`incom.py` 2.0 − Calculating Linguistic Distances and Asymmetries in Auditory Perception of Closely Related Languages

Marius Mosbach[1,2], Irina Stenger[1], Tania Avgustinova[1],
Bernd Möbius[1], Dietrich Klakow[1,2]

[1] Department of Language Science and Technology, Saarland University, Germany
[2] Saarland Informatics Campus

{mmosbach, dlakow}@lsv.uni-saarland.de

ira.stenger@mx.uni-saarland.de

{avgustinova, moebius}@lst.uni-saarland.de

Abstract

We present an extended version of a tool developed for calculating linguistic distances and asymmetries in auditory perception of closely related languages. Along with evaluating the metrics available in the initial version of the tool, we introduce word adaptation entropy as an additional metric of linguistic asymmetry. Potential predictors of speech intelligibility are validated with human performance in spoken cognate recognition experiments for Bulgarian and Russian. Special attention is paid to the possibly different contributions of vowels and consonants in oral intercomprehension. Using `incom.py` 2.0 it is possible to calculate, visualize, and validate three measurement methods of linguistic distances and asymmetries as well as carrying out regression analyses in speech intelligibility between related languages.

1 Introduction

1.1 Background

Individual (receptive) multilingualism is essential in the European cultural and economic area.[1] How can native speakers of L1 spontaneously understand Lx on the basis of L1-Lx relatedness? Common experience shows that the degree of success in intercomprehension differs between spoken and written modalities, due to various linguistic and non-linguistic factors (Gooskens, 2019). While in listening the time available for the auditory input processing is limited, in reading one can jump back at will (Möller and Zeevaert, 2015) during visual input processing. The latter scenario, however, excludes the possibility to check comprehension through interactive communicative feedback. Even though cross-lingual differences are a multidimensional phenomenon (van Heuven, 2008), a simple baseline prediction of speech intelligibility between related languages can be based on the similarity of their phoneme inventories and lexicons.

For example, both similarity of phoneme inventories and lexical similarity were identified as factors predicting the perceptual confusability (and, by implication, similarity) of languages in the Great Language Game (Skirgård et al., 2017). These authors report confusion asymmetries within the set of Slavic languages, i.e., cases of speakers of one language (e.g. Bulgarian) understanding another language (e.g. Russian) better than the other way round. Phonetic and lexical distances are also considered as determinants of mutual intelligibility in the work of Gooskens and colleagues, who use a game-like interface, MICReLa[2], to collect data for Germanic, Romance, and (six) Slavic languages (Gooskens and van Heuven, 2020).

Notions of entropy and surprisal (Shannon, 1948) are employed in the INCOMSLAV framework[3] to measure information density and gauge how language users master high degrees of surprisal due to partial incomprehensibility of linguistic encodings. The key idea is that understanding an unknown but related Lx should be better when the L1 language model adapted for processing the unknown Lx exhibits relatively low average surprisal, or information density (Fischer et al., 2017; Jágrová et al., 2018; Stenger et al., 2017).

---

[1](https://ec.europa.eu/education/policies/multilingualism/about-multilingualism-policy_en)

[2]http://www.micrela.nl/app/
[3]https://intercomprehension.coli.uni-saarland.de

1.2 This Paper

In the present study we extend the `incom.py` toolbox[4] (Mosbach et al., 2019) focusing on mutual intelligibility aspects in oral intercomprehension. First, we compare the available measuring methods for linguistic distances and asymmetries – i.e., Levenshtein distance and word adaptation surprisal – as predictors of mutual intelligibility in auditory perception and add word adaptation entropy as an additional metric for asymmetric intelligibility. At the same time, we consider phonetic aspects, in particular vowel and consonant (dis)similarities, as explaining variables in spoken word intelligibility tests. While the initial version of `incom.py` (Mosbach et al., 2019) provides a baseline to calculate linguistic distances and asymmetries between related languages in visual perception, the current modification, referred to here as `incom.py 2.0`, targets the mutual intelligibility in auditory perception. The contributions of this paper include:

- Evaluation of two metrics for computing distances and asymmetries based on symbolic phonetic notations, including vowel and consonant (dis)similarities.

- Implementation of word adaptation entropy as an additional predictor of asymmetric intelligibility on the word level.

- Validation of presented predictors and explaining variables in spoken word translation tasks for Bulgarian and Russian.

2 `incom.py 2.0`

We borrow notation from Mosbach et al. (2019) and let L denote a language such as Russian or Bulgarian. Each language L has an associated alphabet – a set of characters – $\mathcal{A}(L)$ which includes the special symbol $\emptyset$. We use $w \in L$ to denote a word in language L and $c_i \in w$ to denote the i-th character in word w. Similarly, we will use $\text{IPA}(w)$ to denote the symbolic phonetic representation of w, with $\mathcal{A}_{\text{IPA}}(L)$ being the phonetic alphabet of L and $s_i \in \text{IPA}(w)$ denoting the i-th sound in word w. Moreover, let $\mathcal{V}(\text{IPA}(w))$ denote the set of all vowels in w and $\mathcal{C}(\text{IPA}(w))$ the set of all consonants in w, respectively.

LD and WAS Following Mosbach et al. (2019), `incom.py 2.0` supports computing the Levenshtein distance (LD) and word-adaptation surprisal (WAS) as well as their normalized versions nLD and nWAS between a pair of words w_i, w_j on the orthographic level. We extend these to the phonetic level by computing (n)LD and (n)WAS between $\text{IPA}(w_i)$ and $\text{IPA}(w_j)$, $\mathcal{V}(\text{IPA}(w_i))$ and $\mathcal{V}(\text{IPA}(w_j))$ as well as $\mathcal{C}(\text{IPA}(w_i))$ and $\mathcal{C}(\text{IPA}(w_j))$. We refer to Mosbach et al. (2019) for the mathematical definitions an in-depth discussion of (n)LD and (n)WAS.

Identical Phonetic Correspondences Additionally, `incom.py 2.0` supports the computation of the number of *identical phonetic correspondences* between two words w_i and w_j based on their phonetic representations $\text{IPA}(w_i)$ and $\text{IPA}(w_j)$. We compute the number of identical phonetic correspondences by first applying the LD algorithm of Mosbach et al. (2019) to $\text{IPA}(w_i)$ and $\text{IPA}(w_j)$ to obtain their alignment and then simply counting the number of identical phonetic transcription agreements.

Word Adaptation Entropy Lastly, `incom.py 2.0` supports the computation of the normalized *word adaptation entropy* (nWAE) between two words w_i and w_j. Recall that given a character (sound) $c \in A(L1)$ ($c \in A_{\text{IPA}}(L1)$) and another character (sound) $t \in A(L2)$ ($t \in A_{\text{IPA}}(L2)$), the *character (sound) adaptation surprisal* between s and t is defined as follows: $\text{CAS}(s,t) = -\log_2(P(t|s))$. Now, nWAE between two words w_i and w_j can be computed by first obtaining their aligned sequences $\widetilde{w}_i, \widetilde{w}_j$ using the LD algorithm of Mosbach et al. (2019) followed by computing the average (phonetic) character entropy as follows:

$$\text{nWAE}(\widetilde{w}_i, \widetilde{w}_j) = \frac{1}{L} \sum_{l=0}^{L-1} P(t_l | s_l) * \text{CAS}(s_l, t_l) \tag{1}$$

3 Experimental Setup

Via the INCOMSLAV platform[5], we conducted a series of mutual intelligibility tests in the spo-

---

ken modality with isolated cognate (i.e., historically or etymologically related) words. Being a challenging task on its own, the context-free recognition of cross-lingual cognates is a precondition of successful oral intercomprehension (Gooskens and van Heuven, 2020). Here, we investigate the transparency of Bulgarian–Russian cognates with regard to diachronically motivated sound correspondences that appear to facilitate or hinder human performance. As we are interested in inherent intercomprehension, only people who speak their first language (L1) natively and who do not know the stimulus language (Lx) have been included in the analysis. After completing a background questionnaire, Bulgarian (BG) and Russian (RU) participants were asked to translate randomized 120[6] oral[7] BG and RU stimuli into their respective native language in two series of 60 stimuli each. The items were taken from the written intelligibility tests of Mosbach et al. (2019) in order to obtain a reliable baseline.

The number of RU participants is 29, aged between 16 and 48 years (average age 32)[8] with 23 women, 5 men, and 1 not specified. The number of BG participants is 11, aged between 19 and 37 years (average age 27)[9] with 10 females and 1 male . Even though these two groups differ considerably in size, statistical analyses based on fewer participants are particularly worthwhile for the practice of foreign language learning and teaching and in experiments involving specific target groups (Branets et al., 2020). In this study, when no statistical observations can be made, we report findings as percentages to indicate success rates based on collected material, bearing in mind that our results should be interpreted with caution as general trends.

During the experiments the participants lis-

L1	Spoken stimuli	
	Bulgarian	Russian
Bulgarian		65.58%
Russian	68.42%	

Table 1: Intelligibility scores from free translation tasks by humans in auditory perception.

tened to the stimuli one by one (each word was played twice), and had to provide a written translation within 10 seconds. The order of stimuli was randomized. The time limit was chosen based on the experience from other intercomprehension experiments (Golubović, 2016), 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. The mean percentage of correctly translated items constitutes the intercomprehension score in each language (Table 1). These results show that the RU participants understand BG words at 68.42% and that the BG participants understand RU cognates they are presented with at 65.58%. This suggests that the intelligibility of spoken BG and RU stimuli did not cause major problems for the respective native speakers. For comparison, the intelligibility scores presented in the study of Mosbach et al. (2019) for the written modality are slightly higher, e.g. 71.33% (BG for RU) respectively 74.67% (RU for BG).

4 Results

4.1 Available Measures

The statistical analyses in Mosbach et al. (2019) clearly supported the normalized Levenshtein distance (nLD) as a reliable predictor of orthographic intelligibility on the word level for BG and RU, while the predictive potential of normalized word adaptation surprisal (nWAS) was rather weak despite its modification. From a cross-linguistic perspective, the advantage of surprisal-based methods, in contrast to Levenshtein distance, is in capturing asymmetries (Stenger et al., 2020). In this study, we validate the correlation of nLD and nWAS with the intelligibility scores obtained in oral intercomprehension experiments.

---

in German and English) carried out as challenges in a linguistic game (https://intercomprehension.coli.uni-saarland.de); for more details about intercomprehension experiments see Stenger et al. (2020).

[6]118 nouns and 2 numerals in each language.

[7]BG and RU stimuli were read aloud by Bulgarian and Russian female native speakers and recorded in a professional sound studio.

[8]In total 30 participants took part at the experiments, one male participant noticed the knowledge of Bulgarian and was excluded from the analysis.

[9]In total 13 participants took part at the experiments, two participants noticed the knowledge of Russian and were excluded from the analysis.
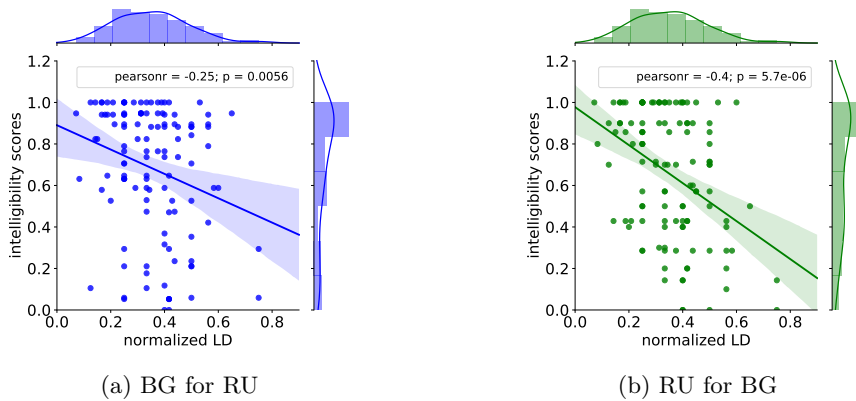
Figure 1: Relationship between speech intelligibility and normalized Levenshtein distance.

Levenshtein Distance. As the simplest version of the Levenshtein algorithm is based on binary differences between alignments, we use a modified version with vowels matched only to vowels and consonants only to consonants. However, it is not always clear how to assign the respective weights in modeling speech intelligibility (Gooskens, 2019). The optimal weighting differs for each language combination, taking into consideration the predictability and generalizability of sound correspondences as well as the human decoding process (Berthele, 2011; Gooskens et al., 2008; Kürschner et al., 2008; Möller and Zeevaert, 2015; Stenger and Avgustinova, 2020). The version that we use in this article is based on a cost matrix[10] including communicatively relevant sound distances for speech intelligibility between BG and RU (Stenger and Avgustinova, 2020).

To reveal the relationship between speech intelligibility and nLD, we correlated the results of the spoken cognate recognition tests with the calculated nLD in each language group. The basic assumption is that small distances would correlate with high intelligibility scores, while large distances are expected to correlate with low intelligibility scores. Figure 1 shows a significant negative correlation in both directions: BG for RU: Pearson's $r = -.252$, $p < .01$ and RU for BG: Pearson's $r = -.401$, $p = 5.68e-06$. This means that cross-lingual spoken word intelligibility is predictable from the nLD between the stimulus and the native language. The larger the distances the more difficult it is to recognize the cognates of a re-

lated language. However, the nLD accounts for 16.08% ($R^2 = .16075$) of the variance in the intelligibility scores for BG listeners and for only 6.33% ($R^2 = .06325$) of the variance in the intelligibility scores for RU listeners, leaving the majority of the variance unexplained. Hence, for RU, the predictive potential of nLD in the oral modality is smaller than in the written one in Mosbach et al. (2019).

Word Adaptation Surprisal. We correlated the nWAS values with the intelligibility scores to find out whether surprisal can predict human performance in spoken cognate recognition. With smaller nWAS values we expect an easier cognate guessing effect. The correlations between nWAS and the intelligibility scores are shown in Figure 2. We found negative correlations between nWAS and the intelligibility of spoken cognates for BG and RU native speakers. However, the negative effect of surprisal values was significant only for BG native speakers: BG for RU: Pearson's $r = -.127$, $p = .17$ and RU for BG: Pearson's $r = -.184$, $p < .05$ (explaining only 3.39% of the variance in the intelligibility data, i.e. $R^2 = .03389$). According to our statistical results, the predictive potential of nWAS is rather weak, which holds for both written (Mosbach et al., 2019) and spoken modalities.

4.2 Vowels Versus Consonants

Additionally, we investigated to what extent (dis)similarities in vowels and consonants may contribute to the cognate recognition process between BG and RU. As Gooskens and van Heuven (2020, p. 376) point out, consonants are better predictors of speech intelligibility

---

[10]The matrix can be found in our repository: `https://github.com/uds-lsv/incompy`
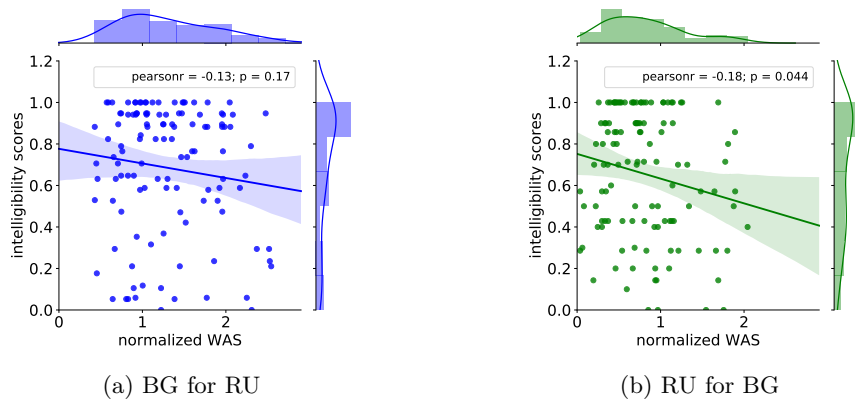
(a) BG for RU

(b) RU for BG

Figure 2: Relationship between speech intelligibility and normaized word adaptation surprisal.

than vowels, e.g., in the intelligibility of Scandinavian dialects for speakers of Standard Danish, consonants are more important than vowels (Gooskens et al., 2008; van Heuven, 2008).

We calculated the normalized vowel and consonant distances (nLD; see Section 2) and correlated these metrics with the intelligibility scores (Figure 3). The negative correlations for the vowels were significantly stronger than for consonants: Pearson's $r = -.222$ ($p < .05$) versus Pearson's $r = -.127$ ($p = .167$) for BG–RU language pair and Pearson's $r = -.390$ ($p = 1.04e-05$) versus Pearson's $r = .019$ ($p = .835$) for RU–BG language pair. Although the correlations are lower than in the case of the full nLD (considering all classes of speech sounds), the vowel distances are significant in both language groups. The BG–RU consonant distance is still negative but not significant. In the opposite direction, the RU–BG consonant distance is not negative as assumed but again not significant. It might be the case that qualitative characteristics of the mismatched consonant correspondences in spoken word recognition are more decisive than their number. In spoken cognate recognition between BG and RU, vowels turn out to be empirically more important than consonants. The impact of the total number of vowels (V) and consonants (C) in the stimulus word and of the identical vowel (Vic) and consonant (Cic) correspondences between the stimulus and the target cognates in spoken word recognition are correlated as explaining variables with the obtained intelligibility scores. Table 2 provides an overview of the statistical results (Pearson's r and p-value).

The statistical results show significantly pos-

itive correlations between the number of vowels and consonants in the stimulus cognates and the intelligibility score of the BG–RU language pair. The positive correlation for consonants is, in this case, significantly stronger than for vowels. However, the correlation for the identical vowel correspondences is significantly stronger, while the correlation for the identical consonant correspondences is positive, but insignificant.

For the RU–BG language pair, statistical results show significantly positive correlations of intelligibility scores with the number of consonants in the stimulus cognates and with identical vowel correspondences. The predictive potential of the number of vowels in auditory perception of RU cognates by BG participants is, however, small and insignificant. The correlation for identical consonant correspondences is not positive as assumed but again insignificant.

How to interpret these results? First of all, in our experimental material, the cognate BG words are slightly shorter than their RU equivalents (4.64 vs. 4.81). Looking at BG and RU sound inventories in more detail, we see an almost equal distribution of consonants between the two languages: 343 vs. 341 (2.86 per word in BG and 2.84 per word in RU). Consonants of the two languages include 224 identical correspondences (38.89%) and 122 mismatched correspondences (21.18%), including missing consonants. At the same time, the distribution of vowels is not so equally represented between BG and RU: 214 vs. 230 (1.78 per word in BG and 1.92 per word in RU), where vowels of the two languages represent only 53 identical correspondences (9.20%) and
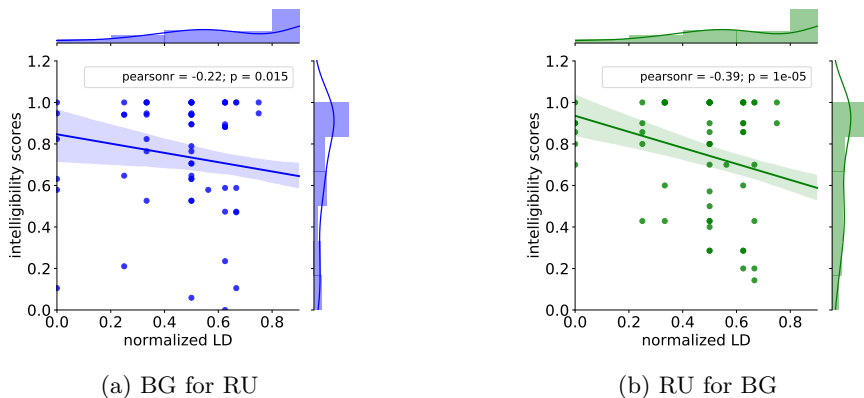
972

(a) BG for RU

(b) RU for BG

Figure 3: Relationship between speech intelligibility and nLD when considering only vowels.

Pair	Explaining variables			
	V	C	Vic	Cic
BG for RU	$.289_{<.005}$	$.317_{<.0005}$	$.225_{<.05}$	$.174_{.58}$
RU for BG	$.021_{.82}$	$.207_{<.05}$	$.306_{<.001}$	$-.005_{.95}$

Table 2: Pearson correlations ($r_{\text{p-value}}$) between intelligibility scores and explaining variables.

177 mismatched correspondences (30.73%), including missing vowels. These facts partly explain the results obtained from correlations and the different contribution of vowels and consonants to understanding spoken cognates.

To conclude, the experimental subjects seemed to rely on the interplay between a vowel (dis)similarity effect and a larger number of consonants with a balanced distribution. On the other hand, it might be the case that participants are communicatively more or less tolerant (or sensible) to different phonetic details (Gooskens et al., 2015), in particular to the qualitative effect of sound correspondences in spoken cognate recognition (Stenger and Avgustinova, 2020).

4.3 Word Adaptation Entropy

Despite all dissimilarities, speech intelligibility between related languages appears to be possible mainly because of regular correspondences. Speakers of L1 can understand a word w_j from a related Lx insofar as they can predict which word w_i of L1 is the best equivalent for w_j. In the case of cognates, the prediction can be based on sound correspondences. Entropy, as the mean of surprisal, gives a quantification of the overall uncertainty involved in making a choice. Let us consider an example. From a RU perspective, the BG–RU cognate pair [dɣp]-[dup] "oak" gives us individual sound adaptation entropies of 0.567 for [d] (BG [d] potentially corresponds to RU [d] and [dʲ]), 1.689 for [ɣ] (BG [ɣ] potentially corresponds to RU [u], [e], [o] and [ɪ]) and 0.523 for [p] (BG [p] potentially corresponds to RU [p] and [pʲ]). An aggregate measure for the entire word thus results in 2.778, which after normalization by the sound correspondence alignment gives us 0.926. From a BG perspective, on the other hand, the RU–BG cognate pair [dup]-[dɣp] "oak" gives us individual sound adaptation entropies of 0.0 for [d] (RU [d] can only correspond to BG [d]) as well as for [p] (RU [p] can only correspond to BG [p]), but 1.296 for [u] (RU [u] potentially corresponds to BG [ɣ], [ə] and [u]). As an aggregate measure for the entire word we get 1.296 and eventually normalized value of 0.432.

Based on sound entropies we have calculated the normalized word adaptation entropy (nWAE) as an aggregate measure for entire words. We assume that the smaller the nWAE, the easier it is to guess the spoken cognate in a related language. We found negative but very low and not significant correlations in both directions: BG for RU: Pearson's $r = -.078$, $p = .40$ and RU for BG: Pearson's $r = -.069$, $p = .46$. The question is why the correlations are so low and not significant? Our analysis is based on a limited number of stimulus words (120 cognates in each direction) and this sample may be too limited for nWAE values (Moberg et al., 2006). Additional linguistic factors may influence the mutual intelligibility of related languages, e.g., the existence of

words that are very similar to the stimulus but differ only in one sound, which leads to disregarding the correct counterpart (Kürschner et al., 2008). We therefore assume that WAE needs to systematically account for such neighborhood density effects as a relevant factor for speech intelligibility.

By using the nWAE values, we can now quantify the uncertainty in the overall adaptation process of cognates. For BG and RU, we found that the mean nWAE for RU given BG (1.267) is higher than for BG given RU (0.807). This means that a RU listener may in general have more difficulties when exposed to BG than a BG speaker exposed to RU. There are 114 BG cognates with higher nWAE values in comparison to 114 RU cognates. This means that RU subjects listening to 114 BG stimuli cognates are predicted to deal with more uncertainty than BG subjects listening to 114 RU stimuli cognates. Only 6 RU cognates have higher nWAE values in comparison to BG cognates. In this case BG subjects adapting 6 RU stimuli cognates are predicted to faca larger uncertainty than RU subjects trying to adapt 6 BG stimulus cognates.

To carry out comparative analyses with the small group of BG participants, we automatically extracted a corresponding sample of 11 RU subjects with matching intelligibility scores. We then calculate the difference in intelligibility between the two groups separately for all cognate pairs. The quantitative data show that 56 BG stimuli have higher intelligibility scores than their RU cognates, 50 RU stimuli have higher intelligibility scores than their BG cognates, and 14 cognate pairs have identical scores. Let us consider six cognate pairs with higher nWAE values for the stimuli, which imply larger difficulties for the listeners. There are indeed three BG stimuli with higher intelligibility scores than their RU cognates, cf. BG–RU [ɔgən]–[ɐgonʲ] "flame/fire" (91% vs. 29%), [sɔɫ]–[solʲ] "salt" (100% vs. 90%) and [kolʲano]–[kɐlʲenə] "knee" (45% vs. 29%). On the other hand, three RU stimuli have higher intelligibility scores than their BG cognates despite higher nWAS values, cf. RU–BG [gɐrox]–[grax] "pea" (50% vs. 27%), [ɪ̯ɪ̯tso]-[jəjtsɛ] "egg" (86% vs. 82%) and [porəx]-[prax] "gunpoweder" (14.0% vs. 9.0%).

To recap, the obtained experimental results indicate that the calculated nWAE values can only partially explain the asymmetric intelligibility of spoken words between the two tested languages.

4.4 Regression Analyses

In this Section we present regression analysis results. The nLD correlates most strongly with intelligibility scores in both directions (see Section 4.1). The negative correlation between the intelligibility score and the nWAS is very small and significant only for BG listeners (see Section 4.1). However, in order to investigate whether the nWAS still has a significant additional contribution to speech intelligibility in case of BG and RU, a multiple regression analysis was performed. Note that even though the nWAE does not correlate significantly with the intelligibility scores in both directions (see Section 4.3) it is included as the third predictor in the regression analysis. Figure 4 and Figure 5 present the results of the regression analysis, conducted first with the method to identify the effect of two predictors in combination: nLD and nWAS, nLD and nWAE, and nWAS and nWAE.

The first regression model containing nLD and nWAS as two predictors could account for 16.12% ($R^2 = .16117$, $p = 3.4271e{-}05$) of the variance in intelligibility of spoken RU words and for only 6.37% ($R^2 = .06370$, $p < .05$) of the variance in intelligibility of spoken BG words. The second model containing nLD and nWAE as two predictors explains the variance to a slightly better extent: RU for BG – 16.30% of the variance ($R^2 = .16296$, $p = 3.0238e{-}05$) and BG for RU – 6.59% ($R^2 = .06592$, $p < .05$). However, the third model of nWAS and nWAE does not explain the variance to a significant extent in both directions: RU for BG – 4.00% of the variance ($R^2 = .03999$, $p = .092$) and BG for RU – 1.64% of the variance ($R^2 = .01637$, $p = .381$). The final model includes all three predictors: nLD, nWAS and nWAE and this results in explaining 16.74% ($R^2 = .16742$, $p = 8.9276e - 05$) of the variance in understanding of RU spoken cognates and in only 6.60% ($R^2 = .06600$, $p < .05$) of the variance in understanding of BG spoken cognates. A combination of nLD, nWAS and nWAE is only a slightly better predictor of speech intelligibility in case
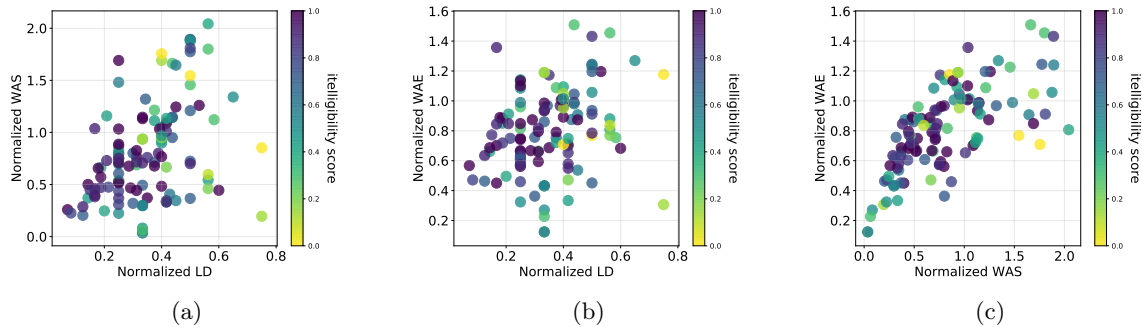
(a)   (b)   (c)

Figure 4: Relationship between speech intelligibility and nLD and nWAS (4a), nLD and nWAE (4b) , and nWAS and nWAE (4c) in the direction RU for BG. Direction BG for RU is shown in the Appendix.

of BG and RU than two predictors: nLD and nWAS or nLD and nWAE, and than nLD alone. Nevertheless, the amount of explained variance is low for all three predictors.

5   Conclusion and Outlook

We presented `incom.py 2.0`, the extended version of the `incom.py` toolkit for calculating similarities and asymmetries between closely related languages, with a focus on auditory perception. Our main conculusions are the following:

- Our statistical analyses clearly support nLD as a reliable predictor of speech intelligibility on the word level, although its predictive potential in the BG–RU setup is smaller for the oral modality when compared to the written one (Mosbach et al., 2019). The predictive potential of nWAS for speech intelligibility in case of BG and RU is rather weak and significant only for BG speakers, which is in line with the results for the written modality (Mosbach et al., 2019).

- Our special attention to the possibly different contributions of vowels and consonants in oral intercomprehension has shown that in the experimental material vowels are more important than consonants in recognizing cognates between BG and RU. A closer look at sound correspondences further reveals that a larger number of consonants with a balanced distribution between the two languages also plays a role in spoken word recognition.

- The additional linguistic asymmetry factor

of nWAE quantifies the overall uncertainty involved in making a choice in spoken word recognition. We correlated nWAE with intelligibility scores and found negative correlations between the nWAE values and the intelligibility scores in both directions. However, the correlations are very low and not significant. Hence, nWAE can only partly explain or predict asymmetric intelligibility of spoken words between the two tested languages, which we attribute to the small empirical base of the present study.

- We carried out regression analyses to find out which combination of presented predictors served better to explain speech intelligibility in our experiments. We found out that the combination of nLD, nWAS and nWAE is a slightly better predictor of speech intelligibility in case of BG and RU than two predictors nLD and nWAS or nLD and nWAE and than nLD alone. Nevertheless, the amount of explained variance is low for three predictors.

As a next step, we plan to extend the range of the studied phenomena beyond mere word similarity and to include predictability of constructional similarity by focusing on cross-lingual correspondences of multi-component units in phrasal and sentential contexts. Additionally, we plan to integrate reaction times into the experiments in order to study the interaction of linguistic and cognitive aspects of human performance.

975

Acknowledgments

References

Raphael Berthele. 2011. On abduction in receptive multilingualism. evidence from cognate guessing tasks. *Applied Linguistics Review*, 2(2011):191–220.

Anna Branets, Daria Bahtina, and Anna Verschik. 2020. Mediated receptive multilingualism: Estonian-Russian-Ukrainian case study. *Linguistic approaches to bilingualism*, 10(3):380–411.

Andrea K. Fischer, Jilles Vreeken, and Dietrich Klakow. 2017. Beyond Pairwise Similarity: Quantifying and Characterizing Linguistic Similarity between Groups of Languages by MDL. *Computación y Systems*, 21(4):829–839.

Jelena Golubović. 2016. *Mutual intelligibility in the Slavic language area*. Ph.D. thesis, University of Groningen.

Charlotte Gooskens. 2019. *Receptive multilingualism*, Language Contact and Bilingualism, pages 149–174. De Gruyter Mouton.

Charlotte Gooskens, Wilbert Heeringa, and Karin Beijering. 2008. Phonetic and Lexical Predictors of Intelligibility. *Int. J. Humanit. Arts Comput.*, 2(1-2):63–81.

Charlotte Gooskens, Renee van Bezooijen, and Vincent van Heuven. 2015. Mutual intelligibility of Dutch-German cognates by children: The devil is in the detail. *Linguistics*, 53(2):255–283.

Charlotte Gooskens and Vincent van Heuven. 2020. How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables? *Linguistic Approaches to Bilingualism*, 10(3):351–379.

Vincent J. van Heuven. 2008. Making Sense of Strange Sounds: (Mutual) Intelligibility of Related Language Varieties. A Review. *International Journal of Humanities and Arts Computing*, 2:39–62.

Klára Jágrová, Tania Avgustinova, Irina Stenger, and Andrea Fischer. 2018. Language models, surprisal and fantasy in Slavic intercomprehension. *Computer Speech & Language*.

Sebastian Kürschner, Charlotte Gooskens, and Renee. van Bezooijen. 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal for Humanities and Arts Computing*, 2(1-2):83 – 100. 2009/c.s.gooskens/pub004.

Jens Moberg, Charlotte Gooskens, John Nerbonne, and Nathan Vaillette. 2006. Conditional entropy measures intelligibility among related languages. *Computational Linguistics in the Netherlands 2006: Selected papers from the 17th CLIN Meeting*, pages 51–66.

Marius Mosbach, Irina Stenger, Tania Avgustinova, and Dietrich Klakow. 2019. incom.py - a toolbox for calculating linguistic distances and asymmetries between related languages. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 810–818, Varna, Bulgaria. INCOMA Ltd.

Robert Möller and Ludger Zeevaert. 2015. Investigating word recognition in intercomprehension: Methods and findings. *Linguistics*, 53(2):313–352.

Claude E. Shannon. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423.

Hedvig Skirgård, Seán G. Roberts, and Lars Yencken. 2017. Why are some languages confused for others? Investigating data from the Great Language Game. *PLOS ONE*, 12(4):1–35.

Irina Stenger and Tania Avgustinova. 2020. How intelligible is spoken Bulgarian for Russian native speakers in an intercomprehension scenario? In *Proceedings of the International Annual Conference of the Institute for Bulgarian Language*, volume 2, pages 142–151, Sofia.

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. European Language Resources Association.

Irina Stenger, Klára Jágrová, Tania Avgustinova, Dietrich Klakow, and Roland Marti. 2017. Modeling the impact of orthographic coding on Czech–Polish and Bulgarian–Russian reading intercomprehension. *Nordic Journal of Linguistics*, 40(2):175–199.

Appendix

Figure 5 shows the relationship between speech intelligibility and the various predictors introduced in Section 2 in the direction BG for RU.
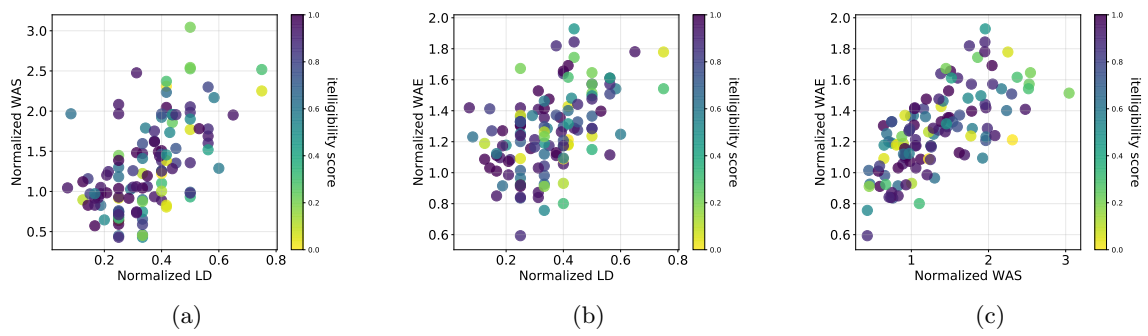
Figure 5: Relationship between speech intelligibility and nLD and nWAS (5a), nLD and nWAE (5b) , and nWAS and nWAE (5c) in the direction BG for RU.

977