and intercultural communication*. Ed. ten Thije, Jan D., and Ludger Zeevaert. Amsterdam: Hamburger Studies in Multilingualism, 2007. 249-264.
- Wells, John C. "SAMPA computer readable phonetic alphabet." *Handbook of Standards and Resources for Spoken Language Systems*. Eds. Dafydd Gibbon, Roger Moore, and Richard Winski. Berlin and New York: Mouton de Gruyter, 1997.
- Zeevaert, Ludger. "Interskandinavische Kommunikation: Strategien zur Etablierung von Verständigung zwischen Skandinaviern im Diskurs." *Philologia: Sprachwissenschaftliche Forschungsergebnisse* 64. Hamburg: Kovač, 2004.