Chapter 5: Linguistic and extra-linguistic predictors of intelligibility among six Slavic languages

Abstract: In this chapter we tried to predict intelligibility using linguistic factors (lexical, orthographic, phonological, morphological and syntactic distances) and extra-linguistic factors (language attitudes and language exposure). Intelligibility was expressed as the number of correct answers in the cloze test or the word translation task, both in the written and the spoken mode. Using stepwise regression analyses we could predict 66% of the variance for the results of the written word translation task, 66% for the spoken word translation task, 44% for the results of the written cloze test and 66% for the spoken cloze test. The best predictor of intelligibility in all four cases was phonological distance. Lexical distance would have been a close second, but due to the high degree of collinearity with phonological distance, it could not account for much of the additional variance. Language exposure also proved to be an important predictor, despite the fact that there is not much exposure among the languages we observed to begin with. Morphological and syntactic distance were significant predictors for intelligibility of texts. Taken together, these results show that intelligibility can be predicted to a great extent using a combination of linguistic and extra-linguistic factors, although the contribution of individual factors may vary depending on the language family.

1. INTRODUCTION

In Chapter 3 we measured intelligibility levels among six Slavic languages on the level of individual words, sentences and larger discourse and across most conditions, we could observe similar patterns. Why can speakers of Czech understand Slovak better than Polish and understand Polish so much better than Bulgarian? Obviously, it is due to differences in language structure, which are generally larger in languages belonging to different sub-families. According to Tang and van Heuven (2015):

"An adequate theory of language should be able to predict the approximate degree of intelligibility of a language A for a native listener of a (related) language B by means of a systematic comparison
of the similarities and differences between the languages concerned in terms of their vocabulary, syntax, morphology, phonology and phonetics.” (p. 285)\textsuperscript{15}

Comparing the linguistic structure of language A and language B could lead to generally accurate results if those languages existed completely independently from one another. But since no language is entirely isolated, we also must take into account factors which go beyond the language structure. Is language A commonly taught in schools across country B? Is there a border area where speakers of language A and language B often interact with each other? Do speakers of language A have a generally negative attitude to speakers of language B? Some of these factors might play a role in intelligibility and the size of that role would depend on the specific language situation.

In this chapter we expanded and refined the regression models from Chapter 4 by adding the results of the translation tasks as the dependent variables and by using both linguistic and extra-linguistic predictors. We showed that the level of intelligibility of both written and spoken language can be predicted to a substantial degree. In §2 we provide an overview of previous research on factors influencing intelligibility; in §3 we present our aims and hypotheses and in section §4 we outline the methods for measuring intelligibility. The linguistic predictors are presented in §5, while the extra-linguistic predictors are described in §6. The results will be presented in §7 followed by discussion and conclusions in §8.

2. PREVIOUS RESEARCH

Attempts to predict intelligibility between language pairs on the basis of their linguistic (dis)similarity have generally been successful. Tang and van Heuven (2007) looked at 15 Chinese varieties and found that subjective judgments of linguistic similarity could predict judged intelligibility to a fairly high degree ($r = 0.89$). Tang and van Heuven (2009) collected intelligibility data using functional testing both on the word and on the sentence level and found that the scores could be predicted to a high extent by lexical similarity and phonological correspondences. In Table 5.1, we provide an overview of studies dealing with the general topic of factors influencing intelligibility.

\textsuperscript{15} In the cited excerpt, Tang and van Heuven were talking about inherent intelligibility, which excludes extrinsic factors such as prior exposure to the other language as predictors. (V. J. van Heuven, personal communication, February 29, 2016)
Tang and van Heuven (2015) included other predictors such as sound inventories, lexical frequencies of the coda consonants, phonetic distance measured by Levenshtein algorithm, etc. They were able to account for around 61% of the variance in intelligibility between 15 Chinese dialects with just two predictors: one lexical (the percentage of cognates between dialect pairs) and one phonological (lexical frequency of syllable rhymes), both computed on a large database. Adding a few other predictors (lexical frequencies based on the Chinese Academy of Social Sciences database, phoneme inventories etc.) resulted in 87% of the total variance explained.

Phonetic distance is the most commonly found significant predictor of intelligibility across the board. Gooskens (2007a) looked at the intelligibility of Danish, Swedish and Norwegian and used language attitudes, language contact and phonetic distances measured by Levenshtein algorithm as potential predictors. She found that phonetic distance alone could account for 66% of the total variance in intelligibility scores. Gooskens (2007b) tested the intelligibility of Dutch, Frisian and Afrikaans as well as Danish, Swedish and Norwegian using a test that involved listening to a text and then answering five open-ended questions about it. Phonetic and lexical distances were used as potential predictors of intelligibility and the results revealed a negative correlation between phonetic distance and intelligibility ($r = -0.80$), but no significant correlation was found between lexical distance and intelligibility. The reason behind this is the fact that lexical distances in both cases were so small that they were not explaining any additional variance.

Gooskens, Heeringa, and Beijering, (2008) also found that phonetic distances were a better predictor of intelligibility than lexical distances in the case of intelligibility of different Scandinavian varieties by native speakers of Danish. Their study focused on phonetic distances in more detail and revealed that the correlation between consonant distances and intelligibility was stronger than the correlation between vowel distances and intelligibility ($r = -0.74$ vs. $r = -0.29$). Lexical distances were excluded as a predictor in a stepwise regression analysis, but since they did correlate significantly with intelligibility, in Table 5.27 we will place the study together with the others which have shown a relationship.

Kürschner, Gooskens, and van Bezooijen (2008) looked at a number of linguistic factors potentially influencing intelligibility. They used a word translation task to measure the intelligibility of Swedish for Danish speakers. The best predictor was once again phonetic distance between the two languages, but other significant predictors include word length, number of syllables, foreign sounds, word frequency, neighborhood density, the use of native language orthography, and the absence ‘stød (glottal stop) in Swedish. Schüppert, Gooskens, Hilton, and van Heuven (2012) discovered even more linguistic factors influencing the intelligibility of Danish for native speakers of Swedish: syllable reduction and articulation speed which is generally higher in Danish than in Swedish.
Hilton, Gooskens, Schüppert (2013) examined the influence of non-native morpho-syntax on the intelligibility of Norwegian for native speakers of Danish in sentence plausibility experiment. They presented Danish participants with sentences where some of them had morpho-syntactic constructions which were grammatical in Norwegian, but not in Danish and measured both the correctness of judgements and response times. They found a significant effect of foreign morpho-syntax, but phonology proved to be an even more important factor influencing intelligibility.

Table 5.1: An overview of research looking into factors influencing intelligibility

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relationship found</th>
<th>No relationship found</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonetic distances</td>
<td>Gooskens (2007a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gooskens (2007b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tang &amp; van Heuven (2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kürschner, Gooskens, &amp; van Bezooijen (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gooskens, Heeringa, &amp; Beijering (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hilton et al. (2013)</td>
<td></td>
</tr>
<tr>
<td>Other linguistic factors:</td>
<td>Kürschner, Gooskens, &amp; van Bezooijen (2008)</td>
<td></td>
</tr>
<tr>
<td>(lexical frequency, phoneme inventories of different varieties, word length, different number of syllables, foreign sounds, neighbourhood density, word frequency, the use of native language orthography, asymmetric perceptions of sound correspondences, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gooskens, Heeringa, &amp; Beijering, (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schüppert, Gooskens, Hilton &amp; van Heuven (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tang &amp; van Heuven (2015)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gooskens et al. (2015)</td>
<td></td>
</tr>
<tr>
<td>Morpho-syntax</td>
<td>Hilton et al. (2013)</td>
<td></td>
</tr>
<tr>
<td>Speech rate</td>
<td>Hilton, Schuppert &amp; Gooskens (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schüppert, Hilton &amp; Gooskens (2015)</td>
<td></td>
</tr>
</tbody>
</table>
Gooskens, van Bezooijen, and van Heuven (2015) investigated which linguistic factors influence the cross-language recognition of Dutch-German cognates. They made every effort to exclude any influence of extra-linguistic factors by looking at children who have not had any exposure to the other language and held equally positive attitudes. Dutch participants were still significantly better at recognizing German cognates than vice versa, which the authors ascribe to the number of lexical competitors, phonetic detail and asymmetric perceptions of sound correspondences.

The relationship between extra-linguistic factors, such as language exposure or language attitudes and intelligibility is much more elusive. The general rationale is clear: the more positive the attitude towards a language and/or its speakers, the more likely a person is to make an effort to understand that language; the more exposure a person has to a particular language, the more likely she is to understand it. Still, the demonstrated relationship between these two potential factors and intelligibility has been weak at best.

In their study on Swedish-Danish intelligibility, Delsing and Lundin Åkesson (2005) cite the differences in language attitudes and language exposure as the main reason for the asymmetric intelligibility between Swedish and Danish, but they do not report correlations between these factors and intelligibility. Gooskens (2006) analyzed a part of their data and found a correlation between the ratings of how beautiful a language is and intelligibility of that language ($r = .56, p = .02$).

Gooskens and van Bezooijen (2006) tested the mutual intelligibility of Dutch and Afrikaans. They found some relationships between intelligibility scores and language attitudes. Nevertheless, language attitudes could explain only up to 20% of the total variance in the intelligibility scores, thus showing that linguistic factors are more important. In the van Bezooijen and Gooskens (2007) study of intelligibility of written Frisian and Afrikaans by Dutch participants, a more positive attitude to these two languages did not mean a better intelligibility score.

Impe (2010) studied the intelligibility of Dutch in different areas of the Netherlands and Belgium. She reported a very low significant correlation between attitudes and intelligibility $r = .11, p = 0.001$. Gooskens and Hilton (2013) investigated the intelligibility of Danish for Norwegian subjects from the north and south of the country using a spoken word translation task. They found no correlation between the mean intelligibility scores and the attitude scores in the data.

Schüppert and Gooskens (2011) presented nouns in a multiple choice picture-pointing task with Swedish and Danish pre-schoolers and elicited their attitudes to the test language. They found no significant correlation between intelligibility and attitudes. On the other hand, after controlling for age in the same dataset, Schüppert, Hilton, and Gooskens (2015) found a low but positive cor-
relation between attitudes and intelligibility. Still, the attitude variable expressed as the extracted attractiveness dimension of the speaker in a matched-guise task, could explain only 3.6% of the variation in the intelligibility scores.

The role of language exposure is relatively straightforward: the more exposure to a related language the participants have, the more likely they are to understand that language. This variable refers to actual learning of a related language in a classroom setting or in a self-paced way and to everyday exposure in the form of listening to people speak a related language, reading materials or listening to music in that language, etc. Still, this relationship was not confirmed in Gooskens (2006), Gooskens (2007a) or Gooskens and Hilton (2013). On the other hand, in Ház (2005) Dutch participants were better at understanding spoken German than vice versa, simply because German is commonly taught in Dutch schools. Here it is worth mentioning one of the rare studies from the Slavic languages area, Dickins (2009) compared the results of his study, with the results of Tejnor (1971). In both cases, only opinion testing was used, but in those four decades, the percentage of Czech speakers who claimed they could understand Slovak soared from 12% to 71%. From the 1970s to the 1990s, Czech and Slovak were spoken in the same country and there was plenty of exposure to both languages. After the breakup of Czechoslovakia, the situation changed slightly, up to the point that young Czech speakers and foreigners, who have not had much exposure to Slovak seem to have some troubles understanding it (Berger, 2003).

3. AIMS AND HYPOTHESES

The main aim of the present paper is to develop a statistical model of both written and spoken cross-language intelligibility of selected Slavic languages. Using both linguistic predictors, pertaining to the languages themselves, and extra-linguistic predictors, pertaining to the participant's opinions and experiences with those languages, we strive to come to a model that can predict the level of intelligibility among six Slavic languages as well as possible. The six languages we work with are Czech, Slovak and Polish (belonging to the West Slavic sub-family) and Croatian, Slovene and Bulgarian (belonging to South Slavic sub-family), with the plan to include other Slavic languages, such as Russian, at a later point.

Based on the results presented in the previous section, it can be concluded that extra-linguistic factors are probably not vital for predicting intelligibility in the Slavic language area: firstly, the relationship between attitudes and intelligibility seems to be relatively weak; secondly, in the case of the six languages from our study, language exposure is relatively limited; and thirdly, including linguistic factors as the only predictors of intelligibility results in a comparatively good model of
meAsuring intelligibility, as we have demonstrated in Chapter 4. Nevertheless, extra-linguistic factors might account for some of the variance in intelligibility results that linguistic factors alone could not explain, which means that including them as potential predictors could mean arriving at a slightly better model of intelligibility than we would with linguistic factors alone. For this reason, we choose to include language attitudes and language exposure in our proposed model of intelligibility.

In the present study, the potential linguistic predictors are lexical distances, phonological distances, orthographic distances, morphological distances and syntactic distances and the potential extra-linguistic predictors are language attitudes and language exposure. We make a distinction between understanding individual words and understanding sentences or texts, as we believe that in the latter linguistic factors such as morphology or syntax might play a role.

We expect all the factors included in the model to be significant predictors of intelligibility. The amount of explained variance will inevitably vary, so in line with previous research we expect that phonological and lexical distances will account for most of the variance in intelligibility scores. Since Slavic languages are characterized by rich inflectional morphology, we hypothesize that morphological distances will also play an important role. The role of syntax as a predictor of intelligibility has not been investigated before, but on the basis of the results from Chapter 4, §3, we expect syntactic distances to be a significant predictor in the present model as well. As for language attitudes, they will certainly not be the most important predictor, but at the very least we do expect to find a statistically significant positive relationship between attitudes and intelligibility. The influence of language exposure in the model depends on the actual amount of exposure: the more exposure there is, the more likely it is that this variable will outweigh all the linguistic factors. Since there is relatively little exposure between the six Slavic languages the present study focuses on, we hypothesize that language exposure will still be a significant predictor, but not the most important one (as might well be the case in a model involving some other languages).

4. MEASURING INTELLIGIBILITY

4.1. The word translation task

4.1.1. Method

Variants of the word translation task have been used extensively in intelligibility research. In the most common version, the participants are simply asked to translate the words they see or hear (Gooskens, Kürschner, & van Bezooijen, 2011; Gooskens & Hilton, 2013), but sometimes the task
is to perform semantic categorization of the presented words (Tang & van Heuven, 2009), or to select the correct translation in a multiple-choice setting (Schüppert, Hilton, & Gooskens, 2015).

We conducted the intelligibility experiments using an online application. The translation task used in the present study represents the most basic version of the method: the participants were asked to type a translation of the words presented to them. Since we were interested both in the written and the spoken language, the task had both versions: in the former, the participants would see the words on the screen, while in the latter they would hear recordings of each word. For more details on the testing material, see Chapter 3, §4.1.

4.1.2. Participants

Our primary target group was young adults. Since the whole experiment was online, we recruited participants through their universities, mailing lists and social networks. Anyone could do the task, but for the actual analysis we filtered the data on the basis of the following criteria: the participants had to be native speakers of Croatian, Slovene, Bulgarian, Czech, Slovak or Polish; aged between 18 and 30; who only speak the language they indicated as their native language at home; have completed at least high school education; and have never learned the test language.

The mean age of participants who did the written word translation task is 23 years and roughly one third of the sample was male. Seventy-four percent of the participants have had some higher education, either academic or vocational. According to the data from the UNESCO Institute for Statistics, the percentage of the total population of each of the relevant six countries who enrolled in tertiary education in the period of up to five years from their high school graduation ranges from 54% in Slovakia to 84% in Slovenia (the data refers to 2013, with the exception of Croatia, for which the most recent numbers available were for 2012). This indicates that our sample is representative of the educated young population of the six countries.

The written word translation task was completed by a total of 999 participants. Each test was completed by at least 22 participants, with the exception of Slovene and Slovak participants translating

written Bulgarian words, where the number of participants is 16 and 19, respectively. A complete overview of the participant characteristics broken down by native language-test language combination can be found in Appendix G.

A generally low number of participants who did the written word translation task in Bulgarian can partly be explained by the fact that Bulgarian is the only language in our interest group which is written in Cyrillic. In order to be able to conduct the written test in Bulgarian, we only assigned it to those participants who stated they could read Cyrillic in the background questionnaire. Since this is a fairly small group of people, the participant numbers were bound to be small as well.

The spoken word translation task was done by a total of 1074 participants. Their mean age was 22.86 years and once again, around 33% of the sample was male. Around 76% of the sample has had some higher education. Each unique task was done by at least 21 participants. A more complete overview of the participant numbers and characteristics is in Appendix G.

4.1.3. Procedure

All the participants began the experiment by completing a background questionnaire where they were asked about their age, sex, level of education, but also information about their language competencies, the geographic proximity of speakers of a different language etc. Next, they would get questions designed to measure their attitudes to other Slavic languages and their amount of exposure to them. More information on the methods for measuring these variables will be provided in §6.

The participants were assigned a test language randomly (except in the case of written Bulgarian, where the condition was familiarity with the Cyrillic script) and they were asked if they had ever learned that language. In case they had, they were excluded from further analysis. They did either the written or the spoken version of the task, never both. The task consisted of translating a random mix of 50 out of 100 words from our list, so each participant translated a unique set of words.

In both the written and the spoken version of the task, the participants were given 10 seconds to translate each word. If they finished more quickly, they could either click on the “Next” button or press Enter on their keyboards. The time limit was carefully piloted with people whose typing speeds varied greatly, and we settled for the time which seemed to be sufficient for typing even the longest words, but not long enough for using a dictionary. In order to deter participants from using online translators, copying and pasting words was disabled.
In the spoken translation task, each word was heard twice, the rationale being that in a real-life situation one could reasonably ask the interlocutor to repeat what was said once, but mostly not more than that. To make sure all participants received the same input, the space reserved for typing a translation would appear only after a word was played both times.

### 4.2. The cloze test

#### 4.2.1. Method

The cloze test is a task whose objective is to fill in the gaps in a text. It could be completely open-ended, but in most cases, the format is either multiple-choice (several options for each gap) or the omitted words are given in a random order and the participant's task is to place them back into the correct gaps in the text. Completing the task successfully implies that the participant can: 1) understand the words in question; and 2) understand the context well enough to be able to place the words appropriately.

The testing material for our version of the cloze test came from B1 level texts found online\(^{17}\) and then adapted by shortening the most complex sentences and making sure that all four texts used consisted of 16 or 17 sentences and were about 200 words long in total. The development of the cloze test material is described in more detail in §4.1 of Chapter 3.

The spoken version of the cloze test was recorded using the same speakers as for the word translation task material. The voices of four speakers which were rated most positively in the voice parade were used in the actual experiment: one voice for each text (see §4.1.1 above).

#### 4.2.2. Participants

A total of 938 participants took part in the written cloze test. Their mean age was about 23 years and a bit more than 35% of participants were male. The percentage of participants who have had some higher education was 73%. Each individual task was done by at least 20 participants, with the exception of Slovene participants who did the task in Bulgarian, where the number is 19. Since we used the same dataset for the analysis on the language level, a complete overview of participant groups can be found in Appendix G.

\(^{17}\) [http://esl-bits.net/pet.htm](http://esl-bits.net/pet.htm) last retrieved on 13.08.2015.
The spoken cloze test is the task that was done by the lowest number of participants, 823 in total. In most groups we still had more than 20 participants, except from Bulgarian native speakers who were the least numerous group. Just as before, the average age of participants is around 23 years and roughly one third of the sample is male. The percentage of participants with a completed or ongoing tertiary education was 77%. The full table is provided in Appendix G.

### 4.2.3. Procedure

After completing the background questionnaire and the language attitude section (see §6.1), the participants would proceed to the task. In the written version of the cloze test, the participants could see the whole text in front of them at all times. They were instructed to move the 12 words to the right gaps in the text by dragging and dropping them within 10 minutes. The participants could get the translations of the target words by pointing at them with their cursor. After placing a word in one of the gaps in the text, that word would becomegreyed out in the selection area. In case the participants wanted to change an answer, they could simply drag and drop a different word into the same gap, which would cause their original word of choice to re-appear in black in the selection area above the text.

In the spoken cloze test the gap took the form of a beep of uniform length (one second, with a 30ms pause before and after it). The spoken cloze test was played in fragments of one or two sentences, where each fragment contained only one gap. The fragments were repeated twice and only then would the participants see 12 words on the screen. A selection had to be made within 30 seconds, otherwise the response was recorded as blank. Any word used was greyed out, but it could be reused if needed – in a similar manner as for the written cloze test.

### 5. LINGUISTIC PREDICTORS

#### 5.1. Lexical distances

We measured the lexical distance among Croatian, Slovene, Bulgarian, Czech, Slovak and Polish by counting the number of lexical non-cognates between each language pair and turning it into a percentage. For our purpose, the term “lexical non-cognates” is taken to mean words with the same root and generally the same meaning.

The dataset used for this measurement is our actual testing material – the list of 100 nouns (which will be used to predict the results of the word translation tasks) and the four texts (which will be used for predicting the results of the cloze test). As in the previous chapters, in the context of
linguistic distance measurements, we shall refer to the list of 100 nouns as the list data and to the four texts as text data. We would begin by using the original word list in one language e.g. Croatian and try to find any counterparts in the other five languages. The main idea behind this approach is to mimic the mind of a participant reading or listening to those words in Croatian and trying to find any counterpart word in their native language even if it is not a perfect translation. Aligning the words for each language separately enabled us to pinpoint asymmetries between language pairs e.g. the lexical distance between Croatian and Polish might be smaller than the other way around. The ultimate goal of making this distinction is creating a more sensitive model of intelligibility.

In Example 1 below, all the words have been aligned to the Croatian word for ‘woman’, žena. In Slovene, Bulgarian, Czech and Slovak the words are perfect counterparts, they are similar in form and share exactly the same meaning. In Polish, the word żona actually means ‘wife’. We decided to include it because it still points the reader/listener in the right direction.

<table>
<thead>
<tr>
<th></th>
<th>Croatian</th>
<th>Slovene</th>
<th>Bulgarian</th>
<th>Czech</th>
<th>Slovak</th>
<th>Polish</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) žena</td>
<td>ženska</td>
<td>жена</td>
<td>žena</td>
<td>žena</td>
<td>żena</td>
<td>żona</td>
</tr>
</tbody>
</table>

In the Example 2, we demonstrate exactly the same procedure, only this time with Polish as the test language. The Polish word for ‘woman’ is kobieta and it does not have a counterpart in any of the other five languages.

<table>
<thead>
<tr>
<th></th>
<th>Polish</th>
<th>Croatian</th>
<th>Slovene</th>
<th>Bulgarian</th>
<th>Czech</th>
<th>Slovak</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) kobieta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

By comparing these two examples it is clear that a Polish speaker reading or listening to the word žena in Croatian or ženska in Slovene has a general idea of a female, while a Croatian or Slovene speaker who has not been exposed to Polish would have no clue what kobieta might mean. With enough cases such as this one, two languages could have an asymmetry in their lexical distance, i.e., language A might share more of language B’s vocabulary, but not vice versa, which in turn might yield asymmetric intelligibility.

Our measurements resulted in a distance matrix, with potentially different values for each language combination (i.e. Croatian-Slovene and Slovene-Croatian distance do not have to be the same). For more information on all the distance measurements, see Chapter 2, §4.
5.2. Orthographic distances

For calculating orthographic distances between the readers’ language and the stimulus language we used the Levenshtein algorithm (Levenshtein, 1966). The Levenshtein distance represents the number of insertions, deletions and substitutions that need to be performed in order to transform one string into another (Kruskal, 1999). We will illustrate the algorithm through a comparison of the Slovak word človek with the Polish word człowiek (‘man’).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slovak</td>
<td>č</td>
<td>l</td>
<td>o</td>
<td>v</td>
<td>e</td>
<td>k</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>c</td>
<td>z</td>
<td>l</td>
<td>o</td>
<td>w</td>
<td>i</td>
<td>e</td>
<td>k</td>
</tr>
</tbody>
</table>

We aligned the words by matching a vowel with a vowel, and a consonant with a consonant. Once the words are aligned, we looked at the operations necessary to get from the Slovak word to the Polish one. The differences in the base are weighed as 1; for example: a versus e, c versus s. But if two characters have the same base, but different diacritics, we weigh this as 0.3. For example: a versus á, s versus š. The choice of 0.3 as weight is somewhat arbitrary, but it is motivated by the idea that diacritical differences confuse the reader to a much smaller extent than the differences in the base.

The normalized Levenshtein distance is calculated when the total number of operations is divided by the total length of the alignment. We normalize for word length because a single change should weigh much more in a three-letter word than it would in a ten-letter one. So in our example the distance would be 3.3/8 = 0.41 or 41%. The distances were calculated only on the basis of counterpart words and the total orthographic distance between pairs of languages was obtained by calculating the average distance for all the words in our dataset. We used the same tables as when measuring lexical distances, which enabled us to pinpoint asymmetries in orthographic distances as well. For each language pair we calculated the aggregated orthographic distance, by calculating the average of the normalized Levenshtein distances of the word pairs which are considered for each language combination. Orthographic distances were calculated both on the basis of list data and on the basis of text data, but in the latter case, the results were presented in terms of stems and affixes.

In case of Bulgarian, which is the only language written in Cyrillic only, we used transliterated data. The participants in the intelligibility experiment only got a task in written Bulgarian if they had previously indicated that they could read Cyrillic. Because of that, we felt that the orthographic distances between transliterated Bulgarian and the other five Slavic languages best reflected the reality of, e.g., a Slovak participant who could read Cyrillic doing the intelligibility task in Bulgarian.
5.3. Stem and affix distances

When modelling intelligibility of a text, morphology might be an important factor, especially in a language family such as Slavic. Heeringa, Swarte, Schüppert and Gooskens (2014) found no significant correlation between the orthographic stem variation and orthographic affix variation in the case of Germanic language family. This might indicate that orthographic stem and affix distances might serve as predictors of intelligibility separately. That is why we decided to break the orthographic distance measurements based on the text data into two. Orthographic stem distances measured on the basis of text data will serve as a potential "orthographic" predictor of intelligibility (as measured by the cloze test). Orthographic affix distances will serve as a measurement of differences in morphology. Herceforth we shall refer to orthographic affix distances as morphological distances, but we would like to remind the reader that these measurements are still orthographic in their nature. When measuring stem and affix distances we used the Levenshtein algorithm in the same way as described in §5.2.

5.4. Phonological distances

Phonological distances are measured by means of Levenshtein algorithm in the same way as orthographic distances, the only difference being that the input are transcribed words, rather than orthography. The transcriptions were made in IPA and the transcribers were native speakers of the six languages who had a background in phonology. The full phonemic inventory of the six languages we focused on can be found in Appendix C. Due to time constraints, the phonological distances were calculated on the basis of list data only and those results will be used as predictors of intelligibility measured by the word translation task as well as the cloze test. In Chapter 2, §3.2, we compared the lexical and orthographic distances obtained on the basis of the list data with those obtained on the basis of the text data. Since the correlation was .90 for lexical distances and .94 for orthographic distances, we proceeded with the use of the list dataset for measuring phonological distances.

The pronunciation distance between two languages is calculated on the basis of the counterpart pairs. The algorithm itself functions in much the same way as with orthography, but the operation weights we used were slightly different. We used graded weights i.e. segment distances, meaning that the pair \[i, ɔ\] is considered to be more different than the pair \[i, ɪ\]. The segment distances are obtained on the basis of Heeringa (2004). It may seem somewhat arbitrary that we used more sensitive weights for phonological distances, but the equivalent of the procedure would be to use weights on the basis of the letter shape for orthographic distances and there is no universally accepted measurement of that available.
5.5. Syntactic distances

Our syntactic distance measurements were based on Nerbonne and Wiersma (2006) and Lauttamus, Nerbonne, and Wiersma (2007). Their method is based on frequency profiles of trigrams of part-of-speech (POS) categories. In short, all the data from our testing material was tagged for parts of speech. Next, the texts were analyzed as sequences of three words (trigrams). Since many trigrams appear relatively often (consider for instance the determiner-adjective-noun trigram such as *a blue dress*) their total number is calculated. A frequency vector for each language was created as a unique combination of a certain number of trigrams. Next, the frequency vectors are correlated, the rationale being the higher the correlation between them, the greater the syntactic similarity between the pair of languages in question. A distance measurement can be easily created by subtracting the correlation coefficients from 1. For more details on the measurement, see Chapter 2, §4.5.

6. EXTRA-LINGUISTIC PREDICTORS

6.1. Language attitudes

We measured language attitudes as the rating of how beautiful the participants found their test language. In order to make sure that all participants were familiar with the language they were rating, first they listened to a short sound fragment. The fragment was the first article of the Universal Declaration of Human Rights, recorder by the same four speakers we used for recording the testing material (each participant would get a random voice out of the four).

The participant’s task was to rate the beauty of the language they had heard on a 5-point semantic differential scale ranging from “very ugly” (1) to “very beautiful” (5).

6.2. Language exposure

Even though the participants may never have learned the test language, they could still have picked up a thing or two simply by listening to the language, reading in it, encountering it on a holiday etc. The language exposure variable was designed to measure exactly that. We asked the participants to say how often they have encountered the test language in the last five years in six different contexts:

- listening to people speaking in their presence (e.g. on a vacation, at work, doing shopping, etc);
- watching television, DVDs, movies;
• playing computer games;
• chatting or surfing on the internet;
• talking to people in person/on the telephone/on Skype
• reading books, newspapers, magazines, text on a computer screen

The participants answered by filling out 5-point semantic differential scales ranging from “never” (1) to “every day” (5). The final results were obtained by using the mean of all six scores.

7. RESULTS

7.1. Tests and predictors

Since we are dealing with two different types of intelligibility tests, one pertaining to the intelligibility of individual words and the other to the intelligibility of sentences and higher discourse, we used slightly different predictors. Before we present the results, we shall briefly focus on these different sets of predictors.

Word translation tasks deal with individual words, all of them nouns, therefore morphological and syntactic distances were not applicable as predictors there. We used lexical and orthographic distances as predictors of intelligibility measured by both methods, only in the case of word translation tasks, these measurements were based on the list data, while for the cloze test, they were based on the text data. As previously mentioned, the results of phonological distances were based on the list data only and as such, they will be used both in the case of the cloze test and the word translation task.

We decided not to distinguish between predictors for the intelligibility of written and spoken language, i.e., in both cases we included phonological and orthographic distances. While this may seem unorthodox at first, we must take into account that alphabets in Slavic languages, regardless of whether they are Cyrillic or Latin, follow the one phoneme-one grapheme rule to a large extent. Dual-route hypothesis of written word recognition proposes that phonological processing has a role in reading aloud (Coltheart, 2005) and this might be even more pronounced in languages where simply following a set of rules often results in a correctly read word. Therefore, we propose that phonological distances might be a good predictor of intelligibility even for tests based on written language.
As for the spoken language, both the word translation task and the cloze test in the spoken mode actually have a written component: in the case of the word translation task, our participants are asked to type out the correct translation of the words they hear, while in the spoken cloze test, the target words which should be placed where the beeps are heard are written on the participants’ screen. Thus orthographic distances might play a role in the intelligibility of spoken words or texts. Another reason for including orthography in the model of spoken intelligibility is that the participants might be using their knowledge of orthography in their native languages in order to decode a related language, as demonstrated in Kürschner, Gooskens, and van Bezooijen (2008) and Schüppert (2011). A complete overview of linguistic predictors used for each of the four texts is provided in Table 5.2.

As for extra-linguistic predictors, we used both language attitude and language exposure as predictors for the results of each of the four tests, but since different participants took part in each of the tests, it is important to note that these scores are actually participant-dependent (unlike all linguistics scores, which are the same for all the speakers of language A exposed to language B).
7.2. Language level vs. participant level

In Chapter 3, we conducted the regression analyses on the level of different languages, i.e. our aggregated dataset consisted of 30 cases, which represented 30 language combinations, e.g. native speakers of Czech reading a text in Slovak was one data point. The decision was made on the basis of our predictors, which vary only on language level (the lexical or syntactic distance from Czech to Slovak is the same for all participants) as well as our desire to make the results comparable with previous research that also relied on aggregated datasets. In this chapter however, we include language attitudes and language exposure as predictors, and they are participant-dependent since participants provided their own individual answers. Therefore, there are two different ways to conduct the analysis. One is at the level of language combination (group level), so we would work with a total of 30 cases and for language attitudes and language exposure we would take the mean for each language combination (the linguistic distances are language combination specific already). This represents the approach from Chapter 4. When aggregating a dataset by looking at the groups rather than individuals and using the mean results, the variation gets flattened, which generally leads to better (but inflated) $R^2$, which is why here we opt for doing the analysis on the participant level. After all, a speaker’s intelligibility is an individual phenomenon and an adequate model of intelligibility should be able to predict scores of individual speakers rather than speaker groups.

7.3. Single predictors of intelligibility

In order to see which one of our predictors correlates with intelligibility results best, we computed correlation coefficients for the results of the translation tasks and the cloze test. Full correlation matrices can be found in Appendix H.

The correlations were calculated separately for each test since we used different predictors depending on the task as well as because different participant groups did all four tests, therefore extra-linguistic factors had different values for each. Also, the correlation between e.g. lexical and phonological distances will inevitably vary slightly depending on the task, due to a different total number of cases. Since the analysis was not done on the aggregated datasets, this also means that the correlation coefficients were generally lower here than in Chapter 4.

In all four tests, the highest correlation was found between phonological distances and intelligibility scores, despite the fact that we measured them on the basis of the list data only. The correlation coefficient was $r = .59$ in the case of the written cloze test and around $.75$ for the other three tasks (all $p < .001$). The second best correlation was between lexical distances (measured on the basis
of list data for the two translation tasks and on the basis of the four texts for the cloze tests) and
the intelligibility scores, with only slightly smaller correlation coefficients. Here it is important to
note that phonological and lexical distances were extremely highly correlated (around .90 in all
four cases), which means that one of these variables will probably have to be excluded from the
intelligibility model.

When it comes to other predictors, we found moderately high correlations between orthographic
distances and intelligibility ($r$ between $-.498$ and $-.625$, $p < .001$) and language exposure and
intelligibility ($r$ between $-.519$ and $-.690$, $p < .001$), whereas the correlations between language
attitudes and intelligibility were always rather low (with $r < .3$, $p < .001$).

### 7.4. Multiple regression analyses

In order to create a model of intelligibility, we ran a series of stepwise regression analyses using the
linguistic and extra-linguistic factors as independent variables and the results on the four tasks as
dependent variables. Since completely different participants took the four tests, we ran four separate
regression analyses, one for each test.

In general, we were able to explain up to 66% of the variance in the intelligibility scores of in-
dividual participants. Here the written cloze test appears to be the odd one out, since there we
could account for only 44% of the variance. An overview of the four regression analyses with the
adjusted $R^2$ values and the predictors for the best models are provided in Table 5.3, while a more
detailed overview with $B$ and beta scores for each predictor and AIC values for every model can
be found in Appendix I.

When it comes to the models for written and spoken word intelligibility, all five variables we
entered (lexical distances, phonological distances, orthographic distances, amount of exposure
and language attitudes) ended up in the final model. The order in which they were entered varies
slightly in the case of amount of exposure, which was the second variable entered in the model
for the written word translation task and the fourth variable entered in the model for the spoken
word translation task.

The only variable which was excluded in the both cloze test models was orthographic distance.
The order in which the variables were entered in the two stepwise regression models was exactly
the same for the written and the spoken cloze test.
Looking at all four regression analyses (see Appendix I), we can observe that phonological distances alone can account for the greatest portion of the variance in all four cases, ranging from 35% in the case of the written cloze test to 57% in the case of the written word translation task. In three out of four cases, language exposure was the next variable to be entered into the model, resulting in an additional 3.5 to 6% of additional variance explained.

Language attitudes were always the last variable entered into all four models, meaning that they can explain the least amount of unique variance. The models were compared using the Akaike Information Criterion (AIC) with a correction for finite sample sizes (AICc). In general, the lower the AICc value, the better the model. In the model related to the results of the spoken word translation task, the AICc change is fewer than two points, which indicates that adding this predictor does not result in a better enough model to justify the complexity of the added independent variable. In the model based on the results of the written cloze test, adding language attitudes as a predictor actually results in a slightly higher AICc (976.09 as opposed to 975.93). So in two out of the four cases, adding language attitudes as a predictor does not improve the overall model.

Table 5.3: Summary of multiple regression analyses with adjusted $R^2$, number of cases and predictors for each model. For full specification of the results see Appendix I.

<table>
<thead>
<tr>
<th>Test</th>
<th>Modality</th>
<th>$R^2$ (N)</th>
<th>Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>Written</td>
<td>.635 (N = 999)</td>
<td>Phonological distance, language exposure, orthographic distance, lexical distance, language attitudes</td>
</tr>
<tr>
<td></td>
<td>Spoken</td>
<td>.664 (N = 1074)</td>
<td>Phonological distance, orthographic distance, lexical distance, language exposure, language attitudes</td>
</tr>
<tr>
<td>Cloze</td>
<td>Written</td>
<td>.438 (N = 938)</td>
<td>Phonological distance, language exposure, morphological distance, syntactic distance, lexical distance, language attitudes</td>
</tr>
<tr>
<td></td>
<td>Spoken</td>
<td>.659 (N = 823)</td>
<td>Phonological distance, language exposure, morphological distance, syntactic distance, lexical distance, language attitudes</td>
</tr>
</tbody>
</table>
Finally, just to check whether our distance measurements were stable and generalizable, we re-ran the regression analyses using the same sets of predictors, only this time with lexical and orthographic distance results which were not based on the testing material in the tasks. This means that we used the list data results to predict the scores of the two cloze tests and text data results to predict the translation tasks scores. The $R^2$ values were higher in three out of four cases, but the differences in the amount of explained variance were never larger than 1%. This indicates that the results are not dependent on the particular sets of the words used in the tasks, which in turn makes them more generalizable.

8. DISCUSSION AND CONCLUSIONS

The main purpose of this chapter was to develop a model of intelligibility of isolated words and texts using the data from six Slavic languages. This work expands on the regression analyses done in Chapter 4 in three important ways. First, we added the results of the written and spoken translation task as the dependent variables in an attempt to predict the intelligibility of isolated words as well as texts. Second, we included extra-linguistic predictors, i.e. language attitudes and language exposure, which were not a part of the analysis in Chapter 4. This proved to be a necessary step since in three out of four models, language exposure turned out to be the second-most important predictor of intelligibility. Third, we conducted the analysis on the level of the individual participant, rather than the level of the language. The change was necessary not only because the extra-linguistic factors are participant-dependent, so it does not make much sense to aggregate them, but also because that way we avoid a potential ecological fallacy.

Our model of word intelligibility could explain 63% of the intelligibility variance in the written mode and 66% of the variance in the spoken model. The model of intelligibility of texts could explain 44% of the variance in the written mode and 66% of the variance in the spoken mode. The regression model with both linguistic and extra-linguistic factors represents an improvement on the model consisting of linguistic factors only from Chapter 4. For the sake of comparison, had we aggregated the data the same way as in Chapter 4 the models here could account for 80% of the variance in the written cloze test (compared to 75% with linguistic factors only) and 89% of the variance in the spoken cloze test (as opposed to 80% in the model with linguistic factors only). Even though using only linguistic factors as predictors of intelligibility does give good results, they can certainly be improved further by including extra-linguistic factors, even in a language area such as Slavic, where exposure to neighboring or at least related languages is somewhat limited.
Why does the model based on the results of the written cloze test explain so much less variance compared to the other three? Looking at the predicted values and the actual intelligibility results as measured by the written cloze test, a few patterns emerged. Our model consistently predicted lower scores than the actual ones for Slovene participants doing the test in Croatian and Polish participants doing the test in Slovak. On the other hand, the scores of Croatian and Slovene participants who had Bulgarian as their test language, as well as the scores of Croatian participants who did the task in Slovak were better than the ones predicted by the model.

Upon closer inspection, it was revealed that in the most extreme cases the participants from these three groups simply had a comprehension score of 0. Surprisingly enough, about 21% of participants who did the written cloze test had a zero score, compared with 16% of participants who did the spoken cloze test, which is generally more difficult. In addition, if we compare the results from this chapter with the ones from Chapter 3 where the aggregated data was used, the difference between the $R^2$ of the two models is much smaller – the model based on the results of the spoken cloze test still had a higher $R^2$, but only by .04, compared to .22 difference found here). We can conclude that the model of intelligibility where we used the scores on the written cloze test as the criterion variable is not as good at predicting individual participant scores as the other three models are, even though this difference almost disappears when trying to predict group performance. At this point we cannot be certain why this is the case. One option might be that some of our participants were able to achieve higher scores than expected due to the fact that the written cloze test provides additional context, which the translation tasks do not. The spoken cloze test has the same structure, but due to the nature of the listening process, sentences are not permanently available and it is impossible to go back and re-listen. In the written cloze test it is possible to go back and forth through the text, just as in actual reading, which in turn might mean that some participants could have been able to maximize their performance by using the overall context well.

As we hypothesized, both lexical distances and phonological distances were highly correlated with the intelligibility scores. What we did not predict is that these two measures were so highly intercorrelated as well. The correlation coefficients were around .9 even in cases where we used the text data for measuring lexical distances and the list data for measuring phonological distances, so the results were based on completely different datasets. At this point we cannot say if this is a feature of Slavic languages in general or not. The consequence of this collinearity is that phonological and lexical distance individually are roughly equally good predictors of intelligibility, it just so happens that phonological distances can explain more unique variance in intelligibility scores. For instance, lexical distance alone could account for 52% of the variance in the results of the spoken cloze test, while phonological distance explains 55% of the variance. A significant correlation between lexi-
cal distance and intelligibility which does not translate into a significant predictor in a regression analysis was found in previous research as well (Gooskens, 2007b).

While it does make sense to include lexical and phonological distances as separate predictors of intelligibility, we must keep in mind that their influence is in fact intertwined. Phonological similarity between language pairs would not have much effect if there were not enough cognates to begin with. In the six languages of our study, the two tend to go hand in hand: the lower the number of cognates the two languages share, the more phonologically different they are likely to be.

Orthographic distance was a significant predictor of intelligibility of words, but not of texts. Since there is no reason to assume that participants use the orthography of their native language more when decoding words in a related language than they do when decoding texts, we propose an alternative explanation. More linguistic variables were used in the model for text intelligibility (five as opposed to three for word intelligibility), therefore it could be the case that orthographic distances simply could not account for any additional variance. In addition, phonological distances reflect the differences between the six languages better than orthographic distances, which are a consequence of different arbitrary decisions on how the sounds of each language should be represented graphically (for a broader discussion on this matter see Heeringa et al. (2013).

Morphological and syntactic distance were significant predictors of text intelligibility. With Slavic languages, the influence of morphology is simply a consequence of the fact that there are so many inflexional affixes. As for syntax, it seems logical to assume that the ways in which clauses and sentences are built in a language will play an important role in the intelligibility of texts. Perhaps this is why most of our participants struggled with Bulgarian, which is a language characterized by many differences in syntax, not found in any other Slavic language from our group (see p. 10 and p. 44).

Since the syntactic measurements did not correlate well either with other linguistic predictors or with genealogical distance between the six Slavic languages, we had reason to believe that our word-order based measurement of syntactic distances was not suitable for a language family such as Slavic, which exhibits relatively free word order of major constituents (see pp. 43-45). Nevertheless, syntax proved to be an important predictor of intelligibility on the Slavic language area, which leads us to conclude that the method still captures the overall pattern. Since syntactic relationships between the six Slavic languages do not fully correlate with the genealogical ones, syntactic distance explains some of the unique variance in our intelligibility model, thus making it a good predictor.
Perhaps future studies will come up with a different way of comparing the syntax of Slavic languages. This might involve a detailed annotation of a corpus for word class, gender, number, case, tense, aspect, syntactic function, etc. and a subsequent alignment for each language separately, similarly to how we measured lexical distances. This way the allowed constructions in each language would be captured more effectively. Another approach might be something akin to the indel method introduced in Heeringa, Swarte, Schüppert, and Gooskens (accepted) which measures the average number of words inserted or deleted in sentences of one language compared to the corresponding sentences in another language.

It should be noted however that the actual model of intelligibility will vary greatly depending on the structure of the languages observed. While we demonstrate that intelligibility of closely related languages can be predicted to a great degree using a combination of linguistic and extra-linguistic factors, the relative contribution of those factors will inevitably vary. So even though the inventory of predictors can probably be generalized across language families, the individual models cannot.

When it comes to extra-linguistic predictors, as expected, language attitudes did not contribute too much to our models of intelligibility. While a positive attitude to a language one is trying to understand could conceivable translate into greater motivation for understanding it, this relationship is generally weak. In all four cases, excluding language attitudes as a predictor would result in .02-.06 lower $R^2$ and a negligible difference in AICs. In two cases, the differences in AICc values between the models with language attitude as a predictor and those without was less than two points, which means that adding language attitude as a predictor in a model of intelligibility does not justify the added complexity. Perhaps our measurement of language attitude, expressed as the rating of language beauty is inadequate. Participants in a testing situation might report slightly different attitudes than the ones they actually hold. Another potential explanation might be that language attitude per se does not play an important role in intelligibility. Instead, it might play an important role in the amount of motivation to understand a related language, which could be the factor we should have been measuring. Positive attitude is one of the components of motivation for learning a foreign language in many different frameworks (Tremblay & Gardner, 1995; Williams & Burden, 1997; Dörnyei, 1998), but we still do not know if this applies to the process of understanding a related variety without any preceding learning. If it does, the relationship between language attitudes and intelligibility might have been too circumstantial, resulting in a small percentage of explained variance.

On the other hand, the other extra-linguistic predictor we used, language exposure, proved to be quite important. We assumed that language exposure will not be an important predictor of
intelligibility simply because there is not much cross-language exposure between the speakers of the six Slavic languages of our study. The only notable exception is Czech and Slovak, but this language pair shares so many linguistic similarities that we believed they alone could account for the high level of intelligibility. Another possible exception is the Croatian-Slovene language pair, where speakers of Slovene are still exposed to (Serbo-)Croatian through tourism and the media. Nevertheless, entering language exposure as a predictor into our models resulted in additional 3-6% of variance explained and a 13-50 point drop in AICc. Therefore, despite the fact that the level of language exposure is not too high in the case of the six Slavic languages observed here, it is not to be neglected as a predictor of intelligibility.

What do all these results tell us? The most important finding is that it is possible to predict the level of intelligibility among closely related languages to a high degree of precision from a combination of linguistic and extra-linguistic factors. The constellation of these factors may vary, but in general the inventory should include linguistic distance, phonological distance, orthographic distance, morphological distance, syntactic distance, language exposure and possibly language attitudes as well.
Mutual intelligibility in the Slavic language area

Other experiments on intelligibility in the Slavic language area with a special focus on Serbian and Croatian