Chapter 3
Introducing the project Mutual Intelligibility between Closely Related Languages (MICReLa)

Throughout this book, there are many references to various investigations by researchers from different parts of the world. Still, the results from the so-called Mutual Intelligibility of Closely Related Languages (MICReLa) project are mentioned most frequently. Therefore, the setup of this project is described in detail in this chapter. It also illustrates in more detail one of the tests presented in the previous chapter, the spoken cloze test (see Section 2.2.3.6). The project aimed to establish the degree of mutual intelligibility of 16 closely related languages within the Germanic, Slavic, and Romance language groups in Europe and to relate the results to measurements of linguistic and extra-linguistic determinants.

The investigation was carried out using a large-scale web-based investigation and included six intelligibility tests. These covered spoken and written communication at different linguistic levels: a cloze test (see Section 2.2.3.6) to test intelligibility at the sentence level, a picture-to-story matching task (see Example 2.17 in Section 2.2.3.5) to test global message understanding at the paragraph level, and a word translation task (see Example 2.2 in Section 2.2.1.1) to test intelligibility of individual words. Over 40.000 people, spread over sixteen first languages (70 combinations of listener and target language), participated in the investigation. For more details about the whole MICReLa investigation and results of the intelligibility tests, see Golubović (2016), Swarte (2016), Gooskens and van Heuven (2017, 2020), and Gooskens et al. (2018).

3.1 Method

In this chapter, only the part of the investigation concerned with the set-up and results of the spoken cloze test is presented, since the focus is on spoken language in this book. Results of the other tests can be found in Gooskens and van Heuven (2017). Of the three spoken tests, the cloze test showed the highest correlation (r = .99) with perceived intelligibility (see Section 2.2.3.1). In addition, the fact that the results cover the entire possible range from almost zero to almost 100 percent correct answers shows that this test differentiates well between different levels of intelligibility. The results could be scored automatically, and therefore, the test was an efficient and objective way of testing text intelligibility. This was valuable since the aim was to test a large number of listeners.

Open Access. © 2024 the author(s), published by De Gruyter. BY-NC-ND This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
https://doi.org/10.1515/9783111134697-003

The results of the spoken cloze test were used to answer the following research questions:

1. What is the general mutual intelligibility of closely related languages within the Germanic, Slavic, and Romance language groups in Europe?
2. What is the inherent mutual intelligibility of closely related languages within the Germanic, Slavic and Romance language groups in Europe?

In addition, linguistic distances and extra-linguistic determinants were quantified, making it possible to answer the following question:

3. How well can general and inherent intelligibility of closely related languages within the Germanic, Slavic and Romance language groups in Europe be explained by means of linguistic and extra-linguistic variables?

3.1.1 Target languages

To ensure a manageable design, the investigation focused on the three major language families in EU member states: Germanic, Slavic, and Romance. The study included all 16 official national languages within these language families (see Figure 3.1), but intelligibility was only tested among speakers of languages within the same language family. For instance, the study examined the mutual intelligibility between Spanish and Portuguese, two Romance languages, and between Czech and Polish, two Slavic languages. However, the mutual intelligibility between Spanish and Czech or between Portuguese and Polish was not tested. If a language is an official language in more than one country, only the variety from the country with most speakers was included. For example, although Dutch is an official national language in both Belgium and the Netherlands, the study only included Netherlandic Dutch as the target language since it has more speakers than Flemish Dutch.

3.1.2 Cloze test

The cloze test was based on four English texts from a set of exercises to prepare students for the Preliminary English Test (PET) at the University of Cambridge. The written forms of the four English texts are presented in Appendix A. The texts all have an intermediate level (B1), as formulated by the Common European Framework of Reference for Languages (Council of Europe 2001), and their contents are culturally neutral. At the B1 level, people can understand the main ideas of clear “standard” speech on familiar topics related to work, school, lei-

Germanic	Romance	Slavic
Danish (DA)	French (FR)	Bulgarian (BU)
Dutch (DU)	Italian (IT)	Croatian (CR)
English (EN)	Portuguese (PT)	Czech (CZ)
German (GE)	Romanian (RO)	Polish (PO)
Swedish (SW)	Spanish (SP)	Slovak (SK)
		Slovene (SL)

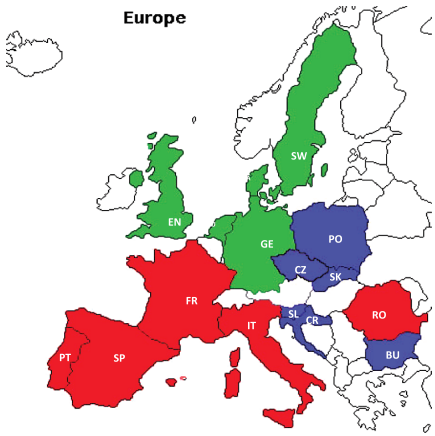

Figure 3.1: Languages included in the MICReLa investigation. Source: Gooskens and van Heuven (2017: 29).

sure, etc. They can also understand the main points of current affairs or personal and professional interest programs on TV and radio, provided that the speech is relatively slow and clear (Council of Europe 2001: 237). The texts were slightly modified to ensure uniformity in total length (around 200 words) and number of sentences (either 16 or 17). Native speakers of the target languages, who were experienced translators with an excellent understanding of English, translated the four texts from English to all the target languages (see Figure 3.1). Per language, at least three translators translated all texts, and the translations were combined into a consensus translation during an interactive session. The translators were instructed to adhere to the original English texts as much as possible while avoiding ungrammatical or unnatural constructions. This approach ensured that the texts were as comparable as possible across languages in terms of content and level of difficulty (see Section 2.1.1).

Each text was divided into 12 audio fragments, consisting of a sentence or a clause. Within each fragment, one word was replaced by a beep (1000 Hz sine wave) of 1 second with 30 milliseconds of silence before and after the beep. Listeners heard all fragments twice and were then given 30 milliseconds to identify the missing word that corresponded with the location of the beep. The twelve response alternatives were continually shown on the screen in three columns: four nouns, four adjectives, and four verbs following the format of the written cloze test shown in Example 2.18 in Section 2.2.3.6. When the listeners hovered their mouse over a word, a translation into their native language was provided. This was implemented to test the intelligibility of the entire text, as unfamiliar

response options could hinder listeners' ability to accurately fill in the gaps, even if they understood the fragment's meaning. To assist the listeners in keeping track of their choices, the inserted words grayed out in the selection area. In case the listeners wished to change their answer, they could select a different word, and the previously selected word would reappear in black.

3.1.3 Speakers and recordings

Recordings were made of six female native speakers for each all the 16 languages in the investigation. All speakers were between 20 and 40 years old and were considered standard speakers of their respective languages. They were instructed to read the texts clearly and at a normal speed. To select the best speakers for the experiment, 16 online surveys were created, each containing six sample recordings from one language. Native listeners of the respective languages were asked to rate each of the six speakers. They rated the speakers on a 5-point scale ranging from "not at all suitable" to "very suitable", answering the question "How suitable is this speaker for presenting the news on national television?". The four best-rated speakers per language were selected for the experiment. The recording of one randomly chosen text was used from each of the selected speakers. This approach helped to minimize the potential influence of voice quality on the results (see Section 2.1.1.5).

3.1.4 Listeners

Since the listeners were tested online, no restrictions concerning their background were set beforehand. Listeners were selected for further analysis afterward by matching the groups according to specific criteria. The study focused on young adults; therefore listeners younger than 18 years and older than 33 years were excluded. The selected listeners originated from the same countries as the speakers (see Figure 3.1). They had all grown up and spent most of their lives in the relevant country and spoke the language of the country as their mother tongue. All listeners had attended or were attending university. Some of the target languages were also school languages. Listeners who had learned the target language for a period exceeding the maximum duration offered in secondary education were excluded because they might not be representative of their peer group. Applying these criteria resulted in the selection of 1833 listeners (426 from the Germanic, 581 from the Romance, and 826 from the Slavic language areas). Among them, 62% of the Germanic, 51% of the Romance, and 43% of the Slavic listeners were male.

Listeners with prior experience with the target language were deliberately included in the sample since the aim was to assess the cross-linguistic intelligibility between the related languages as it is in actual practice, i.e., including the effects of the listeners' education and experience. The obtained results offer an overview of the mutual intelligibility between related languages among a representative group of young, educated Europeans, i.e., the kind of professionals who will meet and want to communicate with one another in international, cross-border contacts. It therefore includes situations of both inherent and acquired intelligibility. This sample is referred to as "general intelligibility" (see Section 1.1).

The researchers conducted a separate analysis of a sub-group of listeners who had not learned the target language at school and had had minimal exposure to it. This allowed for a closer examination of "inherent intelligibility" (see Section 1.1). The intelligibility results from this group of listeners were expected to reflect the traditional genealogic affinities among the languages within each of the three language groups. This subset of listeners included those who had rated their mean exposure below 2.0 (with 1 for no exposure) on six five-point scales and had not learned the target language at school.[10] This selection reduced the number of listeners to 1307. It also reduced the number of language combinations, where more than seven listeners were tested, to 57. This number was regarded as a minimum for a stable analysis. For instance, there were no Dutch listeners between 18 and 33 years who had not learned English in school. Ten of the original 70 language combinations were no longer included, six of which are in the Germanic group. In analyses of inherent intelligibility later in this book, the Romanian listeners were also excluded because their scores were much higher than expected, given the linguistic distances between languages. This may be explained by the fact that most Romanian listeners have greater access to plurilingual resources (see Section 4.2), such as television shows and movies from Spain and Italy with subtitles, compared to other groups of listeners. The scores on the exposure scales, concerning how often they watched television, DVDs, or movies in the target language, were higher for the Romanians (2.30) than for the other Romance listeners (1.50–1.74). Consequently, intelligibility for Romanian listeners cannot be solely attributed to inherent factors.

10 Ideally, listeners with no prior exposure to the target language and no knowledge of other languages that could aid in understanding the target language would be required. However, to obtain an adequate number of listeners, the criteria for participant selection had to be relaxed.

3.1.5 Procedure

The listeners were tested via an online application. They were first asked to complete a questionnaire about their native language, age, gender, and level of education. In addition, they were asked how much exposure they had had to the target language on 5-point scales for six different written and spoken situations (see Section 4.1.1) and for how many years they had learned the language. The responses were used to identify listeners with similar backgrounds, allowing for comparison of the results of the various listeners groups. The answers also made it possible to distinguish post hoc between general and inherent intelligibility (see Section 3.1.4).

In addition, the study included measurements of the listeners' attitudes towards the target language. The listeners were played a short fragment of the Universal Declaration of Human Rights, which was recorded by the same four speakers used for recording the test material. They were then asked to rate the beauty of the target language on a scale ranging from 1 "very ugly" to 5 "very beautiful". The results of the exposure and attitude measurements are detailed in Appendix C.

In the last part of the questionnaire, the listeners judged their own understanding of the target language (opinion testing, see Section 2.2.3.1). Prior to being exposed to any stimuli, they indicated how well they thought they would be able to understand the target language (i.e., judged intelligibility), on a 5-point scale from 1 = "not at all" to 5 = "very well". Whether this question referred to the understanding of spoken or written language was not made explicit. The participants filled in the questionnaire immediately before they did the intelligibility test. Once they had completed the test, they were asked to indicate how well they thought they had understood the target language (i.e., perceived intelligibility) on the same scale as before the stimulus presentation. The results of these measurements are presented in Gooskens and van Heuven (2017). They correlated highly with the results of the spoken cloze test.

Once the listeners completed the questionnaire, the intelligibility test began. The target language was chosen from the same language family as that of the listener, which could be Germanic, Romance, or Slavic (see Figure 3.1). In total, there were 64 different tests (4 texts and 16 languages). Each listener was tested on a randomly selected text and language (never their own native language). Together, the questionnaire and test lasted approximately 15 minutes.

3.2 Results

In Figure 3.2, the results of the cloze test for the three language areas are presented. The table with mean results is presented in Appendix B. In this section,

the intelligibility results are summarized. For a more detailed discussion, see Gooskens et al. (2018) and Gooskens and van Heuven (2020).

The mean general intelligibility score in the Germanic language group involving all listeners is 40%. However, there are significant variations in intelligibility scores between different language combinations. A one-way ANOVA showed that the effect of language pair was highly significant ($F(19, 406) = 74.3$, $p < .001$, $\eta^2 = .777$). Nine listener groups scored below 20%, while four groups had high mean scores (>80%). These latter groups all involve English as the target language. These high scores were expected since English is widely used as a lingua franca and is an important language in the educational system. The lowest scores were observed for language combinations where listeners and target languages belonged to different main branches of the Germanic language tree (West or North), such as Swedish listeners tested with Dutch stimuli and Dutch listeners tested with Swedish stimuli, as well as the intelligibility of Dutch for English listeners. There was also a middle group of seven language combinations with intelligibility scores between 20% and 80%, and four of them involved German as the target language. German is taught as school language in many Germanic countries, which explains why German is well understood even among listeners from the North Germanic group. German listeners understand some Dutch (31% correct responses), and the mutual intelligibility between Swedes and Danes was found to be fairly high (57% and 63% correct).

When examining the subsample of listeners with limited exposure to the target language (as shown by the grey bars in Figure 3.2), the impact of language pair on inherent intelligibility is observed to be smaller than on general intelligibility, as determined by a one-way ANOVA, $F(13, 217) = 12.7$ ($p < .001$, $\eta^2 = .432$). Only Swedes and Danes achieve a moderate level of mutual comprehension (44% and 56%). Germans can understand some Dutch (26%), and receptive multilingualism is commonly used for communication, particularly in the Dutch-German border areas (Ház 2005). However, at first contact, speakers of Dutch and German are typically only able to communicate at a basic level.

The average score of the Romance language group is slightly lower than that of the Germanic group, with a mean of 37%. However, the scores for the various language combinations are more evenly spread across the scale. In several configurations of Romance languages, some degree of mutual intelligibility is present. Spanish is the language that is easiest to understand for all listener groups, with a mean score of 57% correct answers across all listener groups. Romanian is the hardest to understand, with a mean score of 13% for all listeners. The Portuguese listener group has the highest mean score (47%), but it is notable that, on average, Romanians understand the other languages almost as well (45% correct).

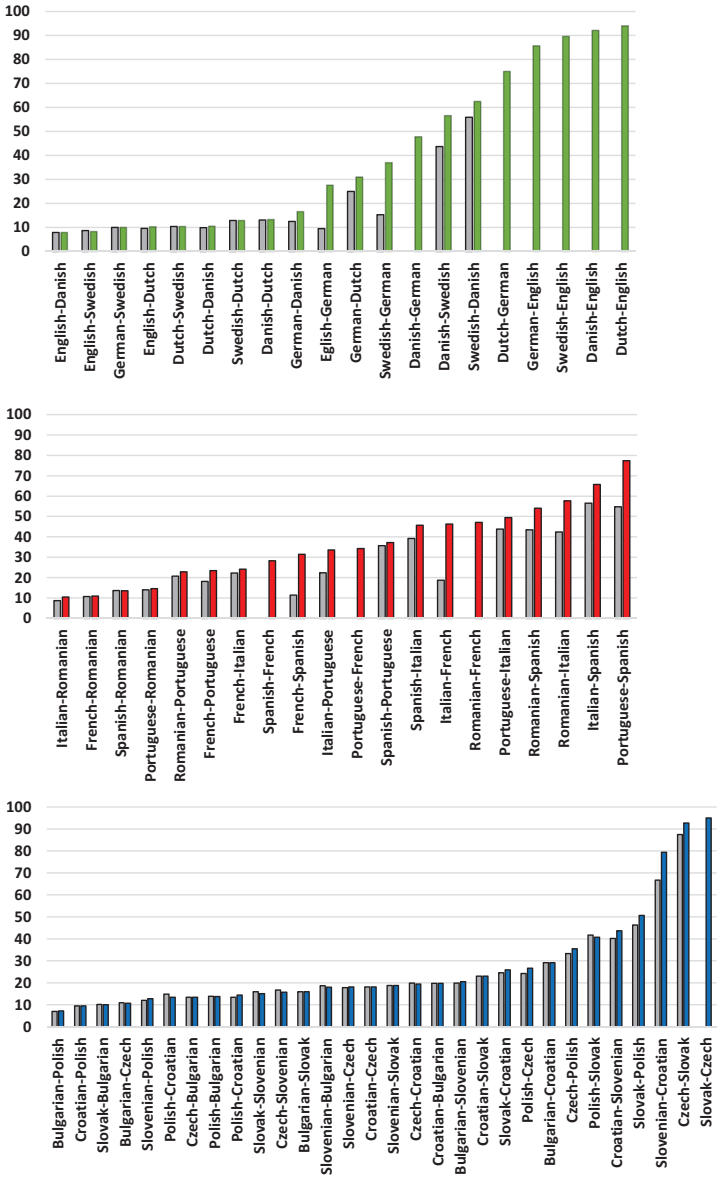

Figure 3.2: Results of spoken cloze tests in the Germanic (top), Romance (middle) and Slavic (bottom) language areas. For each language combination, the listener language is given first and the target language second (e.g., "English-Danish" means "English native speakers listening to Danish"). Colored bars show the mean results of all listeners ordered from lowest to highest percentage of correct answers (general intelligibility), and grey bars show mean results for listeners with minimal exposure (inherent intelligibility). Adapted from Gooskens et al. (2018: 179 and 184).

Upon analyzing the listeners with minimal exposure to the target language (inherent intelligibility), it was found that many groups possess some level of understanding of the target language, particularly when it involves Spanish or Italian. Most groups of listeners tested in French have learned French at school and are excluded from this analysis. Only in the Italian group are there enough listeners left for an analysis, and they have a low intelligibility score (19%) when listening to French.

The Slavic language group has a lower mean intelligibility (28%) than the other language families, with 18 out of 30 language combinations showing scores below 20%. A one-way ANOVA on the 70 language pairs did not show a significant effect of language pair, $F(2, 67) = 1.6$ ($p = .204$, $\eta^2 = .046$). There are some exceptions to the low scores, including high mutual intelligibility between Czech and Slovak (93% and 95%), Croatian and Slovenian (44% and 79%), and moderate mutual intelligibility between Polish and Slovak (41% and 51%). Slovak has the highest general intelligibility score (38% correct across all language combinations), and Slovaks are also the best at understanding other closely related languages (39%). Bulgarian is difficult for all groups to understand (15% correct), and the Bulgarians understand very little of the other Slavic languages in the study (17%).

Listeners from the Slavic area in the sample have very little exposure to the other Slavic languages, and studying them is uncommon (see Appendix C). Therefore, there are hardly any differences between the general and inherent intelligibility results. An exception is Slovenians listening to Croatian. Two-thirds of the Slovenians scored above 2.0 on the exposure scale, while the remaining third had lower scores in the intelligibility test (72%) than the whole group (79%). All Slovaks listening to Czech had a mean exposure score above 2.0 and were excluded from the analysis of listeners with minimal exposure to the target language.

3.3 Conclusions

The MICReLa project aimed to determine the degree of mutual intelligibility between 16 closely related languages (70 language combinations) from three language families in Europe. To achieve this, a spoken cloze test was used, and two sets of data were presented. The first set of data represented the mutual intelligibility between young, educated Europeans (general intelligibility). The second set represented a group of listeners with minimal exposure to the target language (inherent intelligibility).

The first selection of listeners allowed for an examination of the potential for receptive multilingualism as a mode of communication in Europe. The results showed that the Danes, who had no prior exposure to Swedish, scored 44% correct in the cloze test, while the Swedes scored higher at 56%. Previous literature

suggests that with some effort, Swedes and Danes can communicate effectively using their own languages (receptive multilingualism) without prior exposure (Delsing and Lundin Åkesson 2005, see also Section 8.5). Hence, a score of 40% correct could serve as a potential communication threshold and put the remaining results into perspective. Among the Germanic listener groups, eight scored above 40%. However, apart from the Swedish and Danish groups, the other groups learned the target language (English or German) at school. The German and English listener groups scored very poorly and would be unable to engage in a conversation using receptive multilingualism.

Within the Romance group, eight subgroups scored above the tentative 40% threshold for effective communication. Romanian listeners exhibit high scores across all languages except for Portuguese. However, since Romanian is a difficult language for other speakers of Romance languages, their scores are low, making receptive multilingualism an unlikely successful mode of communication with Romanians. The same holds for French speakers. Results indicate that the combination of Italian and Spanish may be the most successful for receptive multilingualism, with the combination of Spanish and Portuguese also showing potential success despite some challenges for Spanish speakers in understanding Portuguese.

Six groups in the Slavic language family have attained scores above the 40% threshold for effective communication. Czech and Slovak speakers display the highest level of mutual intelligibility among all tested European language combinations. Additionally, receptive multilingualism is likely to be a successful mode of communication for Slovenes and Croatians, as well as for speakers of Polish and Slovak. The remaining Slavic listeners have relatively low scores, indicating potential challenges for successful communication through receptive multilingualism.

It is worth noting that the translators and speakers involved in creating the speech materials for the MICReLa investigation were unaware that their recordings would be presented to non-native listeners. This suggests that the speech materials may be more challenging for non-native listeners to process compared to real-life interactions between receptive multilinguals, where speakers can adjust their speech to the listener, and the listener can ask for clarification if needed (see Section 4.6). Furthermore, the duration of the audio fragments may have been too brief for listeners to learn how to adjust their perceptual categories or develop other strategies to cope with deviant speech input. As a result, the findings of the study should be considered conservative estimates of the potential success of receptive multilingual communication in European languages.

Based on the findings, it can be inferred that speakers of several language combinations have the potential to communicate effectively in their respective languages, given the current circumstances. However, it would be feasible to implement receptive multilingualism for many additional language combinations

with little effort. As Golubović (2016) demonstrated, see Section 4.1, a short four-and-a-half-hour instruction was sufficient to significantly improve Croatian comprehension among Czech speakers. Future research is needed to determine which language combinations are suitable for receptive multilingualism and how to achieve the required level of intelligibility for successful communication efficiently. Section 8.5 delves into a more detailed discussion of receptive multilingualism as a means of communication.

The second selection, reflecting inherent intelligibility, makes it possible to look at the linguistic basis for mutual intelligibility. This measure shows how well listeners can understand a language without previous knowledge of the language. Only Swedes and Danes reach the 40% threshold in the Germanic group. In the Romance language family five groups of listeners score higher than 40%. However, these mainly include Portuguese or Romanian listeners, and since there is often an asymmetric relationship in language combinations involving these languages, the inherent basis for mutual intelligibility is still low. For example, Portuguese listeners scored 62% correct when listening to Spanish, but Spanish listeners only scored 36% when listening to Portuguese. In Chapter 6, reasons for asymmetric intelligibility are discussed in detail.

By measuring inherent intelligibility, it is also possible to establish the relationship between intelligibility scores and the genealogic characterization of the European languages included in the investigation. In Section 8.4, the traditional genealogic classifications of the 16 languages involved in the MICReLa investigation are presented by means of language trees, and it is shown how intelligibility is reflected in the traditional family trees.

To conclude, Chapter 3 presented some of the intelligibility measurements in the MICReLa project. To interpret the results and better understand why some closely related languages are easier to understand than others, it is important to identify factors that may explain the intelligibility results. In Chapters 4 and 5, various linguistic and extra-linguistic determinants of intelligibility are discussed. Some of the determinants were also quantified within the MICReLa project, making it possible to correlate them with the intelligibility results. In Chapter 7, the relative contribution of the various determinants is established through regression analyses including general and inherent intelligibility scores as well as quantifications of the determinants. In Appendix B, Appendix C, and Appendix D, tables with general and inherent intelligibility results of the spoken cloze test as well as measures of linguistic (phonetic, lexical, and syntactic distances) and extra-linguistic determinants (attitude and exposure scores) used for these analyses are presented. All the determinants are included in regression analyses to get an impression of the relative contribution of the various determinants to the general and inherent intelligibility of the European languages included in the investigation.